# TRAINING NEURAL MODELS OF NONLINEAR MULTI-PORT ELEMENTS WITHIN WAVE DIGITAL STRUCTURES THROUGH DISCRETE-TIME SIMULATION

*Oliviero Massi, Alessandro Ilic Mezza, Riccardo Giampiccolo, and Alberto Bernardini*

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

## ABSTRACT

Neural networks have been applied within the Wave Digital Filter (WDF) framework as data-driven models for nonlinear multi-port circuit elements. Conventionally, these models are trained on wave variables obtained by sampling the current-voltage characteristic of the considered nonlinear element before being incorporated into the circuit WDF implementation. However, isolating multi-port elements for this process can be challenging, as their nonlinear behavior often depends on dynamic effects that emerge from interactions with the surrounding circuit. In this paper, we propose a novel approach for training neural models of nonlinear multi-port elements directly within a circuit's Wave Digital (WD) discrete-time implementation, relying solely on circuit input-output voltage measurements. Exploiting the differentiability of WD simulations, we embed the neural network into the simulation process and optimize its parameters using gradient-based methods by minimizing a loss function defined over the circuit output voltage. Experimental results demonstrate the effectiveness of the proposed approach in accurately capturing the nonlinear circuit behavior, while preserving the interpretability and modularity of WDFs.

## 1. INTRODUCTION

Data-driven methods have significantly impacted various fields of digital signal processing, driving a gradual convergence between model-based techniques and machine learning approaches [1]. In audio signal processing, Digital Audio Effects (DAFx) have followed a similar trend, particularly in the domain of Virtual Analog (VA) effects, which aim to recreate the behavior of analog audio gear in the digital domain [2]. A fundamental challenge in designing VA effects is the development of efficient techniques that enable real-time emulation of analog audio circuits while preserving the distinctive timbral characteristics introduced by the nonlinear behavior of their constitutive analog components.

Traditionally, VA methods can be categorized into two main approaches: *closed-box* and *glass-box*, which correspond to the earlier terms *black-box* and *white-box*, respectively. Closed-box methods focus on learning the input-output behavior of a circuit using parametric models without explicitly considering its internal structure. Examples of such methods are nonlinear system identification approaches based on Volterra series [3] and, more generally, neural networks [4]. On the other hand, glass-box methods entail solving in the discrete-time domain the algebraic or the Ordinary Differential Equations (ODEs) characterizing the schematic, typically using state-space methods [5], port-Hamiltonian methods [6], or Wave Digital Filters (WDFs) [7]. A key characteristic of glass-box models is their reliance on full knowledge of the circuit under study, a requirement that can be mitigated when experimental sampling of the considered circuit is available.

While a portion of the research in VA modeling has focused on employing ever-evolving deep learning architectures to produce closed-box models of analog audio systems, such as vacuum-tube amplifiers [8], compressor circuits [9, 10], and modulation effects [11], other research lines have explored hybrid approaches combining glass-box methods with ideas from the machine learning domain. Closed-box models based on deep neural networks can be flexible in learning highly complex behaviors directly from data, but they often require large amounts of training data and may struggle with generalization beyond the training set. In contrast, glass-box models inherently incorporate domain knowledge by enforcing physical constraints, such as circuit topology and governing equations. These *inductive biases* [12] not only improve data efficiency, but also enhance model interpretability, and generalization to unseen conditions. In particular, inspired by prior work on Differentiable Digital Signal Processing (DDSP) [13], several research efforts have been devoted to defining differentiable glass-box DSP structures that can be optimized using gradient-based methods to fit physical input-output measurements from a real device [14, 15].

Among glass-box approaches, Wave Digital Filters (WDFs) have emerged as a promising framework for realizing differentiable DSP structures [16, 17]. Initially developed by A. Fettweis in the 1970s for the digital implementation of passive reference analog filters [18], WDFs rely on a port-wise transformation of port voltages and port currents (the so-called Kirchhoff variables) into incident and reflected waves (i.e., wave variables) with the introduction of a free parameter per port called *port resistance*. This domain change allows modeling connection networks and circuit elements separately, as input-output scattering blocks, achieving advantages like high modularity and potentially explicit realizations when stable discretization methods are employed [18, 19]. When interconnecting these blocks, the introduced free parameters are set through the so-called *adaptation* process in order to eliminate implicit relations, avoiding thus Delay-Free Loops (DFLs). Circuits with up to one nonlinear one-port element, described by an explicit mapping, can be realized in the Wave Digital (WD) domain in a fully explicit fashion, i.e., without the need of any iterative solver [18, 20]. Recently, thanks to the use of a vector definition of wave variables [21], the class of circuits that can be

emulated with WDFs in a fully explicit fashion has been extended to accommodate even those circuits featuring one nonlinear multi-port element, provided that an explicit formulation of the nonlinear WD mapping can be obtained [22].

The WD discrete-time simulation within an Automatic Differentiation (AD)-based framework has been used to perform Lumped Element Model (LEM) parameter values estimation via gradient-based methods [23], enabling the optimization of equivalent circuit models to match experimental observations. Recent advancements have integrated neural network-based blocks [22, 24, 25, 26, 27] among traditional WD blocks in order to implement the explicit scattering relation of one-port or multi-port nonlinear elements. Such neural models are pre-trained over wave variables obtained by first sampling the Kirchhoff domain characteristic of the nonlinear element and then integrated into the circuit WDF realization. However, determining exact nonlinear component characteristic curves requires them to be measured in isolation from the circuit, which is not only impractical but can also compromise the integrity of the circuit [28]. Additionally, other nonlinear effects might arise from the interaction of the nonlinear element with the rest of the circuit, which would not be captured by the measurement in isolated conditions.

