EXPLICIT VECTOR WAVE DIGITAL FILTER MODELING OF CIRCUITS WITH A SINGLE BIPOLAR JUNCTION TRANSISTOR

Oliviero Massi, Riccardo Giampiccolo, Alberto Bernardini, and Augusto Sarti

Dipartimento di Elettronica, Informazione e Bioingegneria
Politecnico di Milano
Piazza L. Da Vinci 32, 20133 Milano, Italy
oliviero.massi@polimi.it | riccardo.giampiccolo@polimi.it | alberto.bernardini@polimi.it
augusto.sarti@polimi.it

ABSTRACT

The recently developed extension of Wave Digital Filters based on vector wave variables has broadened the class of circuits with linear two-port elements that can be modeled in a modular and explicit fashion in the Wave Digital (WD) domain. In this paper, we apply the vector definition of wave variables to nonlinear two-port elements. In particular, we present two vector WD models of a Bipolar Junction Transistor (BJT) using characteristic equations derived from an extended Ebers-Moll model. One, implicit, is based on a modified Newton-Raphson method; the other, explicit, is based on a neural network trained in the WD domain and it is shown to allow fully explicit implementation of circuits with a single BJT, which can be executed in real time.

1. INTRODUCTION

Virtual Analog (VA) modeling [1] is an audio signal processing field of research which focuses on the digital emulation of analog audio equipment. Over recent years, a lot of research effort has been dedicated to the faithful and efficient digital implementation of circuit nonlinearities, which concur to the well appreciated timbral characteristics of analog audio gear. Forming the core of countless amplifier models used in a broad range of audio equipment, Bipolar Junction Transistors (BJTs) are arguably the most relevant components in this regard.

In the literature, VA modeling approaches can be generally divided into two categories: *black-box* approaches that infer a global model of a reference circuit relying on pairs of observed input/output data using, e.g., Volterra series [2] or neural networks [3], and *white-box* approaches that emulate the reference circuit by simulating the corresponding system of ordinary differential equations, e.g., using state-space methods [4], the port-Hamiltonian method [5], or Wave Digital Filters (WDFs) [6].

Among white-box techniques, WDFs have proved to be a very promising framework for creating digital models of reference analog circuits. Developed in the 70s by A. Fettweis to derive digital implementations of passive analog filters [7], WDFs rely on a port-wise linear mapping of Kirchhoff pairs of variables (voltage and current) into pairs of wave variables (incident and reflected) with the introduction of a scalar free parameter per port called *port resistance*. Circuit elements and topological connection networks are modeled separately, in a modular fashion. The reference circuit is represented in the Wave Digital (WD) domain as an interconnection of one-port and multi-port WD blocks, characterized by implicit scattering relations between wave variables, called *delay-free loops*. By proper choice of port resistances (through the so-called *adaptation* process) and making use of stable discretization methods [8] (e.g., trapezoidal rule), circuits containing linear one-port elements can be implemented in a fully explicit fashion, i.e., removing all delay-free loops [7]. Furthermore, a WD structure which relies on a scalar definition of wave variables is able to accommodate a single nonlinear one-port circuit element by placing it at the root of a tree-like structure and connecting it to an adapted (reflection-free) port of a WD junction [9, 10]. In order to design an explicit WDF, since nonlinear one-port elements cannot be generally adapted, the reflection-free port of the junction is necessary, otherwise the instantaneous wave reflection from the junction back to the nonlinearity would result in a delay-free loop [9, 10]. Such a WDF design methodology, originally conceived for static nonlinearities, has been further extended to reference circuits with a single nonlinear one-port element with memory [11, 12].

The previous considerations on traditional WDFs with a single nonlinear one-port element are not directly applicable to WDFs with a single nonlinear multi-port element. In fact, in that case, computability issues might arise due to unavoidable delay-free loops. More in general, when both ports of a generic two-port element are connected to the same multi-port junction, as in Fig. 1, a double delay-free loop passing through the element and the junction always arises, and it cannot be eliminated even in the case in which both ports of the junction are locally made reflection-free [13, 14, 15].

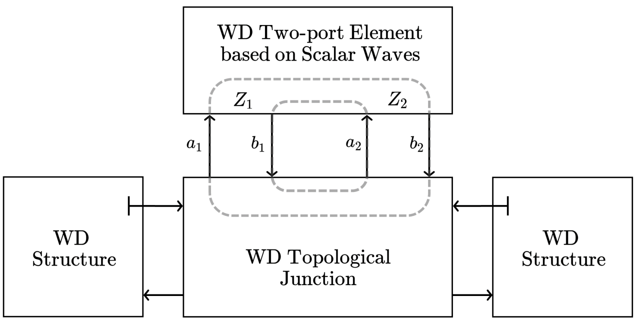

Figure 1: *Example of traditional WDF based on a scalar definition of waves that includes a WD two-port element. The dashed circles indicate two delay-free loops that arise in the WDF.*

Several works in the literature of WDFs are devoted to the modeling of multi-port nonlinearities and specifically to the modeling of BJTs. Werner *et al.* [16] proposed a hybrid Kirchhoff/Wave approach based on the K-method [4, 17, 18] to solve circuits with multiple/multi-port nonlinearities that are grouped at the root of the digital structure. Olsen *et al.* [19] used the Newton-Raphson (NR) method to speed-up the computation of the same digital structures described in [16]. In [20], *et al.* defined a general WD model for 3-terminal devices, whose number of ports can vary from 1 to 6, along with explicit WD realizations of MOSFET and JFET transistors and an implicit WD realization of BJTs based on the Ebers-Moll model and implemented using a robust modified Newton-Raphson solver. Kolonko *et al.* [21] introduced a modified Ebers-Moll model where the forward and reverse conducting diodes are considered individually and then connected through a suitable WD junction.

All the WD models of BJT transistors presented in the literature are based on scalar definitions of port variables and this often causes unavoidable delay-free loops when two-port models are employed. The recently introduced Vector WDFs [22] generalized the scalar port-wise definition of wave variables in linear two-port elements to a vector definition. Eventually, this allowed to overcome several computability issues, such as delay-free loops that are formed when connecting a two-port to a topological junction.

