

## Article

# Neural Network-Assisted DPD of Wideband PA Nonlinearity for Sub-Nyquist Sampling Systems

Mengqiu Liu <sup>1</sup>, Xining Yang <sup>2</sup>, Jian Gao <sup>2</sup>, Sen Cao <sup>2</sup>, Guisheng Liao <sup>1</sup>, Gaopan Hou <sup>1</sup> and Dawei Gao <sup>1,\*</sup><sup>1</sup> Hangzhou Institute of Technology, Xidian University, Hangzhou 311200, China;

lmengqiu@stu.xidian.edu.cn (M.L.); liaogs@xidian.edu.cn (G.L.); hougaoan@xidian.edu.cn (G.H.)

<sup>2</sup> 29th Research Institute of China Electronics Technology Group Corporation, Chengdu 610036, China;

yangxining@cetc.com.cn (X.Y.); swieegaoj@sina.com (J.G.); caosen@cetc.com.cn (S.C.)

\* Correspondence: gaodawei@xidian.edu.cn

**Abstract:** The design of conventional digital predistortion (DPD) requires an analogue-to-digital converter (ADC) with a sampling frequency that is multiple times the signal bandwidth, which is extremely challenging for sub-Nyquist sampling systems with undersampled signals. To address this, this paper proposes a neural network (NN)-assisted wideband power amplifier (PA) DPD method for sub-Nyquist sampling systems, wherein a dual-stage architecture is designed to handle the ambiguity caused by subsampled communications signals. In the first stage, the time-delayed polynomial reconstruction method is employed to estimate the wideband DPD nonlinearity coarsely with the undersampled signals with limited pilots. In the second stage, an NN-based DPD method is proposed for the virtual training of the DPD, which learns the up-sampled DPD behavior by taking advantage of the pre-estimated DPD model and the input data signals, which reduces the length of the training sequence significantly and refines the DPD behavior efficiently. Simulation results demonstrate the efficacy of the proposed method in tackling the wideband PA nonlinearity and its ability to outperform the conventional method in terms of power spectrum, error vector magnitude, and bit error rate.

**Keywords:** digital predistortion (DPD); undersampling; memory polynomial; attention mechanism model



Academic Editor: Yingjie Jay Guo

Received: 27 December 2024

Revised: 2 February 2025

Accepted: 11 February 2025

Published: 12 February 2025

**Citation:** Liu, M.; Yang, X.; Gao, J.; Cao, S.; Liao, G.; Hou, G.; Gao, D. Neural Network-Assisted DPD of Wideband PA Nonlinearity for Sub-Nyquist Sampling Systems. *Sensors* **2025**, *25*, 1106. <https://doi.org/10.3390/s25041106>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

With the rapid development of sixth-generation (6G) wireless communication, the spectrum width is increasing rapidly. This trend necessitates the application of ultra-wideband communication and spectrum sensing technologies. However, the elevated frequency band makes the nonlinearity of power amplifiers (PAs) severer than ever, where the nonlinearity becomes frequency-dependent, and memory effects need to be taken into account [1]. The nonlinear distortion of the wideband PA leads to poor signal detection performance, which needs to be handled appropriately [2]. Digital predistortion (DPD) techniques are proposed to mitigate the nonlinear distortion effectively [3,4], but the design of DPD usually requires three to five times the bandwidth of the input signal. This presents a significant challenge for broadband communications. In particular, when sub-Nyquist-sampled systems are employed for low-cost and efficient implementation using severely undersampled, low-speed ADCs (analogue-to-digital converters), conventional digital predistortion (DPD) methods cannot be applied directly [5]. Therefore, this study investigates the DPD method in the context of sub-Nyquist sampling of communication systems to achieve effective nonlinear correction of amplifiers, even when ADCs are undersampled.

Spectral extrapolation, a commonly used undersampling predistortion method proposed in [6], models the power amplifier (PA) under band-limited sampled data to recover or approximate the original output signal that should be sampled based on the estimated PA model. However, this approach has constraints that necessitate the inclusion of a band-limited filter in the feedback channel, which can be problematic in practical situations due to the performance of the band-limited filter and the challenges in its design. In [7], an undersampling recovery digital predistortion technique (USR-DPD) is proposed, which iteratively recovers the full-band output signal from the undersampled output signal, and then extracts the DPD parameter information using a memory model. However, this method requires the addition of extra DACs to the system, thereby increasing the system's complexity. Moreover, the method requires multiple internal and external iterations to obtain the final estimated DPD signal, which increases the computational cost. Ref. [8] proposed a forward behavior modeling approach that utilizes low-rate aliased PA output signals to estimate the DPD model coefficients. In this approach, the aliased low-rate PA output signals are employed to estimate the model parameters across the full frequency band of the PA. In [5], an undersampling digital predistortion method based on a multi-tone mixing feedback technique (MTM-DPD) is proposed for the processing of multi-tone mixing under undersampling conditions. In [9], a novel undersampled feedback signal iterative learning control (US-ILC-DPD) technique is proposed. This technique still exhibits good PA nonlinearity compensation even at low sampling rates. In [10], a random demodulation-based reduced sampling rate (RDRS) method is proposed to recover the full-band power amplifier (PA) output signal. This method uses only the I-way component of the band-limited feedback signal to reconstruct the full-band signal output. Although the method has a simple structure, it introduces a time alignment problem due to the additional random sequence generator and analog multiplier in the feedback path. This issue still needs to be addressed to ensure accurate signal reconstruction. In [11], a novel predistortion technique is proposed, which consists of a memoryless amplitude-to-amplitude (AM/AM) gain function that can be implemented in the analog domain and a nonlinear model with memory effects in the digital domain, reducing the sampling rate requirements of the forward and feedback paths. However, most of these methods have significant limitations. They are particularly vulnerable to noise and often fail to deliver satisfactory performance under the conditions of low signal-to-noise ratios (SNRs).

Artificial neural networks (NNs) have been rapidly developed in recent years and can be used for PA and DPD modeling as they can exhibit strong nonlinear feature representation without pre-conditioning [12,13]. Ref. [14] proposed a method for combining recurrent neural networks with PA modeling and DPD. Ref. [15] proposed the application of long short-term memory (LSTM) networks in the fields of amplifier modeling and digital predistortion. However, these models do not account for the downsampling environment. In [16], a digital predistorter, modeled by an augmented real-valued time delay neural network (ARVTDNN), was proposed to realize the correction of PA nonlinearity in broadband transmitters. The structure of this network is simple. However, its application scenarios are limited, and its performance decreases in some specific scenarios. In [17], a nonlinear correction method for power amplifiers is proposed, which combines a global optimization search algorithm with adaptive linearization to adaptively find the best parameters for nonlinear correction. In [18], a deep neural network (DNN) model was proposed for low-feedback estimation of DPD. However, this approach first requires the recovery of the model parameters of the PA before solving for the inverse model coefficients. This additional step increases the overall system complexity. Attentional mechanisms are inspired by the human visual system and are widely used in neural network models for their ability to improve model performance and model flexibility and generalization [19]. However,

the method does not account for sub-Nyquist sampling scenarios. Moreover, the sparse characteristics of the undersampled data stream can negatively impact the training process of neural networks (NNs). This issue is demonstrated in the simulation results and needs to be carefully addressed.

This paper proposes an NN-assisted wideband PA DPD method for sub-Nyquist sampling communication systems, where a dual-stage architecture is designed to handle the ambiguity caused by the down-sampled signals. In the first stage, a time-delayed memory polynomial reconstruction (TDMPR)-based method is implemented with the undersampled signals for the preliminary estimation of DPD parameters, which demands a limited number of pilot signals. Then, an attention-based NN is proposed for the refinement of DPD, where virtual training is performed with transmitted data signals, reducing the training overhead significantly. It is demonstrated through simulations that the proposed method can mitigate the wideband PA nonlinearity efficiently for sub-Nyquist sampling systems while providing outstanding performance gains compared to the existing downsampling DPD methods.

## 2. Signal Model

We consider a single-user uplink communications system. Considering the cost of mobile devices, we assume that the user has a single antenna where low-cost PA is employed, resulting in nonlinear distortion with memory effects. The feedback channel is equipped with a low-rate ADC. The system block diagram is shown in Figure 1.