In this manuscript, we introduce a novel approach for training neural models of nonlinear multi-port elements directly within a circuit's WD discrete-time implementation, relying solely on circuit input-output voltage measurements. Exploiting the differentiability of WD simulations, we embed a randomly-initialized neural network into the simulation process and optimize its parameters using gradient-based methods by minimizing a loss function defined over the circuit output voltage. The modular structure of WDFs allows us to encapsulate the circuit nonlinear behavior within the scattering relation of the multi-port element, which we implement explicitly using a neural model. This approach preserves the interpretability and structure of the WDF framework while enabling accurate modeling of complex nonlinear behaviors directly from experimental measurements.

The remainder of the manuscript is organized as follows. Section 2 provides a background concerning WDFs and their extension to a vector definition. Section 3 then addresses the problem of training a neural model within the discrete-time WD simulation. A case study for the application of the developed methodology is presented in Section 4, along with numerical results. Finally, Section 5 concludes this manuscript.

## 2. VECTOR WAVE DIGITAL FILTERS

In this section, we provide the necessary background on Vector Wave Digital Filters (VWDFs), which are WD structures that combine both scalar and vector definitions of wave variables.

### 2.1. Scalar Waves

The design of WDFs relies on a port-wise description of a reference analog circuit. Each pair of Kirchhoff port variables $v_j$, and $i_j$, is transformed into a corresponding pair of voltage wave variables, defined as [18]

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

where $v_j$ and $i_j$ represent the $j$-th port voltage and port current, while $a_j$ and $b_j$ correspond to the $j$-th port incident and reflected waves, respectively. $Z_j \neq 0$ is a real-valued free parameter, referred to as *reference one-port resistance*, which plays a crucial role in solving WD structures [18]. For linear one-port elements, an appropriate choice of $Z_j$ can eliminate the instantaneous dependence of $b_j$ on $a_j$. Specifically, the scattering equation for a general linear one-port element, including dynamic components, in the WD domain is given by [19]

$$b_j[k] = \frac{R_{\mathrm{g},j}[k] - Z_j[k]}{R_{\mathrm{g},j}[k] + Z_j[k]} a_j[k] + \frac{2Z_j[k]}{R_{\mathrm{g},j}[k] + Z_j[k]} V_{\mathrm{g},j}[k]\,. \tag{2}$$

This formulation is obtained by applying (1) to the Thévenin equivalent model of a one-port circuit element in the discrete-time domain

$$v_j[k] = R_{\mathrm{g},j}[k] i_j[k] + V_{\mathrm{g},j}[k]\,, \tag{3}$$

where $k$ represents the sample index, while $R_{\mathrm{g},j}[k]$ and $V_{\mathrm{g},j}[k]$ denote the resistive and voltage parameters, respectively. Setting $Z_j[k] = R_{\mathrm{g},j}[k]$, the instantaneous dependence between $b_j[k]$ and $a_j[k]$ is eliminated [19]. Under this condition, the scattering equation (2) simplifies to $b_j[k] = V_{\mathrm{g},j}[k]$, and the linear one-port element is said to be *adapted* according to WDF theory [18].

### 2.2. Vector Waves

Using the scalar definition of wave variables in (1) to model multi-port elements in the WD domain often leads to unavoidable DFLs which cannot be eliminated through any choice of the free parameters [21, 22]. A possible strategy to address this limitation is to encompass the $N$ ports of a given $N$-port circuit element into vector wave variables, defined as [21]

$$\mathbf{a} = \mathbf{v} + \mathbf{Z}_{1:N}\mathbf{i}\,, \quad \mathbf{b} = \mathbf{v} - \mathbf{Z}_{1:N}\mathbf{i}\,, \tag{4}$$

where $\mathbf{v} = [v_1, \ldots, v_N]^{\mathrm{T}}$ and $\mathbf{i} = [i_1, \ldots, i_N]^{\mathrm{T}}$ are the vectors of the $N$-port voltages and currents, respectively. Similarly, $\mathbf{a} = [a_1, \ldots, a_N]^{\mathrm{T}}$ and $\mathbf{b} = [b_1, \ldots, b_N]^{\mathrm{T}}$ are the vector of waves incident to and reflected from the $N$-port element, respectively, whereas

$$\mathbf{Z}_{1:N} = \begin{bmatrix} Z_{11} & \ldots & Z_{1N} \\ \vdots & \ddots & \vdots \\ Z_{N1} & \ldots & Z_{NN} \end{bmatrix}, \tag{5}$$

is a full-rank $N \times N$ matrix of real-valued free parameters $Z_{\imath\jmath}$, with $\imath, \jmath \in \{1, \ldots, N\}$, referred to as *reference multi-port resistance*.

Similar to the scalar case, an $N$-port linear element can be adapted by appropriately selecting the $\mathbf{Z}_{1:N}$ entries to remove the instantaneous dependence between the vectors of reflected waves and incident waves in the scattering relation [21].