In this paper, we apply Vector WDFs to the implementation of circuits containing a single two-port WD model of a BJT. The nonlinear two-port element can be accomodated at the root of the WDF, which allows to implement explicit WD structures preserving the modularity of Fettweis' traditional WDFs [7]. We first propose an extension of the Ebers-Moll model [23] of a BJT and we present two vector wave-based WD realizations. The first one is implicit and it relies on a modified Newton-Raphson method [20] to solve the characteristic equation derived from the extended Ebers-Moll model. The second one is explicit and it is characterized by a Multi-Layer Perceptron network trained in the WD domain; this accompanies a growing trend in the WDF literature [24, 25, 26], in which data-driven neural models of nonlinear devices are connected to "traditional" WD blocks.

The remainder of this manuscript is organized as follows. In Section 2, the design of WDFs containing a single nonlinear two-port element is discussed. Section 3 introduces the two proposed BJT models based on vector waves. In Section 4, the developed methods are applied for the emulation of the input stage of a guitar distortion pedal. Conclusions are drawn in Section 5.

2. VECTOR WAVE DIGITAL FILTERS

2.1. Scalar Waves

The design of WDFs is based on a port-wise description of a reference analog circuit. Circuit elements and topological connection networks are modeled using one- or multi-port WD blocks characterized by scattering relations. In traditional WDFs based on voltage waves, each pair of Kirchhoff variables at a generic port j of a circuit element, i.e., the port voltage v_j and the port current i_j, is substituted with a pair of scalar WD variables defined as [7]

$$a_j = v_j + Z_{jj} i_j \quad , \quad b_j = v_j - Z_{jj} i_j \, , \tag{1}$$

where a_j is the incident wave, b_j is the reflected wave and $Z_{jj} \neq 0$ is a free real-valued parameter, usually called *reference port resistance* and here renamed as *reference one-port resistance*. This free parameter is set to *adapt* linear one-port circuit elements, thus obtaining explicit WD scattering relations in the discrete-time domain in which the reflected wave does not instantaneously depend on the incident wave [7, 8].

N-port topological connection networks, characterized by a vector of port voltages $\mathbf{v}_{\mathrm{J}} = [v_{\mathrm{J}1}, \ldots, v_{\mathrm{J}N}]^T$ and a vector of port currents $\mathbf{i}_{\mathrm{J}} = [i_{\mathrm{J}1}, \ldots, i_{\mathrm{J}N}]^T$, are modeled in the WD domain using N-port junctions characterized by the wave variables

$$\mathbf{a}_{\mathrm{J}} = \mathbf{v}_{\mathrm{J}} + \mathbf{Z}_{\mathrm{J}} \mathbf{i}_{\mathrm{J}} \quad , \quad \mathbf{b}_{\mathrm{J}} = \mathbf{v}_{\mathrm{J}} - \mathbf{Z}_{\mathrm{J}} \mathbf{i}_{\mathrm{J}} \, , \tag{2}$$

where $\mathbf{a}_{\mathrm{J}} = [a_{\mathrm{J}1}, \ldots, a_{\mathrm{J}N}]^T$ is the vector of waves incident to the junction, $\mathbf{b}_{\mathrm{J}} = [b_{\mathrm{J}1}, \ldots, b_{\mathrm{J}N}]^T$ is the vector of waves reflected by the junction, while $\mathbf{Z}_{\mathrm{J}} = \mathrm{diag}\,[Z_1, \ldots, Z_N]$ is a diagonal matrix having one-port resistances as diagonal entries. The relation between $\mathbf{a}_{\mathrm{J}}$ and $\mathbf{b}_{\mathrm{J}}$ is

$$\mathbf{b}_{\mathrm{J}} = \mathbf{S} \mathbf{a}_{\mathrm{J}} \, , \tag{3}$$

where $\mathbf{S}$ is a $N \times N$ scattering matrix. General formulas for computing the scattering matrix of arbitrary reciprocal or non-reciprocal connection networks in the WD domain are discussed in [27, 15].

When two-port circuit elements are present, a WDF structure based on scalar port-wise wave definition is generally affected by computability problems. In Fig. 1, a generic WDF structure based on scalar wave variables includes a (linear or nonlinear) two-port element whose ports are both connected to the same topological junction. As highlighted by the dashed paths, two delay-free loops involving cross-dependencies between wave variables are unavoidably formed in the WDF: they cannot be eliminated through any choice of the free parameters.

2.2. Vector Waves

With the purpose of encompassing both ports of the same two-port element, we introduce the following *vector definition of wave variables* [22], which generalizes (1):

$$\begin{aligned} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} &= \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} + \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} \\ \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &= \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} - \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} . \end{aligned} \tag{4}$$

$[v_1, v_2]^T$ is the vector of the two port voltages, $[i_1, i_2]^T$ is the vector of the two port currents, $[a_1, a_2]^T$ is the vector of the waves incident to the two-port element, $[b_1, b_2]^T$ is the vector of the waves reflected by the two-port element and

$$\mathbf{Z}_{1,2} = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \tag{5}$$

is a full-rank 2×2 matrix of real free parameters $Z_{\imath\jmath}$, with $\imath \in 1, 2$ and $\jmath \in 1, 2$, which we refer to as *reference two-port resistance*. Since $\mathbf{Z}_{1,2}$ is full-rank, we have

$$|\mathbf{Z}_{1,2}| = \det\,[\mathbf{Z}_{1,2}] = Z_{11} Z_{22} - Z_{12} Z_{21} \neq 0 \, . \tag{6}$$

The inverse mapping from WD variables to Kirchhoff variables can be expressed as

$$\begin{aligned} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} &= \frac{1}{2} \left(\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \right) \\ \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} &= \frac{1}{2|\mathbf{Z}_{1,2}|} \begin{bmatrix} Z_{22} & -Z_{12} \\ -Z_{21} & Z_{11} \end{bmatrix} \left(\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} - \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \right) . \end{aligned} \tag{7}$$

Analogously to one-port linear elements, a two-port WD element based on the proposed vector definition of waves (4) can be adapted through the assignment of $\mathbf{Z}_{1,2}$, which eliminates the instantaneous dependency between the vector of reflected waves $[b_1, b_2]^T$ from the vector of incident waves $[a_1, a_2]^T$ in the scattering relation .