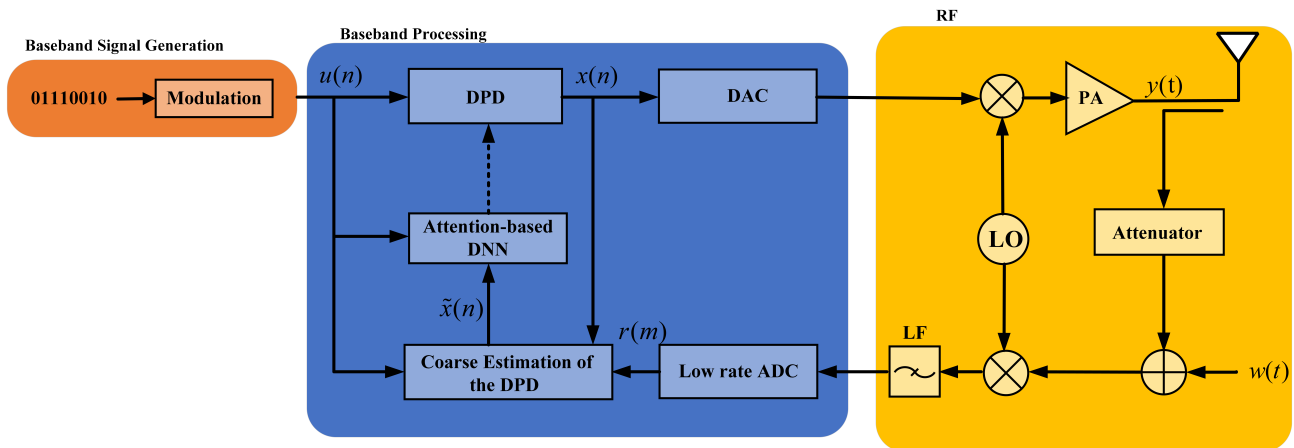


**Figure 1.** A system block diagram of the sub-Nyquist sampling system.

The modulated signal  $u(n)$  is first transmitted through a DPD module to generate  $x(n) = f_{\text{DPD}}(u(n))$ , where  $f_{\text{DPD}}(\cdot)$  denotes the transfer function of DPD, and the learning of the DPD is the focus of this work. The  $x(n)$  is first upconverted, then passed through the DAC, and, finally, fed into the PA. To represent the wideband PA's nonlinear behavior, the memory polynomial (MP) model is employed [20]. The MP model is represented as a baseband signal with a discrete form, i.e.,

$$y(n) = \sum_{k=1}^K \sum_{q=0}^Q h_{kq} x(n-q) |x(n-q)|^{k-1}, \quad (1)$$

where  $Q$  is the maximum memory benefit depth,  $K$  is the maximum nonlinear order, and  $k$  is only odd, while  $h_{kq}$  is the coefficient of the polynomial, and  $n = 0, 1, \dots, N-1$ . The reason for implementing only odd orders is that the main factor of nonlinear distortion is

the spectral aliasing and inter-modulation distortion produced by odd orders, so the effect of even orders can be ignored. (1) can be written in matrix form as follows:

$$\mathbf{y} = \mathbf{X}\mathbf{h}, \quad (2)$$

where

$$\mathbf{y} = [y(0), y(1), \dots, y(N-1)]^T \quad (3)$$

$$\mathbf{h} = [h_{10}, \dots, h_{1Q}, h_{30}, \dots, h_{3Q}, \dots, h_{K0}, \dots, h_{KQ}]^T. \quad (4)$$

Define  $\varphi_{kq}(n) = x(n-q)|x(n-q)|^{k-1}$ , and the matrix  $\mathbf{X}$  is represented as

$$\mathbf{X} = \begin{bmatrix} \varphi_{10}(0) & \varphi_{11}(0) & \cdots & \varphi_{KQ}(0) \\ \varphi_{10}(1) & \varphi_{11}(1) & \cdots & \varphi_{KQ}(1) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{10}(N-1) & \varphi_{11}(N-1) & \cdots & \varphi_{KQ}(N-1) \end{bmatrix}. \quad (5)$$

The PA output signal  $y(n)$  is transmitted through the wireless channel to obtain  $r(n)$ . Therefore, the feedback signal  $r(n)$  at the BS can be expressed as

$$r(n) = y(n) + \omega(n), \quad (6)$$

where  $\omega(n)$  is the noise signal, which is introduced in the feedback channel and is assumed to be Gaussian for simplicity. Due to the nonlinearity of the wideband PA, the spectrum of the feedback signal  $r(n)$  in the receiver is broadened, so an ADC with high sampling frequency is required to achieve unambiguous sampling. However, a sub-Nyquist sampling system, an ADC with a sampling rate that is much smaller than the Nyquist sampling frequency, is used to sample the feedback signal  $r(n)$ . The rate of the undersampled ADC is  $1/D$  ( $D$  is an integer, and  $D > 0$ ) of the Nyquist sampling frequency.  $r(n)$  is processed by the low-rate ADC, resulting in  $r(Dn)$ . Hence, our aim is to restore the transmitted signal  $u(n)$  from the undersampled feedback signals  $r(Dn)$  by designing a novel DPD architecture with limited pilots, as detailed in Section 3.

### 3. Proposed NN-Assisted Dual-Stage DPD for Sub-Nyquist Sampling Systems

#### 3.1. Stage 1: Time-delayed Memory Polynomial Reconstruction-Based Coarse Estimation of the Digital Predistorter

Directly using downsampled feedback signals for training a neural network model will prevent the model from accurately learning the behavior of the Digital Predistortion (DPD) system, leading to poor performance. Therefore, a preliminary estimation of the DPD signals is necessary before model training [8]. As depicted in Figure 1, after the feedback signal passes through a low-rate ADC, it must be combined with the high-rate complex baseband signal and fed into the DPD signal coarse estimation module, the structure of which is shown in Figure 2. The downsampled feedback signal is denoted as  $r(m)$ , and  $r(m) = r(Dn)$ . The complex baseband signal  $x(n)$  is first sampled to obtain  $x(m)$ , where  $x(m) = x(Dn)$ . Subsequently, the PA forward model is constructed using both  $r(m)$  and  $x(m)$ . The expression for the PA forward model is

$$r(m) = \sum_{k=1}^K \sum_{q=0}^Q g_{kq} x(m-q) |x(m-q)|^{k-1}, \quad (7)$$

where  $K$  and  $Q$  are the same as in (1). Therefore, this PA model can also be expressed in a matrix form, like (2):

$$\mathbf{r} = \mathbf{X}_D \mathbf{g}, \quad (8)$$

where

$$\mathbf{r} = [r(0), r(1), \dots, r(M-1)]^T \quad (9)$$

$$\mathbf{g} = [g_{10}, \dots, g_{1Q}, g_{30}, \dots, g_{3Q}, \dots, g_{K0}, \dots, g_{KQ}]^T. \quad (10)$$

The matrix  $\mathbf{X}_D$  is represented as

$$\mathbf{X}_D = \begin{bmatrix} \varphi_{10}(0) & \varphi_{11}(0) & \cdots & \varphi_{KQ}(0) \\ \varphi_{10}(1) & \varphi_{11}(1) & \cdots & \varphi_{KQ}(1) \\ \vdots & \vdots & & \vdots \\ \varphi_{10}(M-1) & \varphi_{11}(M-1) & \cdots & \varphi_{KQ}(M-1) \end{bmatrix}. \quad (11)$$

Using least square (LS) solutions, the coefficients  $\mathbf{g}$  can be obtained as

$$\mathbf{g} = (\mathbf{X}_D^H \mathbf{X}_D)^{-1} \mathbf{X}_D^H \mathbf{r}. \quad (12)$$

$\mathbf{g}$  is a parameter of the new PA model, which is constructed using the downsampling signal. With this new power amplifier model, a reconstructed feedback signal  $r_{rec}(n)$  based on a storage polynomial is obtained according to (2), where the parameter  $\mathbf{h}$  is replaced by  $\mathbf{g}$ . A preliminary DPD model is built using the reconstructed feedback signal  $r_{rec}(n)$  and the high-rate baseband complex signal  $x(n)$ . Like (2) and (8), it can be expressed in matrix form:

$$\mathbf{x} = \mathbf{R}_{rec} \mathbf{a}, \quad (13)$$

where  $\mathbf{x}$  is an  $N$ -dimensional column vector,  $\mathbf{a}$  is a column vector with  $(K+1)/2 * (Q+1)$  elements, and  $\mathbf{R}_{rec}$  is a  $N \times (K+1)/2 * (Q+1)$  matrix, which is derived in the same way as in Section 2. Using least square (LS) solutions, the coefficients  $\mathbf{a}$  can be obtained as

$$\mathbf{a} = (\mathbf{R}_{rec}^H \mathbf{R}_{rec})^{-1} \mathbf{R}_{rec}^H \mathbf{x}. \quad (14)$$

Based on the determined  $\mathbf{a}$ , the coarse estimation of the DPD behavior can be determined, and it can be used for the prediction of the DPD output whenever data are incorporated.