### 2.3. Connection Networks

Topological connection networks are represented as $J$-port WD blocks characterized by the vector $\mathbf{b}_{\mathrm{J}} = [b_1, \ldots, b_J]^{\mathrm{T}}$, which contains the waves incident to the junction and reflected by the elements, and the vector $\mathbf{a}_{\mathrm{J}} = [a_1, \ldots, a_J]^{\mathrm{T}}$, which includes waves reflected by the junction and incident to the elements. The scattering relation between $\mathbf{b}_{\mathrm{J}}$ and $\mathbf{a}_{\mathrm{J}}$ for a $J$-port junction is given by

$$\mathbf{a}_{\mathrm{J}} = \mathbf{S}\mathbf{b}_{\mathrm{J}}\,, \tag{6}$$

where $\mathbf{S}$ is a $J \times J$ scattering matrix which can be computed for reciprocal lossless connection networks using either of the following equivalent expressions [29]

$$\mathbf{S} = 2\mathbf{Q}^{\mathrm{T}}(\mathbf{Q}\mathbf{Z}_{\mathrm{J}}^{-1}\mathbf{Q}^{\mathrm{T}})^{-1}\mathbf{Q}\mathbf{Z}_{\mathrm{J}}^{-1} - \mathbf{I}\,, \tag{7}$$

$$\mathbf{S} = \mathbf{I} - 2\mathbf{Z}_{\mathrm{J}}\mathbf{B}^{\mathrm{T}}(\mathbf{B}\mathbf{Z}_{\mathrm{J}}\mathbf{B}^{\mathrm{T}})^{-1}\mathbf{B}\,, \tag{8}$$

where $\mathbf{I}$ is the $J \times J$ identity matrix, while $\mathbf{Q}$ and $\mathbf{B}$ are the fundamental cut-set and loop matrices, respectively. In VWDFs, which incorporate both $N$-port elements described by vector waves and one-port elements described by scalar waves, the matrix $\mathbf{Z}_{\mathrm{J}}$ is a full-rank block diagonal matrix, which features both reference multi-port and one-port resistances on the main diagonal [21].

### 2.4. VWDFs with a Single *N*-port Nonlinear Element

In the previous subsections, we discussed how reference one-port and multi-port resistances can be set to adapt circuit elements, ensuring that the WD structure remains computable. This is, in general, only applicable to linear elements [18, 21]. However, for circuits containing a single nonlinear one-port or $N$-port element, it has been shown that it is possible to exploit free parameters in order to remove DFLs by adapting the port (or ports) of the junction to which the nonlinear element is connected [19, 22]. Such circuits can be therefore modeled using WDFs with a connection tree structure composed of:

- *Root*, i.e., the nonlinear element;
- *Nodes*, i.e., the topological junctions, which have one adapted (scalar or vectorial) port and multiple non-adapted (scalar or vectorial) ports;
- *Leaves*, i.e., the linear elements, which have just one adapted (scalar or vectorial) port.

Such tree-like WD structures are fully-explicit, as long as the one-port or $N$-port nonlinear block is characterized by an explicit mapping. Adapted ports ensure that no implicit relations form at the interconnections among WD blocks. The only sufficient condition for applying such modeling approach is that there exists a reference multi-port resistance that adapts the ports of the junction to which the $N$-port nonlinear block is connected [26].

Assume, without loss of generality, that the entire connection network of a reference circuit, featuring a single $N$-port nonlinear element, can be modeled using a single $J$-port junction (single node). This junction is characterized, at each sampling index $k$, by the vectors of wave variables $\mathbf{b}_{\mathrm{J}}[k] = [b_1[k], \ldots, b_J[k]]^{\mathrm{T}}$ and $\mathbf{a}_{\mathrm{J}}[k] = [a_1[k], \ldots, a_J[k]]^{\mathrm{T}}$. The $N$-port nonlinear element, described by means of vector waves, is connected to ports with indexes from 1 to $N$, while the linear elements (leaves) are assumed to be connected to the remaining $J - N$ ports. The discrete-time simulation of the corresponding VWDF structure comprises the following stages at each sampling index $k$:

- **Leaves Scattering Stage**: the waves reflected by the leaves, corresponding to the last $J - N$ entries of $\mathbf{b}_{\mathrm{J}}[k]$, are computed using their adapted scattering relation, which in the case of one-port elements can be written as

$$b_j[k] = V_{\mathrm{g},j}[k] \quad j \in \{N+1, \ldots, J\}\,. \tag{9}$$

- **Forward Scattering Stage**: the vector of waves incident to the $N$-port root is computed as

$$\mathbf{a}[k] = \mathbf{S}_{1,N}\mathbf{b}_{\mathrm{J}}[k]\,, \tag{10}$$

where $\mathbf{S}_{1,N}$ is the sub-matrix formed by the first $N$ rows of the scattering matrix $\mathbf{S}$.

- **Root Scattering Stage**: the vector of waves reflected by the root is computed according to the vector scattering relation $\mathbf{f}$ characterizing the nonlinear element

$$\mathbf{b}[k] = \mathbf{f}(\mathbf{a}[k])\,, \tag{11}$$

where $\mathbf{b}[k]$ constitutes the first $N$ entries of $\mathbf{b}_{\mathrm{J}}[k]$.