The approaches described for the WD modeling of topological connection networks can be generalized for the design of topological multi-port WD junctions based on mixed scalar and vector definitions of waves [22]. An N-port topological junction which is connected to other WD blocks through P two-port connections and $N-2P$ one-port connections is characterized by the same wave variable definition in (2), where $\mathbf{Z}_\mathrm{J}$ is in this case a *full-rank block-diagonal matrix* of free parameters (not simply a diagonal matrix as in WDFs solely based on scalar waves). Assuming, without loss of generality, to number the P two-port connections before the $N-2P$ one-port connections, $\mathbf{Z}_\mathrm{J}$ can be written as

$$\mathbf{Z}_\mathrm{J} = \begin{bmatrix} \mathbf{Z}_{1,2} & \dots & \mathbf{0} & [0,0]^T & \dots & [0,0]^T \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{Z}_{2P-1,2P} & [0,0]^T & \dots & [0,0]^T \\ [0,0] & \dots & [0,0] & Z_{2P+1} & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ [0,0] & \dots & [0,0] & 0 & \dots & Z_N \end{bmatrix} \tag{8}$$

where $\mathbf{Z}_{1,2}, \dots, \mathbf{Z}_{2P-1,2P}$ are 2×2 full-rank submatrices and $Z_{2P+1}, \dots, Z_N$ are scalar parameters different from zero.

The resulting WD junctions can be used to interconnect one-port elements based on traditional scalar waves and two-port elements based on vector waves, thus solving many of the computability problems that would arise using just scalar definitions, while preserving modularity, i.e., modeling circuit elements and topology in a separate fashion.

2.3. WDFs with a Single Vector Nonlinear Two-Port Element

WDFs with a single nonlinearity can be organized into tree-like structures called *connection trees* [10, 8]. Three types of constitutive blocks can be identified in such WD connection trees: the *root*, i.e., the nonlinearity, which has no upward-facing ports and whose downward-facing ports cannot be adapted; *nodes* (typically multi-port topological junctions), which have one upward-facing port and one or more downward-facing ports; *leaves*, which have upward-facing ports and no downward-facing ports [10, 8]. A WD structure of the sort can be solved without employing any iterative solver if the nonlinearity is characterized by an explicit mapping in the WD domain. In fact, delay-free loops can be removed through the *adaptation* of all the leaves and upward-facing ports of junctions. It follows that a WDF based on traditional scalar waves and containing a single nonlinear one-port element with an explicit scattering relation can be implemented in a fully explicit fashion. As outlined in the above description of Fig. 1, a WDF based on scalar waves with a single nonlinear two-port placed at the root would instead be characterized by unavoidable delay-free loops. However, the adoption of a vector definition of waves for modeling the nonlinear two-port element allows us to make also this kind of WD structures *fully explicit*, again, under the assumption that the nonlinearity is characterized by an explicit WD mapping. In fact, as shown in Fig. 2, through proper assignment of the free parameters Z_{11}, Z_{12}, Z_{21} and Z_{22} constituting the reference two-port

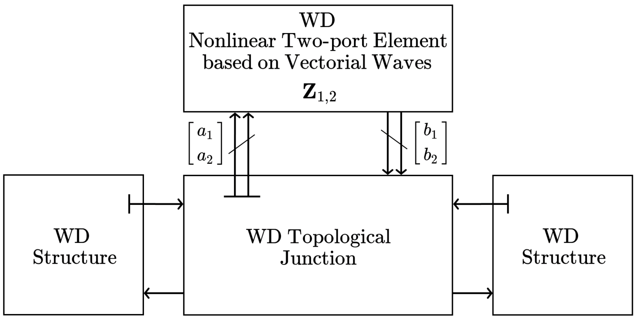

Figure 2: *WD structure featuring a WD nonlinear two-port element based on vector waves connected to a topological junction. The T-shaped stub indicates port adaptation.*

resistance $\mathbf{Z}_{1,2}$, it is possible to make the pair of junction ports at which the nonlinear two-port element is connected reflection-free. This means that the nonlinear element does not introduce any delay-free loop despite being connected to the junction through a double port connection.

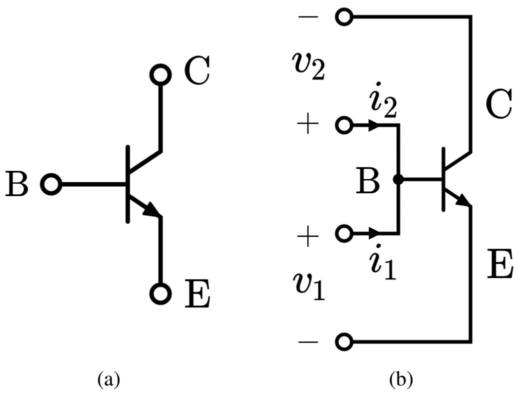

Figure 3: *(a) Symbol of a BJT as a three-terminal device and (b) corresponding two-port element definition.*

3. BJT MODELS BASED ON VECTOR WAVES

A BJT is an electronic device characterized by three terminals called *base*, *emitter* and *collector* - shown in Fig. 3(a) as node B, E and C, respectively. Through the modeling approach discussed in [20], it can be described as a two-port element, as shown in Fig. 3(b). In fact, its behavior is completely described by the voltages v_1 (across base-emitter port) and v_2 (across base-collector port). In the following, we introduce an extension to the Ebers-Moll model (EMM) [23], probably the most widespread large signal BJT model appearing in the literature. As shown in Fig. 4, a series and a parallel resistor are introduced at each of the two ports: $(R_{\mathrm{s1}}, R_{\mathrm{p1}})$ at port 1 and $(R_{\mathrm{s2}}, R_{\mathrm{p2}})$ at port 2. The series resistances encapsulate the contribution of several structural resistances of the device, while the parallel ones are mainly due to current leakage at the two *p-n* junctions forming the BJT. Typically, R_{p1}, R_{p2} have high values, while R_{s1}, R_{s2} are very low [28]. The introduced resistors also attenuate numerical problems that might

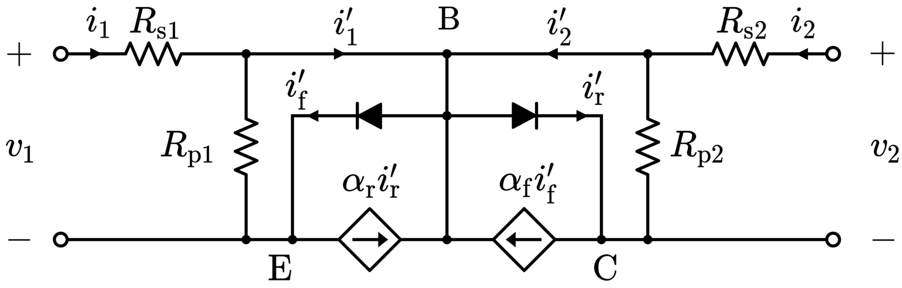

Figure 4: *Proposed extended Ebers-Moll model.*

arise when regions of the v-i characteristic curves of the diodes with extremely high (or extremely low) slopes are visited, especially when dealing with large-amplitude signals [8, 29].