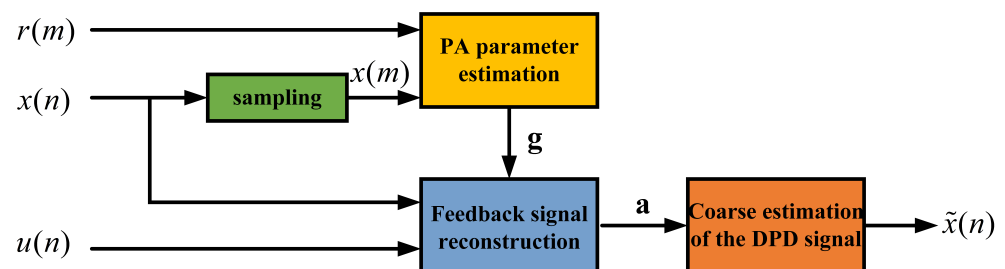


**Figure 2.** The structure of the proposed attention-based NN for virtual training.

### 3.2. Stage 2: An Attention-Based NN for the Virtual Training of DPD

Based on the coarse estimate of DPD from Stage 1, we propose an attention-based NN to refine the DPD, where virtual training is implemented with the input data and the trained MP model. We assume we have the new data  $\tilde{u}(n)$  coming in. Before refining the

DPD model, the DPD output can be predicted based on  $\mathbf{a}$ , which is calculated in the first stage, as expressed below:

$$\tilde{x}(n) = \sum_{k=1}^K \sum_{q=0}^Q a_{kq} u(n-q) |u(n-q)|^{k-1}. \quad (15)$$

With the data pairs  $\{(\tilde{\mathbf{u}}(n), \tilde{\mathbf{x}}(n))\}_{n=0}^{N_0-1}$ , we proposed an attention-based NN to refine the DPD behavior and used the pairs for the virtual training of the NN.

Figure 3 illustrates the structure of the proposed NN. The operation of each layer is given in the following subsections.

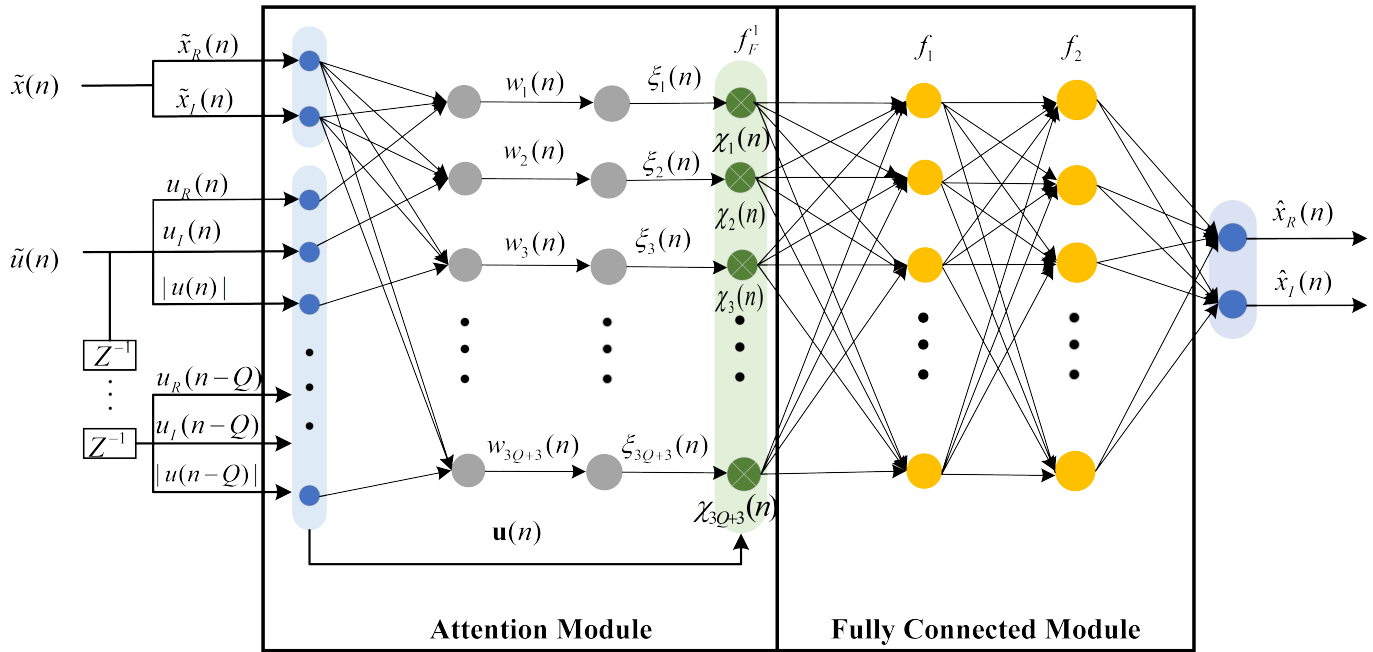


**Figure 3.** The structure of the attention-based NN proposed for virtual training.

### 3.2.1. Inputs and Outputs of the Proposed NN

The model's inputs encompass the real, imaginary, and norm of the current and historical input signals of the DPD, as well as the real and imaginary output signals of the DPD estimated from Stage 1. Consequently, the inputs can be represented as

$$\tilde{\mathbf{u}}(n) = [u_R(n), u_I(n), |u(n)|, \dots, |u(n-Q)|]^T, \quad (16)$$

where  $u_R(n)$  represents the real part of  $u(n)$ ,  $u_I(n)$  represents the imaginary part of  $u(n)$ ,  $|u(n)|$  is denoted as the norm of  $u(n)$ , and  $Q$  signifies the depth of memory. The NN's output is the estimated DPD signal with the real and imaginary parts separately, i.e.,

$$\tilde{\mathbf{x}}(n) = [\tilde{x}_R(n), \tilde{x}_I(n)]^T. \quad (17)$$

### 3.2.2. Attention Module

The front part of the model comprises an attention mechanism featuring two layers. The first one is a nonlinear layer with the activation function  $\tanh$ , where the size is equal to the inputs  $\tilde{\mathbf{u}}(n)$ . Hence, the  $i$ -th output from the first layer of the attention module is represented as

$$w_i(n) = \tanh(\alpha_i^T \tilde{\mathbf{u}}(n) + \beta_i^T \tilde{\mathbf{x}}(n) + b_i), \quad (18)$$

where  $w_i(n)$  denotes the correlation between the  $i$ -th input and the pre-estimated DPD output, and  $\mathbf{u}(n)$  denotes the data input at the time instant  $n$ .  $\{\alpha_i, \beta_i, b_i\}$  are the parameters in the attention module. The second layer is a weight calculation layer, which is utilized to calculate the weight of each input neuron and reduce the importance of the redundant estimated signals. The output from the second layer of the attention module is denoted as

$$\xi_i(n) = \frac{\exp(w_i(n))}{\sum_{i=1}^{3Q+3} \exp(w_i(n))}, (i = 1, 2, \dots, 3Q + 3). \quad (19)$$