- **Backward Scattering Stage**: the vector of waves reflected by the node (incident to the leaves) is computed as

$$\mathbf{a}_{\mathrm{J}}[k] = \mathbf{S}\mathbf{b}_{\mathrm{J}}[k]\,. \tag{12}$$

At each sampling index $k$, the values of the Kirchhoff variables at each scalar or vectorial port can be obtained by applying the inverse linear transformation of (1) and (4), respectively.

### 2.5. Explicit WD *N*-port Nonlinear Element Modeling

The VWDF discrete-time simulation of reference circuits containing a single $N$-port nonlinear element is fully explicit, provided that the Root Scattering Stage, expressed in (11), can also be solved explicitly. In [22, 26], it is demonstrated that a parametric nonlinear model, specifically a neural network, can be employed to regress the explicit nonlinear WD mapping for an $N$-port nonlinear element. The general form of this explicit WD mapping is given by

$$\hat{\mathbf{b}}[k] = \mathbf{f}_{\boldsymbol{\theta}}(\mathbf{a}[k])\,, \tag{13}$$

where $\mathbf{f}_{\boldsymbol{\theta}}$ is the neural approximation of function $\mathbf{f}$, whose parameters $\boldsymbol{\theta}$ are obtained training the model to estimate the vector of reflected waves $\hat{\mathbf{b}}$ given the corresponding vector of incident waves $\mathbf{a}$ as input. These pairs of vectors of wave variables are derived from Kirchhoff domain samples of voltages $[v_1, \ldots, v_N]$ and currents $[i_1, \ldots, i_N]$ of the nonlinear element characteristic. Each entry is then transformed using the vector definition of wave variables in (4), where $\mathbf{Z}_{1:N}$ is chosen to ensure adaptation at the ports to which the $N$-port WD block is connected.

## 3. TRAINING NONLINEAR ELEMENT MODELS THROUGH VWDF SIMULATION

By using (13) in place of (11), the computational flow for the VWDF discrete-time simulation of a reference circuit with a single $N$-port nonlinear element can be defined at each sampling index $k$ by the input-output relation

$$\hat{y}[k] = \varphi(x[k]; \boldsymbol{\theta})\,, \tag{14}$$

where $x[k]$ is the $k$-th sample of the input voltage, and $\hat{y}[k]$ is the estimate of output voltage $y[k]$. The circuit output is determined by the set of parameters $\boldsymbol{\theta}$, which, in this case, are associated with the representation $\mathbf{f}_{\boldsymbol{\theta}}$ of the $N$-port nonlinear element. By implementing $\varphi$ in Python using the PyTorch Autograd engine [30], each operation involved in the discrete-time simulation process is organized within a Directed Acyclic Graph (DAG). This structure enables the efficient computation of gradients of any loss function with respect to the model parameters using reverse-mode AD [23]. We define the training objective as finding $\boldsymbol{\theta}^{\star}$ such that

$$\boldsymbol{\theta}^{\star} = \arg\min_{\boldsymbol{\theta}} \mathcal{L}(\mathbf{y}, \hat{\mathbf{y}})\,, \tag{15}$$

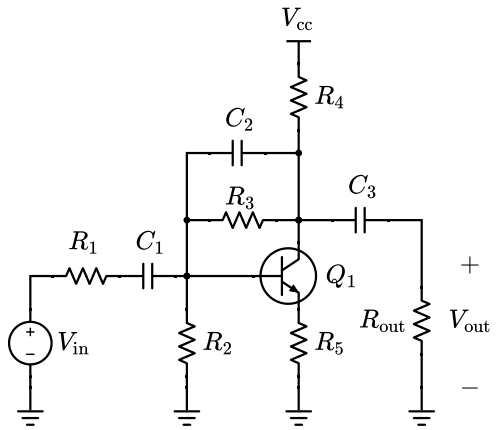


Figure 1: *Big Muff Pi input stage.*

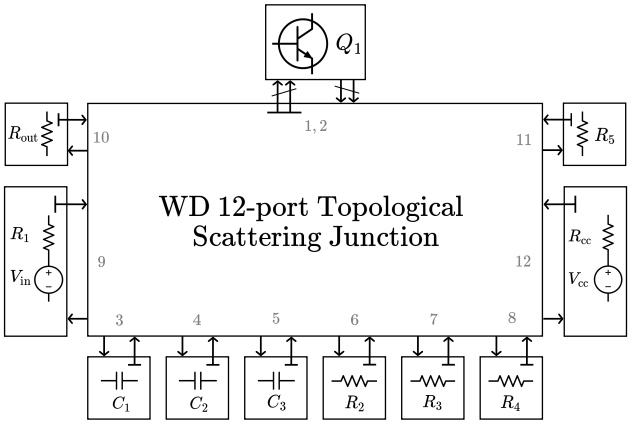


Figure 2: *VWDF implementation of the circuit shown in Figure 1.*

where $\mathbf{y} = [y[k], \dots, y[k-K+1]]^{\mathrm{T}}$ represents segments of the ground-truth circuit voltage output windowed using a rectangular window of size $K$, and $\hat{\mathbf{y}} = [\hat{y}[k], \dots, \hat{y}[k-K+1]]^{\mathrm{T}}$ contains the model predictions for a given input sequence $\mathbf{x} = [x[k], \dots, x[k-K+1]]^{\mathrm{T}}$. Each input sequence $\mathbf{x}$ is pre-padded with zeros to ensure the correct initialization of dynamic elements and allow the circuit to reach a steady-state condition. During the processing of these zero-padding steps, which we define as *warm up*, the PyTorch Autograd engine is disabled to prevent unnecessary gradient computations, and the corresponding outputs are subsequently discarded.