Our extended EMM is mathematically described by the following system of equations

$$\begin{cases} i_1 = i'_1 + \frac{v_1 - i_1 R_{\mathrm{s1}}}{R_{\mathrm{p1}}} \\ i_2 = i'_2 + \frac{v_2 - i_2 R_{\mathrm{s2}}}{R_{\mathrm{p2}}} \end{cases}, \tag{9}$$

where

$$\begin{aligned} i'_1 &= I_{\mathrm{s1}} \left(e^{\left(\frac{v_1 - i_1 R_{\mathrm{s1}}}{\eta_{\mathrm{f}} V_{\mathrm{t}}} \right)} - 1 \right) - \alpha_{\mathrm{r}} I_{\mathrm{s2}} \left(e^{\left(\frac{v_2 - i_2 R_{\mathrm{s2}}}{\eta_{\mathrm{r}} V_{\mathrm{t}}} \right)} - 1 \right) \\ i'_2 &= I_{\mathrm{s2}} \left(e^{\left(\frac{v_2 - i_2 R_{\mathrm{s2}}}{\eta_{\mathrm{r}} V_{\mathrm{t}}} \right)} - 1 \right) - \alpha_{\mathrm{f}} I_{\mathrm{s1}} \left(e^{\left(\frac{v_1 - i_1 R_{\mathrm{s1}}}{\eta_{\mathrm{f}} V_{\mathrm{t}}} \right)} - 1 \right). \end{aligned} \tag{10}$$

In equations (9) and (10), i_1 is the current through the base-emitter port, i_2 is the current through the base-collector port, α_{f} is the forward common-base current gain, α_{r} is the reverse common-base current gain, v_1 is the base-to-emitter voltage, v_2 is the base-to-collector voltage, V_{t} is the thermal voltage, I_{s1} is the saturation current of the base-emitter *p-n* junction, I_{s2} is the saturation current of the base-collector *p-n* junction, η_{f} is the ideality factor of the base-emitter *p-n* junction and η_{r} is the ideality factor of the base-collector *p-n* junction.

3.1. Implicit WD Model based on Modified NR Method

Reformulating (9), the BJT two-port model in Fig. 4 can be characterized by the following system of nonlinear equations in the Kirchhoff domain

$$\mathbf{h}\left(\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}, \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} \right) = \begin{bmatrix} i'_1 + \frac{v_1 - i_1 R_{\mathrm{s1}}}{R_{\mathrm{p1}}} - i_1 \\ i'_2 + \frac{v_2 - i_2 R_{\mathrm{s2}}}{R_{\mathrm{p2}}} - i_2 \end{bmatrix} = \mathbf{0}, \tag{11}$$

where $[v_1, v_2]^T$ is the vector of port voltages, $[i_1, i_2]^T$ is the vector of port currents. Since (11) is composed of two coupled transcendental implicit equations, finding an explicit scattering relation in the WD domain is not possible in general, even though some explicit WD mappings based on simplifications have been proposed in some particular cases [30]. To overcome that difficulty, we adjust the NR method to the local resolution of two-port nonlinear WD blocks based on a vector wave definition. To this aim, relationships between port vector variables, (4) and (7), are rewritten as follows [30],

$$\begin{bmatrix} i_1 \\ i_2 \end{bmatrix} = \frac{1}{|\mathbf{Z}_{1,2}|} \begin{bmatrix} Z_{22} & -Z_{12} \\ -Z_{21} & Z_{11} \end{bmatrix} \left(\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} - \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \right), \tag{12}$$

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = 2 \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} - \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}. \tag{13}$$

Replacing (12) into (11) leads to a vector nonlinear equation $\mathbf{g}(\boldsymbol{\varphi}) = \mathbf{0}$, where $\boldsymbol{\varphi} = [v_1, v_2]^T$. Given the vector wave $[a_1, a_2]^T = \left[a_1^{(k)}, a_2^{(k)}\right]^T$ incident to the nonlinear two-port element and the reference two-port resistance $\mathbf{Z}_{1,2}$, which is set in such a way to make reflection-free the pair of junction ports to which the nonlinear two-port element is connected, the equation $\mathbf{g}(\boldsymbol{\varphi}) = \mathbf{0}$ can be solved using the NR algorithm characterized by the following update rule

$$\boldsymbol{\varphi}^{(k+1)} = \boldsymbol{\varphi}^{(k)} - \left[\mathbf{J}\left(\boldsymbol{\varphi}^{(k)}\right)\right]^{-1} \mathbf{g}\left(\boldsymbol{\varphi}^{(k)}\right), \tag{14}$$

where $\boldsymbol{\varphi}^{(k)} = \left[v_1^{(k)}, v_2^{(k)}\right]^T$ and $\boldsymbol{\varphi}^{(k+1)} = \left[v_1^{(k+1)}, v_2^{(k+1)}\right]^T$ indicate the values of $\boldsymbol{\varphi}$ evaluated at iterations k and $k+1$, respectively, and $\left[\mathbf{J}\left(\boldsymbol{\varphi}^{(k)}\right)\right]^{-1}$ is the inverse of the Jacobian matrix of $\mathbf{g}$ evaluated at $\boldsymbol{\varphi}^{(k)}$. Solving $\mathbf{g}(\boldsymbol{\varphi}) = \mathbf{0}$ requires:

- taking a suitable initial guess $\boldsymbol{\varphi}^{(0)}$;
- repeating (14) up to convergence, i.e., up to the case in which $||\boldsymbol{\varphi}^{(k+1)} - \boldsymbol{\varphi}^{(k)}|| < \epsilon_v$ and $||\mathbf{g}\left(\boldsymbol{\varphi}^{(k+1)}\right)|| < \epsilon_g$, where ϵ_v and ϵ_g are small positive scalar thresholds.