Multiplying the calculated weights by the input  $U_n$  yields the weighted signal, as shown below:

$$\chi_i(n) = \xi_i(n) \tilde{u}_i(n). \quad (20)$$

### 3.2.3. Fully Connected Module

The weighted signal output from the attention module is then input to a module with a fully connected layer for regression. The module contains three hidden layers. All hidden layers use the  $\tanh$  activation functions. The output of the fully connected module can be simply expressed as

$$\hat{\mathbf{x}}(n) = \tanh(\mathbf{C}_2 \tanh(\mathbf{C}_1 \chi(n) + \mathbf{b}_1) + \mathbf{b}_2), \quad (21)$$

where  $\chi(n) = [\chi_1(n), \chi_2(n), \dots, \chi_{3Q+3}(n)]$ , and  $\mathbf{C}_1, \mathbf{C}_2, \mathbf{b}_1, \mathbf{b}_2$  denotes the parameters of the hidden layer, where  $\mathbf{C}_1$  can be denoted as

$$\mathbf{C}_1 = \begin{bmatrix} c_{1,11} & c_{1,12} & \cdots & c_{1,1(3Q+3)} \\ c_{1,21} & c_{1,22} & \cdots & c_{1,2(3Q+3)} \\ \vdots & \vdots & & \vdots \\ c_{1,L_1 1} & c_{1,L_1 2} & \cdots & c_{1,L_1(3Q+3)} \end{bmatrix}. \quad (22)$$

and  $\mathbf{b}_1$  can be denoted as

$$\mathbf{b}_1 = [b_{1,1}, b_{1,2}, b_{1,3}, \dots, b_{1,L_1}]^T. \quad (23)$$

Similarly,  $\mathbf{C}_2$  is expressed as

$$\mathbf{C}_2 = \begin{bmatrix} c_{2,11} & c_{2,12} & \cdots & c_{2,1(3Q+3)} \\ c_{2,21} & c_{2,22} & \cdots & c_{2,2(3Q+3)} \\ \vdots & \vdots & & \vdots \\ c_{2,L_2 1} & c_{2,L_2 2} & \cdots & c_{2,L_2(3Q+3)} \end{bmatrix}, \quad (24)$$

and  $\mathbf{b}_2$  is expressed as

$$\mathbf{b}_2 = [b_{2,1}, b_{2,2}, b_{2,3}, \dots, b_{2,L_2}]^T, \quad (25)$$

where  $L_1$  and  $L_2$  are the number of nodes in the first hidden layer and the second hidden layer, respectively.  $\tanh(\cdot)$  is the activation function, which has good approximation ability for DPDs [21], and it can be expressed as

$$\tanh(x) = \frac{1 - \exp(-2x)}{1 + \exp(-2x)}. \quad (26)$$

The proposed attention-based NN is trained using the Adam optimization algorithm and the loss of mean square error (MSE) to tune the parameters of the model, where the loss function is

$$Loss = \frac{1}{2} \frac{1}{N_0} \sum_{n=0}^{N_0-1} \sum_{i=1}^2 (\hat{x}_i(n) - \tilde{x}_i(n))^2. \quad (27)$$

The learning rate is set to 0.085. The specific steps of the proposed method are outlined in Algorithm 1.

---

**Algorithm 1** The Proposed Dual-Stage DPD Method for sub-Nyquist Sampling Systems

---

**Input**

1. Pilot signals: PA input signal  $\mathbf{x}$ , Downsampled PA input signal  $\mathbf{x}_D$ , feedback signal  $\mathbf{r}$ .
2. Data:  $\{\tilde{\mathbf{u}}(n)\}$ .

**Stage 1:**

1. PA nonlinear model reconstruction using sub-nyquist Nyquist sampling data and calculate the polynomial coefficients:

$$\mathbf{g} = (\mathbf{X}_D^H \mathbf{X}_D)^{-1} \mathbf{X}_D^H \mathbf{r}.$$

2. Reconstruction of the feedback signal from down-mining using the estimated PA model parameters:

$$\mathbf{r}_{\text{rec}} = \mathbf{X} \mathbf{g}.$$

3. The pre-estimated DPD parameters are obtained by building a PA inverse model based on the reconstructed feedback signal:

$$\mathbf{a} = (\mathbf{R}_{\text{rec}}^H \mathbf{R}_{\text{rec}})^{-1} \mathbf{R}_{\text{rec}}^H \mathbf{x}.$$

**Stage 2:**

Virtual training:

1. Generate expected DPD outputs using data:
 
$$\tilde{\mathbf{x}}(n) = \tilde{\mathbf{u}}(n) \mathbf{a}, n = 0, 1, \dots, N_0 - 1.$$
2. Virtual training of the DPD:
  - (1) Formulate the inputs and outputs of the proposed NN using (16) and (17);
  - (2) Calculate the output of the attention module as described in (20);
  - (3) Determine the output of the proposed NN based on (21);
  - (4) Update the hyper parameters according to the loss function in (27).

**Predicting**

Using the train NN to predict  $\hat{x}(n)$ .

---

## 4. Simulation Results

In this section, we conduct simulations to demonstrate the effectiveness of the proposed NN-assisted wideband PA DPD method. We consider a single-user uplink communications system. The 16-QAM modulated signal is generated. The baseband waveforms are generated by up-sampling five times and then passing through a raised cosine filter with a roll-off coefficient of 0.25. For the wideband PA modeling, the MP model is simulated with nonlinear order  $K = 5$ , and its memory depth is  $Q = 2$ . According to [22], the coefficients are  $[h_{10}, h_{30}, h_{50}, h_{11}, h_{31}, h_{51}, h_{12}, h_{32}, h_{52}] = [1.0513 + 0.0904j, -0.0542 - 0.2900j, -0.9657 - 0.7028j, -0.0680 - 0.0023j, 0.2234 + 0.2317j, -0.2451 - 0.3735j, 0.0289 - 0.0054j, -0.0621 - 0.0932j, 0.1229 + 0.1508j]$ . For the MP-based DPD estimation in Stage 1, the order and the memory remain the same. At the feedback channel, the feedback signal is undersampled by a factor of  $D = 4$ , which means that the ADC sampling rate only needs to be 1/4 of the Nyquist sampling frequency. For the proposed NN, the number of nodes in the hidden layers is [10, 5]. Regarding the number of training sets, 1000 and 100,000 sets of data are

used for validation. The number of Monte Carol runs is 500. For the evaluation metrics, the EVM is defined as

$$\text{EVM} = \sqrt{\frac{\frac{1}{N} \sum_{n=0}^{N-1} |\hat{x}(n) - x_t(n)|^2}{\frac{1}{N} \sum_{n=0}^{N-1} |x_t(n)|^2}}, \quad (28)$$

where  $N$  denotes the number of symbols measured by the EVM, and  $\hat{x}(n)$  denotes the  $n$ -th normalized feedback symbol, while  $x_t(n)$  denotes the ideal value of the  $n$ -th symbol. The SNR is defined as

$$\text{SNR(dB)} = 10 \lg \left( \frac{P_{\text{sig}}}{P_{\text{noi}}} \right), \quad (29)$$

where  $P_{\text{sig}}$  indicates signal power, and  $P_{\text{noi}}$  indicates noise power. Furthermore, the adjacent channel power ratio (ACPR) can be used to describe the power spectrum leakage into neighboring channels. It is expressed as the ratio of the leakage power in the neighboring channel to the signal power in the reference channel. A larger ACPR value indicates more severe leakage and, consequently, more serious nonlinear distortion. Since leakage can occur in both the upper and lower neighboring channels, the ACPR can be calculated separately for each. The upper neighborhood ACPR is expressed as

$$\text{ACPR}_u(\text{dB}) = \frac{P_{ACH}}{P_{REFC}} = 10 \lg \frac{\int_{ACH} P(f) df}{\int_{REFC} P(f) df}, \quad (30)$$

where  $P_{ACH}$  is the power of the signal in the upper neighboring channel, and  $P_{REFC}$  is the signal power in the reference channel. Similarly, the lower neighborhood ACPR can be expressed as

$$\text{ACPR}_l(\text{dB}) = \frac{P_{ACL}}{P_{REFC}} = 10 \lg \frac{\int_{ACL} P(f) df}{\int_{REFC} P(f) df}, \quad (31)$$

where  $P_{ACL}$  is the power of the lower neighboring channel signal.