The loss function in (15) is defined as the weighted sum of three terms: the time-domain Normalized Mean Squared Error (NMSE) (17), spectral convergence (18), and the $L^1$-norm of the regularized log-magnitude error (19). Specifically, it is formulated as [9, 31]

$$\mathcal{L} = \mathcal{L}_{\text{td}} + \lambda \left( \mathcal{L}_{\text{sc}} + \mathcal{L}_{\log} \right) , \tag{16}$$

where

$$\mathcal{L}_{\text{td}} = \frac{\|\mathbf{y} - \hat{\mathbf{y}}\|_2^2}{\|\mathbf{y}\|_2^2} , \tag{17}$$

$$\mathcal{L}_{\text{sc}} = \frac{\| \, |\text{STFT}(\mathbf{y})| - |\text{STFT}(\hat{\mathbf{y}})| \, \|_F}{\| \, |\text{STFT}(\mathbf{y})| \, \|_F} , \tag{18}$$

$$\mathcal{L}_{\log} = \frac{1}{FT} \left\| \log\left( \frac{|\text{STFT}(\mathbf{y})| + \epsilon}{|\text{STFT}(\hat{\mathbf{y}})| + \epsilon} \right) \right\|_1 . \tag{19}$$

Here, $|\text{STFT}(\,\cdot\,)| \in \mathbb{R}^{F \times T}$ represents the magnitude of the Short-Time Fourier Transform (STFT), $\|\cdot\|_1$ denotes the $L^1$-norm, $\|\cdot\|_2$ the $L^2$-norm, and $\|\cdot\|_F$ the Frobenius norm. The term $\epsilon$ is a small constant added to prevent numerical errors, and $\lambda = 10^{-3}$ is the spectral weighting factor.

Solving the optimization problem in (15) through the VWDF discrete-time simulation allows to characterize the nonlinear behavior of the $N$-port nonlinear element directly from input-output circuit voltage measurements. It is worth noting that implementing (13) does not necessarily require the use of neural networks, as any differentiable parametric nonlinear model can be used to address this task. Finally, depending on the specific nonlinear device being modeled within the circuit, either static or dynamic models can be employed to characterize its peculiar nonlinear behavior.

## 4. CASE STUDY

As an example of application, let us consider the input stage of the Big Muff Pi circuit[1] depicted in Figure 1. This circuit includes a single npn 2N5089 Bipolar Junction Transistor (BJT). Following the approach in [22], the transistor can be represented using its two-port element model. The corresponding VWDF connection-tree structure, illustrated in Figure 2, consists of a single WD 12-port junction (node) to which all one-port linear elements are connected as leaves. The BJT, acting as the root of the structure, is connected to the junction with $N = 2$ vectorial ports. The input voltage signal is represented by the voltage source $V_{\text{in}}$, while the output voltage $V_{\text{out}}$ is measured across the terminals of resistor $R_{\text{out}}$. The circuit parameter values are summarized in Table 1.

### 4.1. Dataset and Training

The reference circuit is implemented as a Mathworks Simscape model, with the BJT represented using the extended Ebers-Moll model (EMM) [22]. The parameters for the 2N5089 BJT used in this circuit are listed in Table 2.

To construct a synthetic dataset, we start from a 30-second clean guitar audio excerpt. The input voltage signal $V_{\text{in}}[k]$ is built by applying a multiplicative factor $A \in \{0.5, 1.0, 1.5, 2.0\}$ to the original signal. We add also five 2-second exponential sweeps from 20 Hz to 8 kHz, with amplitudes from 0.1 V to 0.5 V. The total duration of the input voltage signal amounts to 130 s. We obtain the corresponding $V_{\text{out}}[k]$ using the Simscape model, ensuring that it is simulated starting from a steady-state condition. All signals are sampled at $f_{\text{s}} = 44.1$ kHz. Input sequences $\mathbf{x} = [V_{\text{in}}[k], \dots, V_{\text{in}}[k-K+1]]^{\mathrm{T}}$ are obtained by shifting a rectangular window of length $K = 8192$ and hop size 512, and they are assigned with the corresponding ground truth sequences $\mathbf{y} = [V_{\text{out}}[k], \dots, V_{\text{out}}[k-K+1]]^{\mathrm{T}}$. Such sequences are assembled in batches of 64 elements. From the entire dataset, 90% of the entries are used for training, while the remaining 10% is held out for validation purposes.

The nonlinear root scattering $\mathbf{f}$ in the VWDF is implemented using a Multi-Layer Perceptron (MLP) with two hidden layers of 32 units each (1218 parameters) and Exponential Linear Unit

[1]The schematic is taken from: `https://www.electrosmash.com/big-muff-pi-analysis`

Table 1: *Parameter values for the Big Muff Pi input stage circuit shown in Figure 1.*

| $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ | $R_{\text{out}}$ | $C_1$ | $C_2$ | $C_3$ | $V_{\text{cc}}$ |
|---|---|---|---|---|---|---|---|---|---|
| 39 kΩ | 47 kΩ | 470 kΩ | 10 kΩ | 100 Ω | 1 kΩ | 1 µF | 470 pF | 1 µF | 9 V |