Iterative solvers based on the NR method are arbitrarily accurate and generally more efficient than tabulation methods [31]. The main drawback of such iterative solvers is that their convergence is generally not ensured, since it strongly depends on the chosen initial guess $\boldsymbol{\varphi}^{(0)}$ and it is further compromised by numerical issues related to finite word length representation. To mitigate the effect of such problems, the modified NR (MNR) method introduced in [20] enforces control over the variables $v_1^{(k)}$ and $v_2^{(k)}$ iteration by iteration, through the use of compensation functions, ϕ_1 and ϕ_2, designed to prevent overshooting and to increase the NR method convergence rate (we hereby omit their definition for reasons of space; the reader is kindly referred to [20] for a detailed analysis on the topic). The update equation (14) is thus modified according to

$$\tilde{\boldsymbol{\varphi}}^{(k+1)} = \boldsymbol{\varphi}^{(k)} - \left[\mathbf{J}\left(\boldsymbol{\varphi}^{(k)}\right)\right]^{-1} \mathbf{g}\left(\boldsymbol{\varphi}^{(k)}\right), \tag{15}$$

where $\tilde{\boldsymbol{\varphi}}^{(k+1)} = \left[\tilde{v}_1^{(k+1)}, \tilde{v}_2^{(k+1)}\right]^T$, $\boldsymbol{\varphi}^{(k)} = \left[v_1^{(k)}, v_2^{(k)}\right]^T$, and the two components of the vector $\boldsymbol{\varphi}^{(k+1)}$ are given by

$$v_1^{(k+1)} = \phi_1\left(\tilde{\boldsymbol{\varphi}}^{(k+1)}\right), \quad v_2^{(k+1)} = \phi_2\left(\tilde{\boldsymbol{\varphi}}^{(k+1)}\right). \tag{16}$$

Once the NR convergence condition is met, the two-port voltages are set to $[v_1, v_2]^T = \left[v_1^{(k)}, v_2^{(k)}\right]^T = \boldsymbol{\varphi}^{(k)}$ and the vector of waves $[b_1, b_2]^T$ reflected by the two-port nonlinear element is computed by means of (13).

3.2. Explicit WD Model based on Neural Network

Even though a vector definition of waves allows us to eliminate all the delay-free loops of WD structures containing up to a single nonlinear two-port element, in the previous subsection, we have shown that we still need a local iterative solver to compute the implicit vector scattering equation of the nonlinear WD block characterized by the extended EMM. In this subsection we aim at designing an explicit WD BJT model, thus obtaining a fully explicit WD structure with a single two-port nonlinearity. We propose a data-driven model whose WD mapping approximates the nonlinear scattering relation

$$\mathbf{b} = \mathbf{f}\left(\mathbf{a}\right), \tag{17}$$

which relates vectors of incident waves $\mathbf{a} = [a_1, a_2]^T$ to vectors of reflected waves $\mathbf{b} = [b_1, b_2]^T$ in a fully explicit fashion.

The two-port element behavior can be characterized through suitable simulation or measurement campaigns: sets of port voltages $[v_1, v_2]^T$ and port currents $[i_1, i_2]^T$ measures can be collected forming a Kirchhoff domain dataset for the BJT model. The vector definition of wave variables (4), where, again, $\mathbf{Z}_{1,2}$ is set in such a way to make reflection-free the pair of junction ports to which the nonlinear two-port is connected, allows us to transform the Kirchhoff domain dataset into a vector WD domain dataset of known solutions to the nonlinear vector function (17). The task of modeling the function $\mathbf{f}\left(\cdot\right)$ can be consequently seen as a regression problem, i.e.,

$$\hat{\mathbf{b}} = \tilde{\mathbf{f}}(\mathbf{a}; \theta_{\mathrm{WD}}) \tag{18}$$

where $\tilde{\mathbf{f}}$ is the function approximation given by a suitable neural network architecture [32, 33] whose parameters θ_{WD} are obtained by training the network to predict the current value of the reflected vector waves $\hat{\mathbf{b}}$ given the current value of the incident vector waves $\mathbf{a}$ as input.

Since the adopted reference model of the BJT is characterized by an instantaneous (static) nonlinearity, it is not necessary to rely on neural network structures with memory, such as Recurrent Neural Networks [34] or LSTM networks [35]. For this reason, in the next section, we employ a Multi-Layer Perceptron (MLP) network composed of 2 fully-connected layers with 16 hidden units each and Rectified Linear Unit (ReLU) activation functions (354 trainable parameters).

4. CASE STUDY

As a case study, we develop a Virtual Analog model of the common-emitter amplifier constituting the input stage of the Big Muff Pi distortion pedal - see Fig. 5. The circuit is characterized by a single nonlinear element, namely a 2N5089 BJT, whose extended EMM parameters are reported in Table 1.

The implemented WD structure is shown in Fig. 6 and it is characterized by series/parallel adaptors ($\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{P}_1$, $\mathcal{P}_2$, $\mathcal{P}_3$) designed according to [7] and a 6-port WD topological junction $\mathcal{R}_1$ implemented following [15], based on the vectorial definition of waves at the ports of the BJT. Indeed, the matrix of free parameters $\mathbf{Z}_{\mathcal{R}_1}$ of the junction $\mathcal{R}_1$ has a 2×2 submatrix $\mathbf{Z}_{1,2}$ positioned

Table 1: *2N5089 BJT: values of the extended EMM parameters.*

Parameter	Value	Description
V_{t}	25.85 mV	thermal voltage
α_{f}	0.9993	BJT forward current gain
α_{r}	0.5579	BJT reverse current gain
$R_{\mathrm{s1}}, R_{\mathrm{s2}}$	10^{-5} Ω	BJT series port resistance
$R_{\mathrm{p1}}, R_{\mathrm{p2}}$	10^{11}, 10^{8} Ω	BJT parallel port resistance
$\eta_{\mathrm{f}}, \eta_{\mathrm{r}}$	1	BJT ideality factor
I_{s1}	5.9151 fA	B-E junction saturation current
I_{s2}	10.595 fA	B-C junction saturation current

at the ports where the BJT is connected. All the linear elements at the leaves and all the upward facing ports of the junctions are adapted according to traditional WDF principles [7]. The nonlinear two-port element Q_1 is the root of the designed connection tree structure. The free parameters of the port resistance matrix $\mathbf{Z}_{1,2}$ are set to zero out the four entries of the first 2×2 block on the diagonal of the junction scattering matrix $\mathbf{S}_{\mathcal{R}_1}$. As previously mentioned in Section 2.3, this makes the pair of junction ports at which Q_1 is connected reflection-free.