In order to validate the effectiveness of the proposed method, the proposed method is compared with the augmented real-valued time delay neural network (ARVTDNN) method proposed in [16]. The ARVTDNN is an augmented real-valued time delay neural network. The ARVTDNN consists of an input layer, a hidden layer, and an output layer. The inputs to the network are the Cartesian components (I/Q) of the input signals and the envelope correlation. The network has only one hidden layer with 17 nodes, and the output of the network consists of the I and Q paths of the estimated signal. The simulation setups of the ARVTDNN in the simulations are the same as in [16]. We also simulated the method presented in [8], for which we employed time-delayed memory polynomial reconstruction for sub-nyquist-sampled DPD, denoted as TDMPR.

Figure 4 depicts the power spectrum density (PSD) of various DPD methods, including the proposed NN-based method, the ARVTDNN, TDMPR, a method based on the direct application of the DNN proposed in Stage 2, and a method without a DPD function, where the SNR is set as 15 dB.

It can be clearly seen that without DPD, the out-of-band spectrum growth cannot be suppressed at all, which is significantly detrimental to the communication performance. The PSD with the TDMPR-based DPD achieves a marginal performance gain compared to the one without DPD, which is due to the numerical instability of the MP, even with the aid of virtual data sequences. For the direct DNN-based method, the lack of processing of the downsampled feedback signals results in the network being unable to correctly learn the nonlinear behavior of the DPD, leading to a poor nonlinear correction capability. In contrast, our proposed NN-assisted DPD method provides significant performance gains

compared to the other methods. And the spectrum leakage is significantly improved with the proposed method compared to the method in [16].

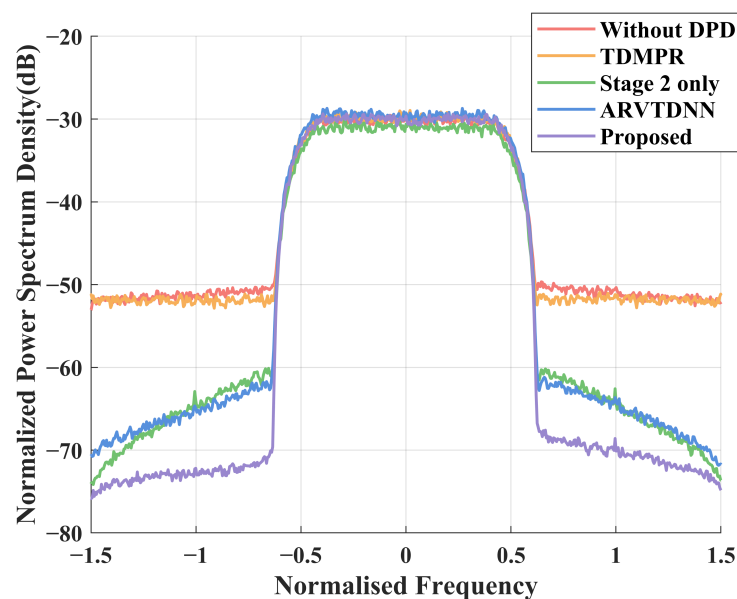


Figure 4. PSD performance of different methods at  $SNR = 15$  dB.

Table 1 compares the ACPR and EVM of DPD for various methods at  $SNR = 15$  dB. It can be seen that the proposed method has outstanding performance and reduces the ACPR from  $-20$  dB to  $-42$  dB. The ARVTDNN reduced the ACPR by 16 dB, which is effective but still limited compared to the proposed method. The performance of the DNN proposed in the method based on Stage 2 is close to that of the ARVTDNN, while the lack of coarse estimation leads to performance loss, which speaks to the efficiency of the proposed two-stage methods. In addition, TDMPR is poorly corrected at low SNRs, with ACPR being reduced by only 2 dB. In conclusion, it is evident that the proposed method has excellent corrective capabilities compared to the state-of-the-art methods.

To observe the detailed variations of these DPD methods under different SNRs, we simulated the proposed DPD method, TDMPR, and the method without DPD, and the results are shown in Figure 5. The SNR is from 5 dB to 35 dB. It can be seen that the proposed NN-assisted method has spectral bands without leakage along various SNR points, which demonstrates the robustness of the proposed method even under low SNRs. Although ARVTDNN also has good noise immunity, this method does not perform as well as the proposed method. In contrast, TDMPR provides unstable PSD performance and marginal performance gain. This suggests that relying solely on TDMPR for estimating the DPD signal is insufficient for mitigating the nonlinear effects induced by the PA, particularly when the SNR is low and the signal is undersampled.

Table 1. ACPR and EVM ( $SNR = 15$  dB).

| Model        | Lower ACPR/dB | Upper ACPR/dB | EVM%   |
|--------------|---------------|---------------|--------|
| Without DPD  | −18.6474      | −21.5771      | 10.25  |
| TDMPR        | −22.6645      | −22.5484      | 8.08   |
| Stage 2 only | −34.2174      | −34.0762      | 5.9943 |
| ARVTDNN      | −36.3610      | −36.0028      | 5.9746 |
| Proposed     | −43.4470      | −41.0365      | 1.3134 |

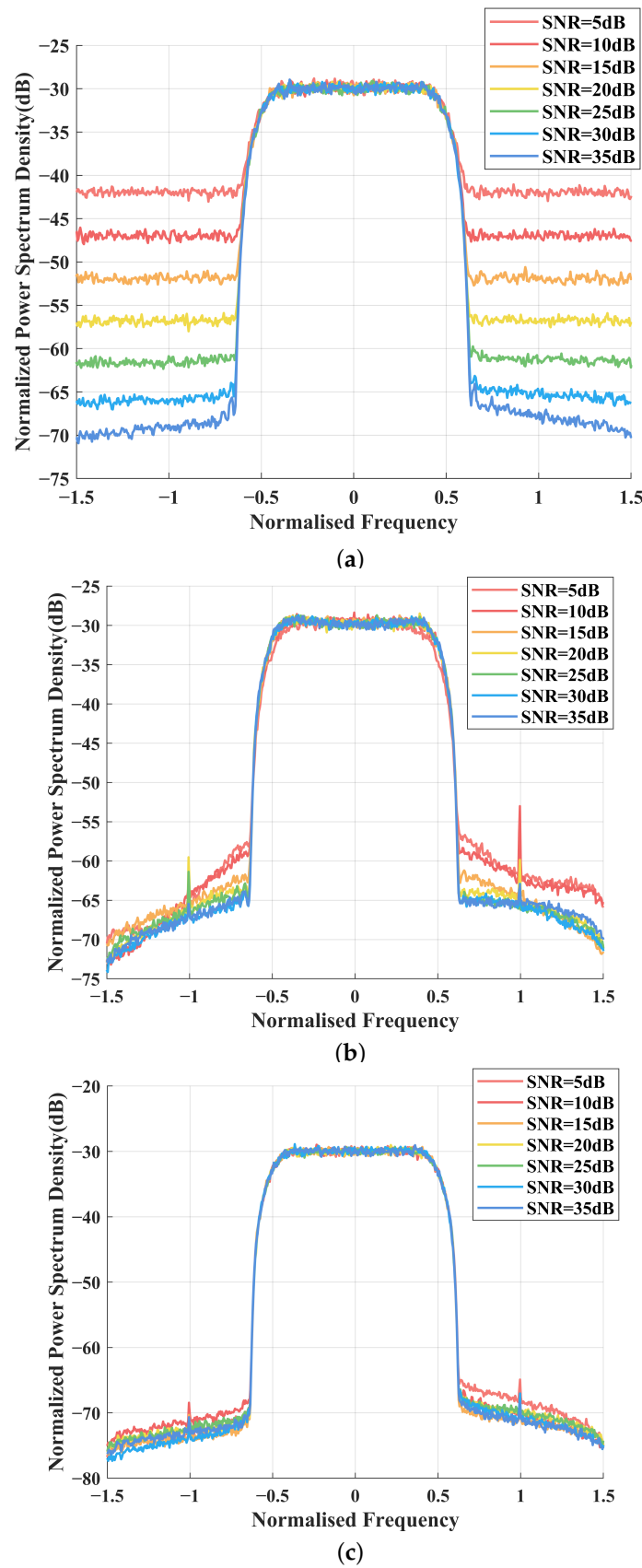


**Figure 5.** PSD performance of various DPD methods versus different SNRs: (a) PSD of PA output with TDMPR, (b) PSD of PA output with ARVTDNN, and (c) PSD of PA output with proposed NN-assisted DPD.