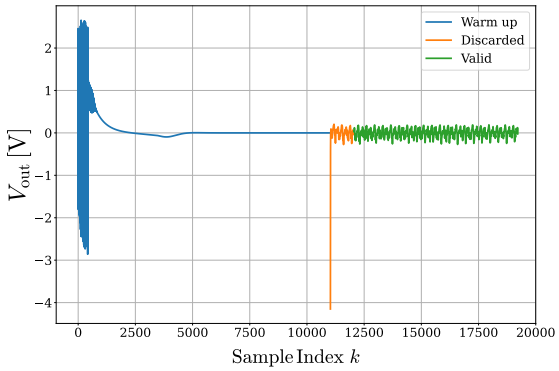


Figure 3: *Example of VWDF discrete-time simulation associated to an output sequence* $\hat{\mathbf{y}}$*. The first* 11,025 *samples (in blue), obtained with a zero input, are disregarded to prevent the initial transient affecting the learning process. The next* 1024 *samples (in orange), corresponding to the onset of the input signal, are also discarded to account for the relative transient. The remaining* 7168 *samples (in green) are considered valid and are used in the computation of the loss function defined in* (16).

(ELU) activation functions. We train the VWDF connection-tree modified to handle batch processing for 15 epochs using the Adam optimizer [32] with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and a learning rate of $10^{-3}$ to minimize the loss function in (16). Each input sequence is preceded by a warm up phase consisting of 11,025 zero samples (0.25 s). To ensure the loss is not influenced by the initial transient behavior of the VWDF following the warm up phase, we compute it using only the last 7168 samples (0.16 s) of the predicted and ground truth output sequences. Figure 3 shows an illustrative example of a predicted output sequence: the initial segment (in blue) corresponds to the warm up phase, and it is disregarded along with the subsequent early portion of the sequence (in orange). Only the final segment (in green) is retained for computing the loss function.

### 4.2. Numerical Results and Discussion

In this section, we present the numerical results obtained simulating the trained VWDF model. The training procedure is designed to characterize the nonlinear behavior of its root element, i.e., the BJT, which is modeled through an MLP. We refer to the proposed training approach as VWDF-MLP. Specifically, we evaluate its discrete-time simulation performance using held-out guitar audio waveforms as input voltage. To provide a comparative analysis, we consider two baseline approaches:

1. **VWDF with externally-trained MLP (MLP-2023)** – The same VWDF implementation based on an MLP (two layers with 32 units and ELU activation functions), but trained following the method in [22]. This approach assumes full knowledge of the Kirchhoff domain nonlinear characteristic of the element and employs it in a dedicated training phase to characterize the WD mapping of the $N$-port nonlinear element.
2. **Fully closed-box model (GRU)** – A Gated Recurrent Unit (GRU) model [8] with 16 hidden units (881 parameters), trained end-to-end using the same data as VWDF-MLP but without warm up. This approach directly models the circuit input-output behavior, without requiring any prior knowledge of the underlying schematic.

Table 2: *2N5089 BJT: extended EMM parameter values.*

| **Param.** | **Value** | **Param.** | **Value** |
|---|---|---|---|
| $V_{\text{th}}$ | 25.85 mV | $I_{\text{s1}}$ | 5.9 fA |
| $\alpha_{\text{f}}$ | 0.9993 | $I_{\text{s2}}$ | 10.6 fA |
| $\alpha_{\text{r}}$ | 0.5579 | $R_{\text{s1}}$ | $10^{-5}$ Ω |
| $\eta_{\text{f}}$ | 1 | $R_{\text{s2}}$ | $10^{-5}$ Ω |
| $\eta_{\text{r}}$ | 1 | $R_{\text{p1}}$ | $10^{11}$ Ω |
| | | $R_{\text{p2}}$ | $10^{8}$ Ω |

Table 3: *Mean Absolute Error (MAE) values for the three compared models.*

| | **VWDF-MLP** | **MLP-2023** | **GRU** |
|---|---|---|---|
| **MAE** [V] | $1.45 \times 10^{-3}$ | $1.11 \times 10^{-3}$ | $1.71 \times 10^{-3}$ |

As an evaluation metric, we use the Mean Absolute Error (MAE)

$$\text{MAE}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{K} \|\mathbf{y} - \hat{\mathbf{y}}\|_1 \,, \tag{20}$$

where $\mathbf{y}$ is a vector containing the samples of the ground-truth output voltage $V_{\text{out}}[k]$, $\hat{\mathbf{y}}$ denotes the model predictions, and $K$ is the length (in samples) of the considered sequences. For each model, we compute the MAE over the entire validation set. The results are summarized in Table 3. Overall, the three methods achieve comparable levels of accuracy. Among them, the MLP-2023 baseline yields the lowest MAE, which can be attributed to the fact that the externally-trained MLP is exactly tuned on the static nonlinear characteristic of the BJT. In contrast, both VWDF-MLP and GRU approaches rely exclusively on circuit input-output measurements. In particular, VWDF-MLP is tasked with learning the behavior of the nonlinear element from the overall circuit dynamic behavior, which poses a more complex identification problem.