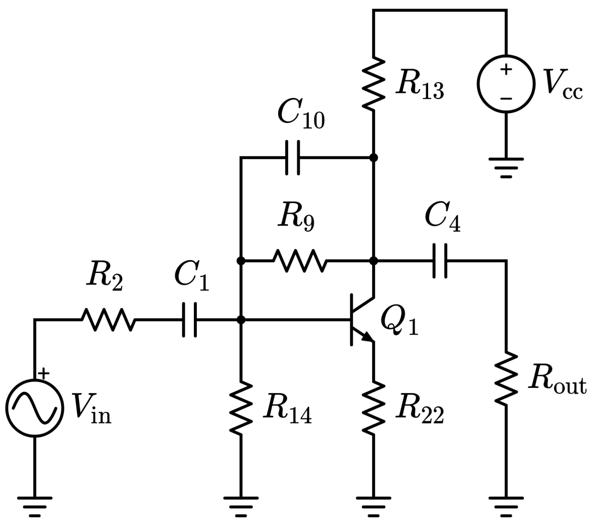

Figure 5: *Big Muff Pi input stage.*

4.1. Dataset and Model Training

In order to generate a suitable Kirchhoff domain dataset for the 2N5089 BJT, we implement the extended EMM using Mathworks Simscape and the model parameters reported in Table 1. The nonlinear characteristic of the model is sampled devising specific port voltage signals to reproduce common device operating ranges. In particular, the base-to-collector voltage v_2 is assumed to vary in the range $[-10, 0]$ V, while the base-to-emitter voltage v_1 varies in the range $[-0.8, 0.8]$ V. To sample the device, v_2 is set to 10^3 different equally spaced DC values extracted from its operating range and, for each of these values, v_1 is linearly increased for a duration of 0.1 seconds to cover its entire operating range; the corresponding base-to-collector current i_2 and base-to-emitter current i_1 are acquired with a sample rate $f_s = 96$ kHz. The

Table 2: *Values of the parameters of the Big Muff Pi input stage circuit shown in Fig. 5.*

R_2	R_9	R_{13}	R_{14}	R_{22}	R_{out}	C_1	C_4	C_{10}	V_{cc}
39 kΩ	470 kΩ	10 kΩ	47 kΩ	100 Ω	101 kΩ	1 µF	1 µF	470 pF	9 V

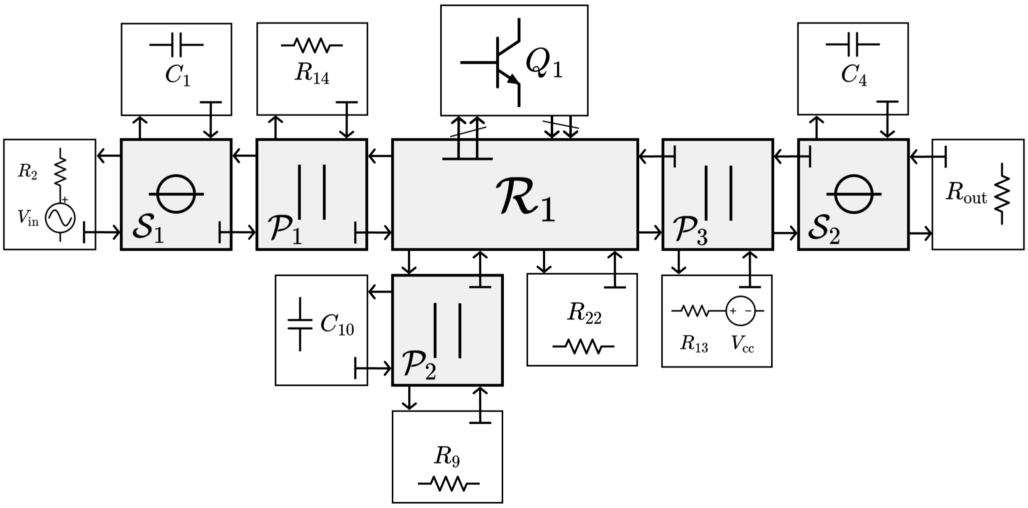

Figure 6: *WDF realization of the circuit shown in Fig. 5.*

complete simulation of the Kirchhoff domain signals v_1, i_1, v_2, i_2 amounts to 100 seconds.

The formed dataset can be expressed in the WD domain by using the vector wave transformation (4), where the free parameters Z_{11}, Z_{12}, Z_{21} and Z_{22} are set according to the already discussed port adaptation condition. From the total amount of data, the 80% are used for training, while the 20% are held out for evaluation purposes. The pairs of corresponding input and output wave variables $\left([a_1, a_2]^T, [b_1, b_2]^T\right)$ are assembled in batches containing 256 elements.

The MLP network described in Section 3.2 is implemented in Python using Pytorch [36] and it is trained for 500 epochs using Adam [37] to minimize the following loss function:

$$\mathcal{L} = \mathcal{E}\left(b_1, \hat{b}_1\right) + \mathcal{E}\left(b_2, \hat{b}_2\right) \tag{19}$$

where

$$\mathcal{E}\left(y, \hat{y}\right) = \frac{\sum_k \left(y_k - \hat{y}_k\right)^2}{\sum_k y_k^2} \tag{20}$$

is the Normalized Mean Squared Error (NMSE). Notably, the two terms summed in the loss function (19) are related to the two components of the output vector. To evaluate the model, we compute the model predictions over the evaluation set. The NMSE computed over those predictions is equal to 1.02×10^{-8}.

4.2. Results

In this subsection, we discuss the numerical results obtained from the simulation of the nonlinear WD structure shown in Fig. 6 using the two different WD BJT models presented in Sec. 3.1 and Sec. 3.2, respectively. Remarkably, thanks to the modularity of WDFs, it is possible to test both the WD BJT implementations employing the same WD structure and just substituting the nonlinear WD block at the root. In case the MNR method-based model is used, an iterative solver is needed to solve the two-port nonlinearity, while, in case the neural network-based model is used, the WD structure can be implemented in fully explicit fashion, since the nonlinearity is also expressed as an explicit wave mapping.