The effects of DPD can be more intuitively observed in Figure 6. The blue constellation diagram represents the signal before correction, showing significant distortion after passing through the PA. In contrast, the red constellation diagram represents the signal after correction. It is evident that the proposed method's distortion correction is significantly better than that of ARVTDNN, indicating superior PA correction performance.

Figure 7 shows the EVM performance of various DPD methods across different SNRs. It is evident that the scenario without DPD corresponds with deteriorated EVM performance. TDMPR offers only a modest performance improvement compared to the scenario without DPD. The ARVTDNN performs well when the SNR is lower, but it plateaus at around 8% and provides less efficient performance compared to TDMPR when the SNR is above around 18 dB. In contrast, the proposed NN-assisted DPD method delivers outstanding performance across all SNR levels, achieving an EVM of 1.014% at 35 dB.

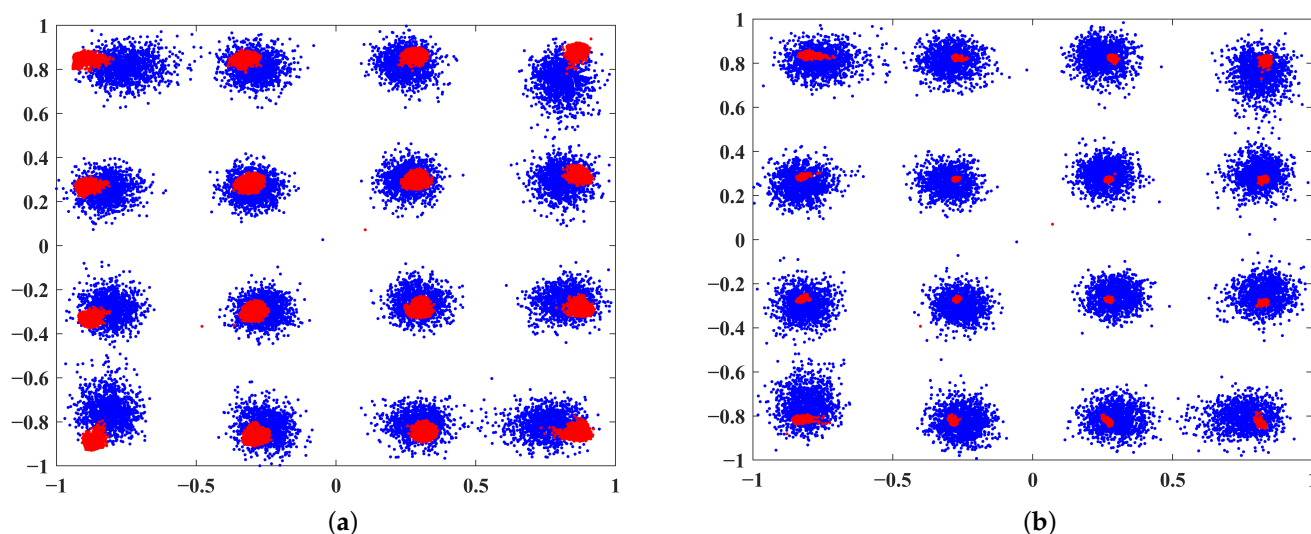


Figure 6. Constellation diagrams at  $SNR = 15$  dB: (a) ARVTDNN and (b) the proposed method.

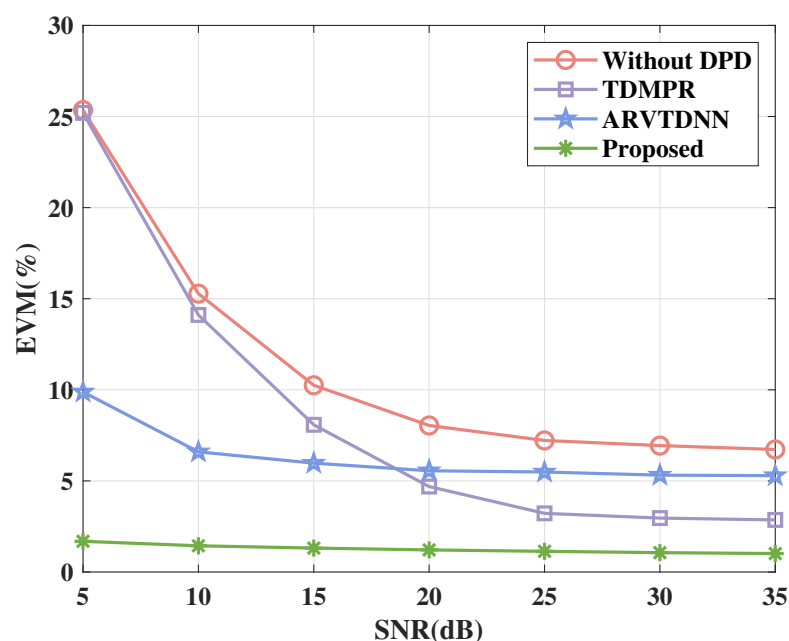


Figure 7. The EVM performance of various DPD methods versus different SNRs.

Figure 8 illustrates the bit error rate (BER) of various DPD methods across different SNRs. Specifically, both the proposed method and the ARVTDNN outperform the TDMPR-based method at low SNRs, validating the effectiveness of virtual training in noisy environments. Moreover, the proposed method demonstrates good nonlinearity correction performance at both low and high SNRs. Additionally, the BER performance of the system without DPD saturates easily around  $2 \times 10^{-2}$ , which significantly affects the communication performance of the sub-Nyquist sampling system.

Considering that the nonlinearity of the amplifier is affected by temperature and humidity, three different PA models are simulated in this paper to simulate the nonlinearity of the PA under different temperatures and humidity conditions. We still use the memory polynomial model to model the PA, which is shown in (1). As in [23], from experiments, we drew the conclusion that the higher humidity and temperature generally increase the nonlinear behavior of a PA. Therefore, by modifying the MP coefficients  $[h_{10}, h_{11}, h_{12}, h_{30}, h_{31}, h_{32}, h_{50}, h_{51}, h_{52}]$ , we can simulate various PAs with different levels of nonlinearity: normal nonlinearity for low temperature and low humidity, medium nonlinearity for mild temperature and mild humidity, and severe nonlinearity for high temperature and high humidity. Table 2 demonstrates the model parameters used at different temperatures and humidities.

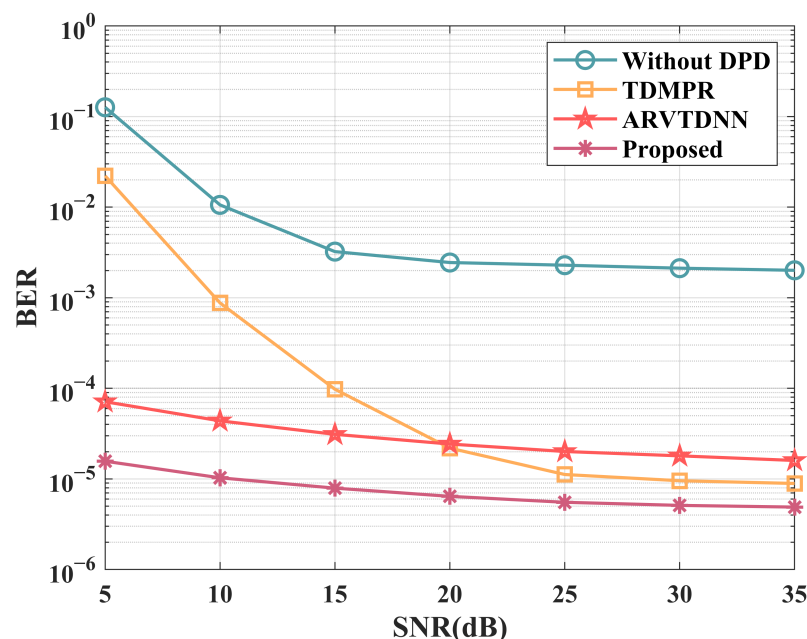


**Figure 8.** The BER performance of various DPD methods versus different SNRs.

Figure 9 shows the PSD performance of the proposed DPD with three different power amplifiers. It can be seen that the proposed DPD provides excellent performance even with high nonlinearity. In contrast, the performance without considering DPD (denoted as “W/o DPD” in the legends) degrades with the temperature and humidity change. We also simulated the constellation diagrams for these three different power amplifiers under nonlinear conditions. As can be seen in Figure 10a–c, the constellation points are well separated, which demonstrates the robustness and adaptability of the proposed NN-based DPD method.