To further assess model performance, Figure 4 provides a time-domain comparison for each approach on a representative segment of the validation waveform. In each case, the predicted output is plotted against the ground truth, while the corresponding MAE curve is shown in the lower accompanying plot. These plots highlight how closely each model is able to follow the real circuit

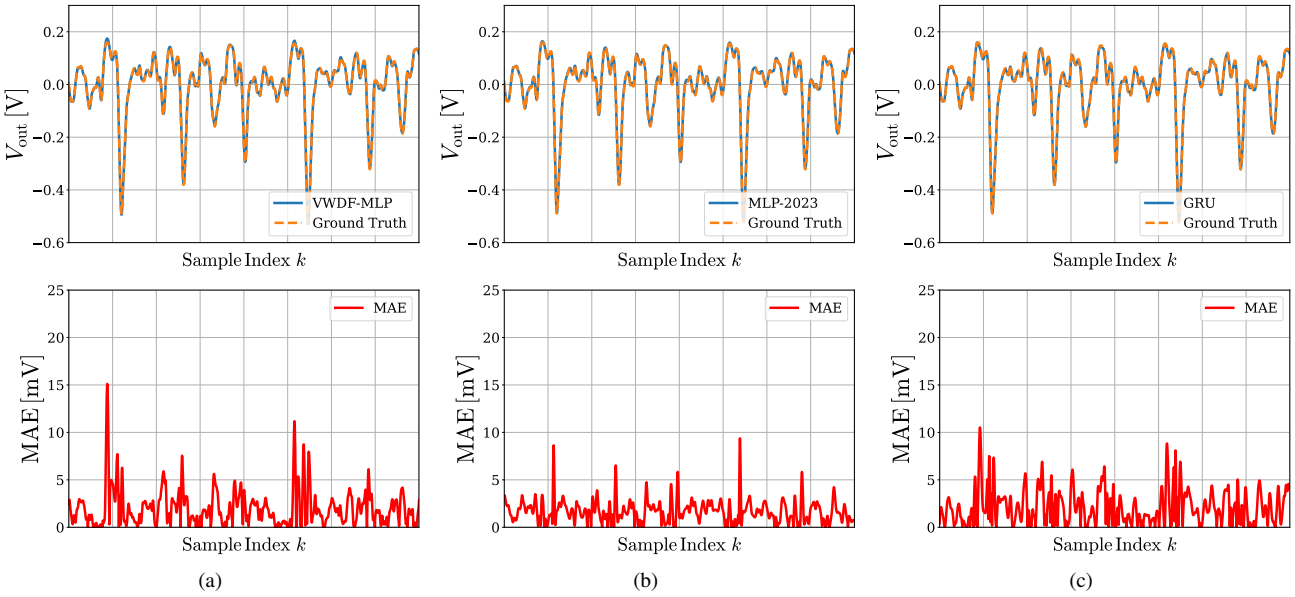


Figure 4: *Voltage measured across resistor $R_{out}$ for the three considered models: (a) VWDF-MLP model predictions (blue) vs. the ground truth (orange). (b) MLP-2023 model predictions (blue) vs. the ground truth (orange). (c) GRU model predictions (blue) vs. the ground truth (orange). The Mean Absolute Error (MAE) curve for each model is displayed in the corresponding bottom subplot (red).*

response in a general application scenario. Notably, VWDF-MLP tends to exhibit larger prediction errors in regions of the input signal where sharp positive transients or amplitude spikes occur—portions of the signal that are more likely to be clipped by the circuit's nonlinear behavior. This can be attributed to the limited presence of such high-energy events in the training dataset. Indeed, the dataset was not designed to uniformly span the full nonlinear behavior of the BJT, as in the case of the MLP-2023 baseline. Nonetheless, VWDF-MLP maintains strong generalization capabilities and achieves performance comparable to the other methods. Like MLP-2023, it retains the key advantage of being physically interpretable within the WDF framework. At the same time, it further extends this benefit to scenarios where the nonlinear element cannot be characterized in isolation and only input-output measurements from the complete circuit are available.

## 5. CONCLUSIONS

In this paper, we have introduced an approach for training neural models of nonlinear multi-port elements within a circuit's WD discrete-time implementation. This is accomplished integrating a differentiable parametric description of the $N$-port nonlinear element, namely a neural network, into the VWDF connection-tree structure. Since the discrete-time simulation of such structure is differentiable, the neural network parameters can be optimized through gradient-based methods to minimize a loss function defined on the circuit's output voltage. A key advantage of this approach is that it eliminates the need for prior knowledge about the nonlinear characteristics of the element being modeled, making it applicable to scenarios where directly characterizing a nonlinear multi-port in isolation from the rest of the circuit might not be feasible. This is achieved while maintaining the interpretability and modular structure inherent in the WDF framework. Our results showcase the effectiveness of the proposed methodology in capturing the nonlinear behavior of a reference circuit for VA applications featuring a single nonlinear BJT. The performance is comparable to that of the two considered baselines, namely the discrete-time VWDF simulation based on a pre-trained model of the BJT and a fully closed-box GRU implementation.

Future work might investigate the integration of Lipschitz-bounded neural architectures into VWDF structures modeling circuits containing one or more nonlinear multi-port elements. Such architectures provide a promising strategy for enforcing formal passivity and stability guarantees in the resulting discrete-time system, while also potentially supporting convergence guarantees for iterative methods in the WD domain [33]. The approach could also be extended to model circuit elements with dynamic nonlinearities by incorporating parametric nonlinear models with memory into the VWDF framework. Lastly, the developed methodology could be applied to the modeling of nonlinear acoustic systems, such as electroacoustic transducers, whose behavior is typically characterized solely through input-output signal pairs, without the possibility of isolating individual nonlinear contributions.