The circuit is tested with an input signal $V_{\text{in}} = A\sin(2\pi k f_0/f_s)$, where k is the sampling index, while $f_s = 96$ kHz and $f_0 = 1$ kHz are the sampling frequency and the fundamental frequency, respectively. According to the circuit analysis in [38], we set the input signal amplitude to $A = 0.7$ V to force the input stage into saturation, therefore causing asymmetric clipping. The input signal has a duration of one second. All the simulation algorithms of the WD structures are implemented as MATLAB scripts and are run on a laptop-mounted Intel Core i5-1240P 1.70 GHz CPU.

The simulation results related to the last 5 periods of the input signal are reported in Fig. 7(a) and Fig. 7(b). Both the WD implementations of the BJT closely match the same reference signal resulting from a Mathworks Simscape simulation of the circuit in Fig. 5, where the BJT has been modeled employing the extended EMM. The deviation between the reference Simscape signal and the neural network-based WD simulation is quantifiable by a NMSE of 2.32×10^{-5}, while the MNR method-based WD simulation is able to achieve an NMSE of 1.88×10^{-6}.

For measuring the computational cost associated to the two different WD implementations, $\Gamma = 1000$ identical simulations of the two WD structures are executed and the Real Time Ratio (RTR) is computed. The RTR is a dimensionless quantity indicating how fast the simulation is with respect to real time. If we consider Γ simulations having X samples as input and an execution time t_{c_i} for a single iteration over the input, the RTR can be computed as $\text{RTR} = (1/\Gamma)\sum_{i=1}^{\Gamma}\left(t_{c_i}/\left(X/f_s\right)\right)$, where f_s is the sampling frequency. The algorithm runs in real time if $\text{RTR} < 1$.

We obtain $\text{RTR} \simeq 1.65$ for the MNR method-based WD implementation and $\text{RTR} \simeq 0.21$ for the neural network-based WD implementation. While the MNR method guarantees arbitrarily

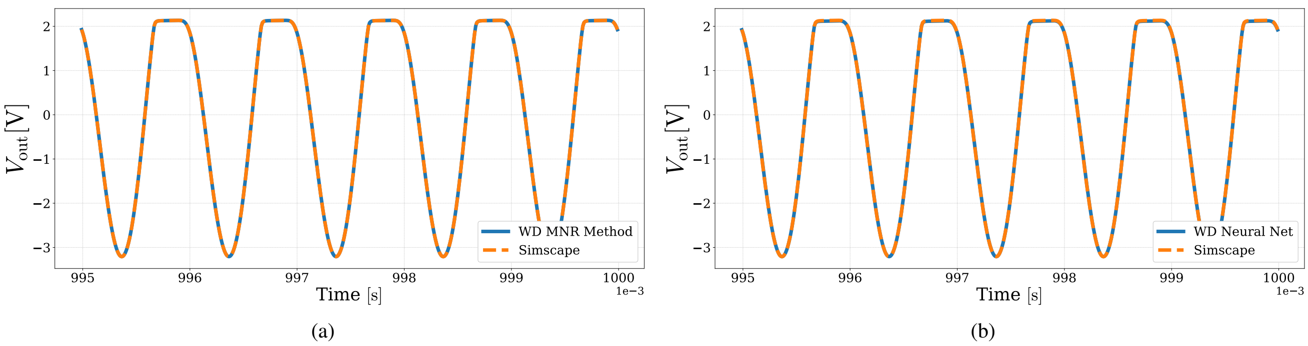

Figure 7: *Voltage V_{out} measured across the resistor R_{out}. (a) Comparison between the MNR method-based BJT WD implementation and Mathworks Simscape. (b) Comparison between the neural network-based BJT WD implementation and Mathworks Simscape.*

good results in terms of accuracy, the iterative solver considerably increases the algorithm execution time. However, with almost comparable accuracy results, the WD structure containing the explicit neural network-based WD BJT model proves to be far more efficient as far as execution time is concerned, being able to run in real time in the MATLAB environment.

5. CONCLUSIONS

In this paper, we have shown that, by adopting a vector definition of waves, it is possible to design WDFs which contain a single two-port nonlinearity connected to a topological junction through a pair of ports with no delay-free loops. It follows that, when the two-port nonlinearity is characterized by an explicit WD mapping, the resulting WDF is fully explicit. The proposed approach allows us to preserve the desirable modularity property of traditional scalar WDFs, according to which circuit elements and connection networks are modeled with separate input-output blocks.

We proposed two vector wave-based models of BJT transistors described with an extended EMM. The first one is implicit, and it relies on a MNR method to solve its nonlinear scattering relation. The second is fully explicit, as it relies on a neural network model which approximates the explicit scattering equation relating the input vector of incident waves to the corresponding vector of reflected waves. We finally provided an accuracy and performance comparison of the two WD BJT models with reference to a specific VA application scenario. The fully explicit WD structure including as root a neural network-based model of a two-port BJT proved to be an excellent compromise between accuracy and computational cost. This shows that the integration of efficient data-driven blocks of nonlinear devices into the WDFs framework might lead the way towards the real time implementation of increasingly complex nonlinear circuits.

Future work might concern the application of vector WDFs for the modeling of generic N-port nonlinear elements. A noteworthy extension would also be considering data-driven methods for the characterization of multi-port nonlinearities with memory, leveraging models or experimental measurements of transistors or vacuum tubes. Finally, the proposed strategy should be validated from a broader perspective, comparing it to other approaches developed in the literature.

6. REFERENCES

[1] V. Välimäki, S. Bilbao, J. O. Smith, J. S. Abel, J. Pakarinen, and D. Berners, *Virtual analog effects*, pp. 473–522, John Wiley & Sons, United Kingdom, 2nd edition, 2011.

[2] D. Bouvier, T. Hélie, and D. Roze, “Phase-based order separation for Volterra series identification,” *International Journal of Control*, vol. 94, no. 8, pp. 2104–2114, 2021.

[3] A. Wright, E.P. Damskägg, V. Välimäki, et al., “Real-time black-box modelling with recurrent neural networks,” in *Proc. 22nd Int. Conf. Digital Audio Effects (DAFx-19)*, Birmingham, UK, 2019, University of Birmingham.

[4] G. Borin, G. De Poli, and D. Rocchesso, “Elimination of delay-free loops in discrete-time models of nonlinear acoustic systems,” *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 5, pp. 597–605, 2000.

[5] A. Falaize-Skrzek and T. Hélie, “Simulation of an analog circuit of a wah pedal: a port-Hamiltonian approach,” in *Proc. 135th Audio Engineering Society Convention*, New York, USA, 2013, Audio Engineering Society.