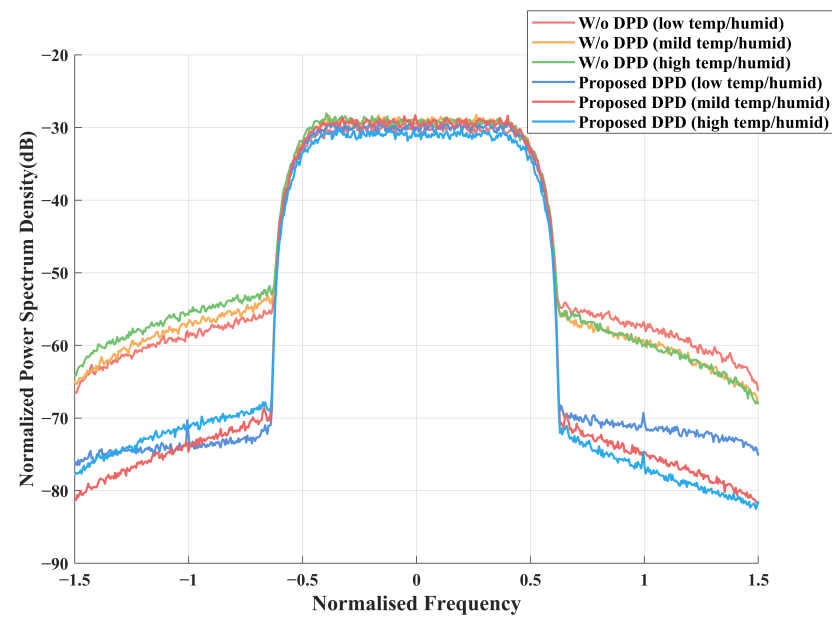


**Figure 9.** PSD of PA output with the proposed NN-assisted DPD with different levels of temperature and humidity.

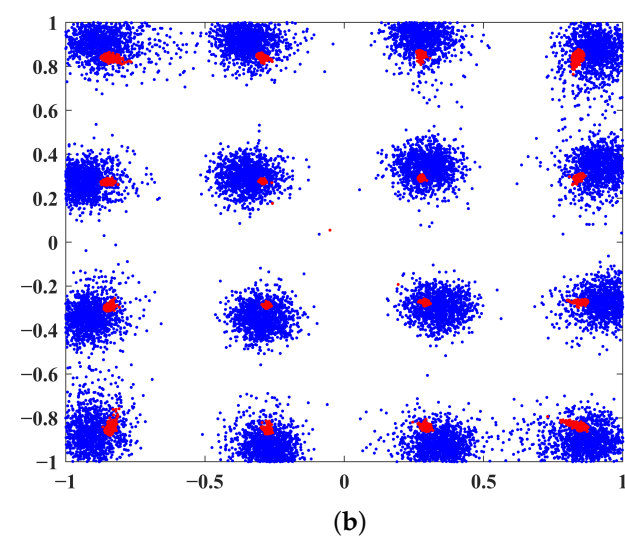
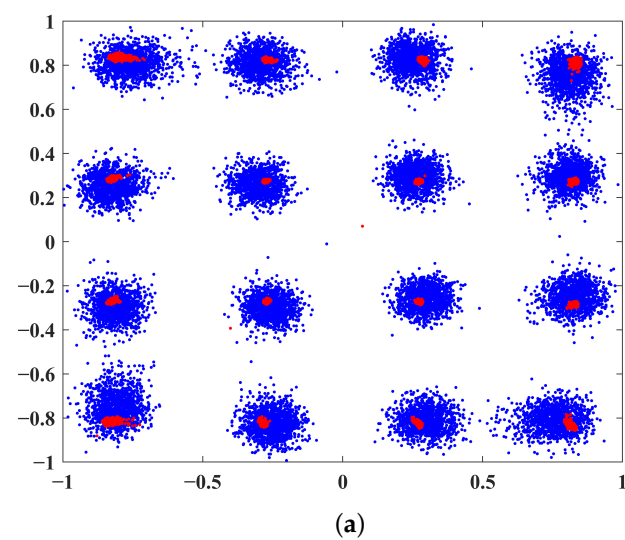


**Figure 10.** Cont.

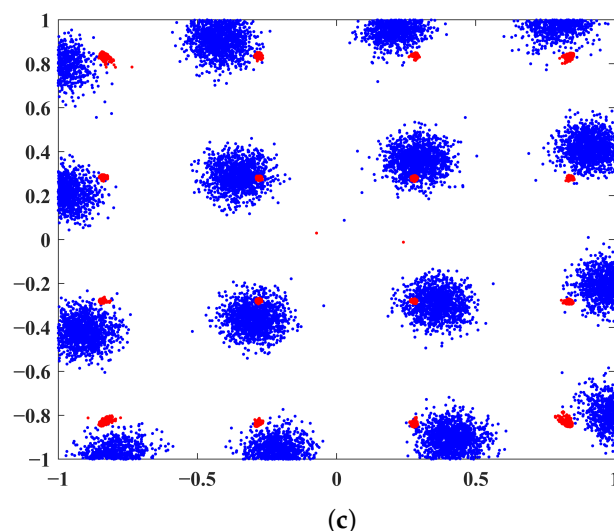


**Figure 10.** Constellation diagrams at different temperature/humidity levels: (a) low temperature/low humidity, (b) mild temperature/mild humidity, and (c) high temperature/high humidity.

**Table 2.** Memory polynomial coefficients for PA modeling.

| Simulating Various PA Nonlinearities Due to Temperature Drift and Humidity Variation | Memory Polynomial Coefficients for PA Modeling                                                                                                                             |
|--------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Simulate low temperature/low humidity (normal nonlinearity)                          | $[1.0513 + 0.0904j, -0.068 - 0.0023j, 0.0289 + 0.0054j, -0.0542 - 0.29j, 0.2234 + 0.2317j, -0.0621 - 0.0932j, -0.9657 - 0.7028j, -0.2451 - 0.3735j, 0.1229 + 0.1508j]$     |
| Simulate mild temperature/mild humidity (medium nonlinearity)                        | $[1.3883 + 0.1264j, 0.0082 + 0.0090j, -0.0017 - 0.0058j, 0.0996 - 0.2196j, -0.0818 + 0.01696j, -0.0205 + 0.0161j, 0.0071 - 0.0006j, -1.4735 - 1.2327j, -0.0209 - 0.0196j]$ |
| Simulate high temperature/high humidity (severe nonlinearity)                        | $[1.9067 + 0.1804j, 0.0075 + 0.0195j, 0.0108 - 0.0024j, -0.1403 + 0.0081j, -0.2134 - 0.0190j, 0.0437 + 0.0081j, -1.0739 - 0.1562j, -0.1196 + 0.1244j, 0.2187 + 0.1474j]$   |

## 5. Conclusions

In this paper, we proposed an NN-assisted DPD method tailored for sub-Nyquist sampling systems. To address the ambiguity problem resulting from undersampling, a two-stage architecture was introduced. In the first stage, a time delay MP reconstruction-based approach was employed to roughly estimate the broadband DPD nonlinearity using undersampled signals, requiring only a limited number of pilot signals. Subsequently, for the virtual training of DPDs, we proposed an attention-based NN DPD method in the second stage. This learning-based method leverages the transmitted data to learn the reconstructed DPD behaviors, significantly reducing the length of the training sequence and effectively refining the DPD behaviors. Our simulation results demonstrate that the proposed method can effectively mitigate the nonlinear distortion of broadband power amplifiers and provides outstanding performance gain compared to the state-of-the-art methods.