## 6. REFERENCES

[1] T. Adali, "An exciting juncture: The convergence of machine learning and signal processing [from the editor]," *IEEE Signal Processing Magazine*, vol. 41, no. 6, pp. 3–5, 2024.

[2] 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.

[3] T. Helie, "Volterra series and state transformation for real-time simulations of audio circuits including saturations: Application to the Moog ladder filter," *IEEE Transactions on*

*Audio, Speech, and Language Processing*, vol. 18, no. 4, pp. 747–759, 2010.

[4] T. Vanhatalo, P. Legrand, M. Desainte-Catherine, P. Hanna, A. Brusco, G. Pille, and Y. Bayle, "A review of neural network-based emulation of guitar amplifiers," *Applied Sciences*, vol. 12, no. 12, 2022.

[5] 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.

[6] A. Falaize-Skrzek and T. Hélie, "Simulation of an analog circuit of a wah pedal: a port-Hamiltonian approach," in *Proceedings of the 135th Audio Engineering Society (AES) Convention*, New York, USA, 2013, Audio Engineering Society.

[7] 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.

[8] A. Wright, E.P. Damskägg, V. Välimäki, et al., "Real-time black-box modelling with recurrent neural networks," in *Proceedings of the 22nd International Conference on Digital Audio Effects (DAFx-19)*, Birmingham, UK, 2019, University of Birmingham.

[9] V. Huhtala, L. Juvela, and S. J. Schlecht, "KLANN: Linearising long-term dynamics in nonlinear audio effects using Koopman networks," *IEEE Signal Processing Letters*, vol. 31, pp. 1169–1173, 2024.

[10] R. Simionato and S. Fasciani, "Modeling time-variant responses of optical compressors with selective state space models," *Journal of the Audio Engineering Society*, vol. 73, pp. 144–165, April 2025.

[11] A. Wright and V. Välimäki, "Neural modelling of periodically modulated time-varying effects," in *Proceedings of the 23rd International Conference on Digital Audio Effects (DAFx2020)*, Vienna, Austria, 2020, pp. 281–288, Vienna University of Music and Performing Arts.

[12] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Reviews Physics*, vol. 3, no. 6, pp. 422–440, 2021.

[13] J. Engel, L. Hantrakul, C. Gu, and A. Roberts, "DDSP: Differentiable digital signal processing," in *Proceedings of the 8th International Conference on Learning Representations (ICLR)*, 2020.

[14] F. Esqueda, B. Kuznetsov, and J. D. Parker, "Differentiable white-box virtual analog modeling," in *Proceedings of the 24th International Conference on Digital Audio Effects (DAFx)*, Vienna, Austria, 2021, pp. 41–48, IEEE.

[15] S. Nercessian, A. Sarroff, and K. J. Werner, "Lightweight and interpretable neural modeling of an audio distortion effect using hyperconditioned differentiable biquads," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 890–894.

[16] L. Kolonko, J. Velten, and A. Kummert, "Automatic differentiating wave digital filters with multiple nonlinearities," in *2020 28th European Signal Processing Conference (EUSIPCO)*, 2021, pp. 146–150.

[17] 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.

[18] A. Fettweis, "Wave digital filters: Theory and practice," *Proceedings of the IEEE*, vol. 74, pp. 270–327, 1986.

[19] 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.

[20] 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.

[21] 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.

[22] O. Massi, R. Giampiccolo, A. Bernardini, and A. Sarti, "Explicit vector wave digital filter modeling of circuits with a single bipolar junction transistor," in *Proceedings of the 26th International Conference on Digital Audio Effects (DAFx23)*, Copenhagen, Denmark, 2023, pp. 172–179, Aalborg University.

[23] A. I. Mezza, R. Giampiccolo, and A. Bernardini, "Data-driven parameter estimation of lumped-element models via automatic differentiation," *IEEE Access*, vol. 11, pp. 143601–143615, 2023.

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

[25] 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.

[26] R. Giampiccolo, S. C. Gafencu, and A. Bernardini, "Explicit modeling of audio circuits with multiple nonlinearities for virtual analog applications," *IEEE Open Journal of Signal Processing*, pp. 1–9, 2025.

[27] O. Massi, S. Yang, R. Giampiccolo, and A. Bernardini, "Explicit neural network-based modeling of time-varying circuits with a single BJT in the wave digital domain," in *2025 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2025, pp. 1–5.

[28] B. Holmes and M. van Walstijn, "Physical model parameter optimisation for calibrated emulation of the Dallas Rangemaster Treble Booster guitar pedal," in *Proceedings of the 19th International Conference on Digital Audio Effects (DAFx-16)*, Brno, Czech Republic, 2016, pp. 47–54, Brno University of Technology.

[29] 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.

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

[31] A. I. Mezza, M. Amerena, A. Bernardini, and A. Sarti, "Hybrid packet loss concealment for real-time networked music applications," *IEEE Open Journal of Signal Processing*, vol. 5, pp. 266–273, 2024.

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

[33] O. Massi, E. Manino, and A. Bernardini, "Wave digital modeling of circuits with multiple one-port nonlinearities based on Lipschitz-bounded neural networks," in *Proceedings of the 27th International Conference on Digital Audio Effects (DAFx24)*, Guildford, UK, 2024, pp. 119–126, University of Surrey.