[6] G. De Sanctis and A. Sarti, “Virtual analog modeling in the wave-digital domain,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 4, pp. 715–727, 2010.

[7] A. Fettweis, “Wave digital filters: Theory and practice,” *Proceedings of the IEEE*, vol. 74, no. 2, pp. 270–327, 1986.

[8] A. Bernardini, P. Maffezzoni, and A. Sarti, “Linear multistep discretization methods with variable step-size in nonlinear wave digital structures for virtual analog modeling,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 11, pp. 1763–1776, 2019.

[9] K. Meerkotter and R. Scholz, “Digital simulation of nonlinear circuits by wave digital filter principles,” in *1989 IEEE International Symposium on Circuits and Systems (ISCAS)*, Portland, USA, 1989, vol. 1, pp. 720–723, IEEE.

[10] A. Sarti and G. De Sanctis, “Systematic methods for the implementation of nonlinear wave-digital structures,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 56, no. 2, pp. 460–472, 2009.

[11] T. Felderhoff, “A new wave description for nonlinear elements,” in *Proc. IEEE Int. Symp. Circuits Syst.* Sep, 1996, vol. 3, pp. 221–224, IEEE.

[12] A. Sarti and G. De Poli, "Toward nonlinear wave digital filters," *IEEE Transactions on Signal Processing*, vol. 47, no. 6, pp. 1654–1668, 1999.

[13] H. Carlin, "Singular network elements," *IEEE Transactions on Circuit Theory*, vol. 11, no. 1, pp. 67–72, 1964.

[14] A. Bernardini and A. Sarti, "Biparametric wave digital filters," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. PP, pp. 1–13, 03 2017.

[15] A. Bernardini, K. J. Werner, J. O. Smith, III, and A. Sarti, "Generalized wave digital filter realizations of arbitrary reciprocal connection networks," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 66, no. 2, pp. 694–707, 2019.

[16] K. J. Werner, V. Nangia, J. O. Smith III, and J. S. Abel, "Resolving wave digital filters with multiple/multiport nonlinearities," in *Proc. 18th Int. Conf. Digital Audio Effects (DAFx-15)*, Trondheim, Norway, 2015, pp. 387–394, Norwegian University of Science and Technology.

[17] D. T. Yeh, J. S. Abel, and J. O. Smith, "Automated physical modeling of nonlinear audio circuits for real-time audio effects—part I: Theoretical development," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 4, pp. 728–737, 2010.

[18] D. T. Yeh, "Automated physical modeling of nonlinear audio circuits for real-time audio effects—part II: Bjt and vacuum tube examples," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 4, pp. 1207–1216, 2012.

[19] M. J. Olsen, K. J. Werner, and J. O. Smith III, "Resolving grouped nonlinearities in wave digital filters using iterative techniques," in *Proc. 19th Int. Conf. Digital Audio Effects (DAFX-16)*, Brno, Czech Republic, 2016, pp. 279–286, Brno University of Technology.

[20] A. Bernardini, A. E. Vergani, and A. Sarti, "Wave digital modeling of nonlinear 3-terminal devices for virtual analog applications," *Circuits, Systems, and Signal Processing*, vol. 39, no. 7, pp. 3289–3319, 2020.

[21] L. Kolonko, B. Musiol, J. Velten, and A. Kummert, "A split-modular approach to wave digital filters containing bipolar junction transistors," in *2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)*, Lansing, 2021, pp. 840–843, IEEE.

[22] A. Bernardini, P. Maffezzoni, and A. Sarti, "Vector wave digital filters and their application to circuits with two-port elements," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 68, no. 3, pp. 1269–1282, 2021.

[23] J. J. Ebers and J. L. Moll, "Large-signal behavior of junction transistors," *Proceedings of the IRE*, vol. 42, no. 12, pp. 1761–1772, 1954.

[24] J. Chowdhury and C. J. Clarke, "Emulating diode circuits with differentiable wave digital filters," in *19th Sound and Music Computing Conference*, Saint-Étienne, France, 2022, pp. 2–9, SMC Network.

[25] C. C. Darabundit, D. Roosenburg, and J. O. Smith, "Neural net tube models for wave digital filters," in *Proc. 25th Int. Conf. Digital Audio Effects (DAFx20in22)*, Vienna, Austria, 2022, pp. 153–160, Vienna University of Music and Performing Arts.

[26] O. Massi, A. I. Mezza, R. Giampiccolo, and A. Bernardini, "Deep learning-based wave digital modeling of rate-dependent hysteretic nonlinearities for virtual analog applications," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2023, no. 1, pp. 12, 2023.

[27] K. J. Werner, A. Bernardini, J. O. Smith, and A. Sarti, "Modeling circuits with arbitrary topologies and active linear multiports using wave digital filters," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 65, no. 12, pp. 4233–4246, 2018.

[28] M. G. Villalva, J. R. Gazoli, and E. R. Filho, "Comprehensive approach to modeling and simulation of photovoltaic arrays," *IEEE Transactions on Power Electronics*, vol. 24, no. 5, pp. 1198–1208, 2009.

[29] A. Bernardini, K.J Werner, P. Maffezzoni, and A. Sarti, "Wave digital modeling of the diode-based ring modulator," in *Proc. 144th Audio Engineering Society Convention*, Milan, Italy, 2018, Audio Engineering Society.

[30] A. Bernardini, K. J. Werner, A. Sarti, and J. O. Smith, III, "Modeling nonlinear wave digital elements using the Lambert function," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 63, no. 8, pp. 1231–1242, 2016.

[31] A. Bernardini, E. Bozzo, F. Fontana, and A. Sarti, "A wave digital Newton-Raphson method for virtual analog modeling of audio circuits with multiple one-port nonlinearities," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 2162–2173, 2021.

[32] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural networks*, vol. 4, no. 2, pp. 251–257, 1991.

[33] C. Huang, "Relu networks are universal approximators via piecewise linear or constant functions," *Neural Computation*, vol. 32, no. 11, pp. 2249–2278, 2020.

[34] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.

[35] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, pp. 1735–80, 12 1997.

[36] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al., "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.

[37] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[38] Electrosmash, "Big Muff Pi analysis," Available at https://www.electrosmash.com/big-muff-pi-analysis, accessed March 31, 2023.