**Author Contributions:** Conceptualization, M.L. and D.G.; methodology, M.L. and D.G.; software, M.L. and X.Y.; validation, M.L., D.G. and J.G.; formal analysis, S.C.; investigation, X.Y. and J.G.; resources, S.C.; data curation, M.L. and D.G.; writing—original draft preparation, M.L. and D.G.; writing—review and editing, X.Y. and J.G.; visualization, G.H.; supervision, X.Y. and S.C.; project administration, G.L.; funding acquisition, G.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the “Pioneer” and “Leading Goose” R&D Program of Zhejiang under grant No. 2024C01079 and the National Natural Science Foundation of China under grant No. 62301394 and grant No. 62431021.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data are contained within the article.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

- Chen, X.; Jin, Z.; Zhang, S.; Xu, S.; Wang, Y. An Improved Digital Predistortion Mechanism via Joint Baseband and Radio Frequency Optimization. *IEEE Commun. Lett.* **2022**, *26*, 439–443. [\[CrossRef\]](#)
- Chen, W.; Liu, X.; Chu, J.; Wu, H.; Feng, Z.; Ghannouchi, F.M. A Low Complexity Moving Average Nested GMP Model for Digital Predistortion of Broadband Power Amplifiers. *IEEE Trans. Circuits Syst. I Regul. Pap.* **2022**, *69*, 2070–2083. [\[CrossRef\]](#)
- Crespo-Cadenas, C.; Madero-Ayora, M.J.; Becerra, J.A. Upgrading Behavioral Models for the Design of Digital Predistorters. *Sensors* **2021**, *21*, 5350. [\[CrossRef\]](#) [\[PubMed\]](#)
- Tian, S.; Yang, D.; Kuang, J.; Long, Z.; Chen, D. A Full Link Wideband Predistortion Based on Under-sampled Feedback Signal for Satellite Communications. In Proceedings of the 2020 International Conference on Space-Air-Ground Computing (SAGC), Beijing, China, 4–6 December 2020; pp. 54–59. [\[CrossRef\]](#)
- Peng, J.; You, F.; He, S. Under-Sampling Digital Predistortion of Power Amplifier Using Multi-Tone Mixing Feedback Technique. *IEEE Trans. Microw. Theory Tech.* **2022**, *70*, 490–501. [\[CrossRef\]](#)
- Ma, Y.; Yamao, Y.; Akaiwa, Y.; Ishibashi, K. Wideband Digital Predistortion Using Spectral Extrapolation of Band-Limited Feedback Signal. *IEEE Trans. Circuits Syst. I Regul. Pap.* **2014**, *61*, 2088–2097. [\[CrossRef\]](#)
- Liu, Y.; Yan, J.J.; Dabag, H.T.; Asbeck, P.M. Novel Technique for Wideband Digital Predistortion of Power Amplifiers With an Under-Sampling ADC. *IEEE Trans. Microw. Theory Tech.* **2014**, *62*, 2604–2617. [\[CrossRef\]](#)
- Wang, Z.; Chen, W.; Su, G.; Ghannouchi, F.M.; Feng, Z.; Liu, Y. Low Feedback Sampling Rate Digital Predistortion for Wideband Wireless Transmitters. *IEEE Trans. Microw. Theory Tech.* **2016**, *64*, 3528–3539. [\[CrossRef\]](#)
- Xu, Z.; Zhang, Q.; Zhang, L.; Yu, Z.; Yu, C.; Zhai, J. Iterative Learning Control for Digital Predistortion with Undersampled Feedback Signal. In Proceedings of the 2021 IEEE MTT-S International Wireless Symposium (IWS), Nanjing, China, 23–26 May 2021; pp. 1–3. [\[CrossRef\]](#)
- Zhang, Q.; Niu, J.; Zhang, L.; Yu, Z.; Yu, C.; Zhai, J. A Novel Undersampling Architecture for Wideband Digital Predistortion. In Proceedings of the 2020 IEEE MTT-S International Wireless Symposium (IWS), Shanghai, China, 20–23 September 2020; pp. 1–3. [\[CrossRef\]](#)
- Ahmed, S.; Ahmed, M.; Bensmida, S.; Hammi, O. Power Amplifier Predistortion Using Reduced Sampling Rates in the Forward and Feedback Paths. *Sensors* **2024**, *24*, 3439. [\[CrossRef\]](#) [\[PubMed\]](#)
- Liu, Z.; Hu, X.; Liu, T.; Li, X.; Wang, W.; Ghannouchi, F.M. Attention-Based Deep Neural Network Behavioral Model for Wideband Wireless Power Amplifiers. *IEEE Microw. Wirel. Components Lett.* **2020**, *30*, 82–85. [\[CrossRef\]](#)
- Jiang, Y.; Vaicaitis, A.; Dooley, J.; Leeser, M. Efficient Neural Networks on the Edge with FPGAs by Optimizing an Adaptive Activation Function. *Sensors* **2024**, *24*, 1829. [\[CrossRef\]](#) [\[PubMed\]](#)
- Luongvinh, D.; Kwon, Y. A Fully Recurrent Neural Network-Based Model for Predicting Spectral Regrowth of 3G Handset Power Amplifiers With Memory Effects. *IEEE Microw. Wirel. Components Lett.* **2006**, *16*, 621–623. [\[CrossRef\]](#)
- Liu, T.; Ye, Y.; Yin, S.; Chen, H.; Xu, G.; Lu, Y.; Chen, Y. Digital Predistortion Linearization with Deep Neural Networks for 5G Power Amplifiers. In Proceedings of the 2019 European Microwave Conference in Central Europe (EuMCE), Prague, Czech Republic, 13–15 May 2019; pp. 216–219.

16. Wang, D.; Aziz, M.; Helaoui, M.; Ghannouchi, F.M. Augmented Real-Valued Time-Delay Neural Network for Compensation of Distortions and Impairments in Wireless Transmitters. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 242–254. [[CrossRef](#)] [[PubMed](#)]
17. Wang, T.; Li, W.; Quaglia, R.; Gilabert, P.L. Machine-Learning Assisted Optimisation of Free-Parameters of a Dual-Input Power Amplifier for Wideband Applications. *Sensors* **2021**, *21*, 2831. [[CrossRef](#)] [[PubMed](#)]
18. Hu, X.; Liu, Z.; Wang, W.; Helaoui, M.; Ghannouchi, F.M. Low-Feedback Sampling Rate Digital Predistortion Using Deep Neural Network for Wideband Wireless Transmitters. *IEEE Trans. Commun.* **2020**, *68*, 2621–2633. [[CrossRef](#)]
19. Chen, X. The Advance of Deep Learning and Attention Mechanism. In Proceedings of the 2022 International Conference on Electronics and Devices, Computational Science (ICEDCS), Marseille, France, 20–22 September 2022; pp. 318–321. [[CrossRef](#)]
20. Lajnef, S.; Boulejfen, N.; Abdelhafiz, A.; Ghannouchi, F.M. Two-Dimensional Cartesian Memory Polynomial Model for Nonlinearity and I/Q Imperfection Compensation in Concurrent Dual-Band Transmitters. *IEEE Trans. Circuits Syst. II Express Briefs* **2016**, *63*, 14–18. [[CrossRef](#)]
21. Haykin, S. *Neural Networks: A Comprehensive Foundation*, 1st ed.; Prentice Hall PTR: Paramus, NJ, USA, 1994.
22. Ding, L.; Zhou, G.; Morgan, D.; Ma, Z.; Kenney, J.; Kim, J.; Giardina, C. A robust digital baseband predistorter constructed using memory polynomials. *IEEE Trans. Commun.* **2004**, *52*, 159–165. [[CrossRef](#)]
23. Zhou, S.; Jha, A.K. Characteristics Modeling of GaN Class-AB Dual-Band PA Under Different Temperature and Humidity Conditions. *IEEE Access* **2021**, *9*, 121632–121644. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.