B-Spline Neural Network-Based Multiuser MIMO-OFDM Nonlinear Uplink

Sheng Chen, *Life Fellow, IEEE*, Pengyu Wang, *Member, IEEE*, Mingkun Li, Emad F. Khalaf, Ali Morfeq, and Naif D. Alotaibi

***Abstract*—Multiple-input multiple-output (MIMO) technology in conjunction with orthogonal frequency division multiplexing (OFDM) transmission is widely adopted in fifth-generation mobile networks to support multiple users. However, in these mobile communication systems, high power amplifiers (HPAs) at user terminals' transmitters are driven into their saturation regions, which makes the multiuser frequency-selective MIMO-OFDM uplink channel nonlinear and renders the standard multiuser detection (MUD) at the base station (BS) ineffective. In this paper machine learning is employed to combat the distortions in the uplink of this multiuser frequency-selective MIMO-OFDM communication system. More specifically, a powerful complex-valued B-spline neural network (BSNN) based design is developed to simultaneously identify the system's channel impulse response (CIR) matrix and the BSNN model for the nonlinear transmitters' HPA together with the BSNN inversion for the nonlinear HPA at transmitters. This enables the BS to effectively implement MUD by utilizing the estimated MIMO-OFDM CIR matrix as well as to compensate for the transmitter HPAs' saturation distortions using the estimated BSNN inversion. A simulation study is included to evaluate the effectiveness of this novel BSNN assisted design in combating multiuser and dispersive channel interference as well as nonlinear distortions for multiuser MIMO-OFDM nonlinear uplink.**

***Index Terms*—Machine learning, B-spline neural network, nonlinear identification, nonlinear inversion, multiple-input multiple-output, dispersive channel, nonlinear transmit power amplifier, multiuser detection.**

I. Introduction

SINCE the standardization of the four-generation mobile communication system, orthogonal frequency division multiplexing (OFDM) [1] combined with multi-bit-per-symbol quadrature amplitude modulation (QAM) scheme [2] has become the dominate transmission technology, owing to its high spectrum efficiency, resisting fading and frequency selective distortions. However, OFDM signals with QAM signaling impose high peak-to-average power ratio (PAPR) and also require high average transmission power, which drives the transmitter high power amplifier (HPA) into its saturation region, causing serious nonlinear distortion to the transmitted signal. In the fifth-generation and future mobile networks, the multi-input multi-output (MIMO) enabled spatial-domain non-orthogonal multiple access plays a crucial role in supporting the massive access demands [3]. Historically, most existing designs for MIMO communication systems assume the linear MIMO channel model [4], [5], [6], [7], [8], [9], [10], [11]. However, this assumption is only valid if the transmitter HPA can operate in its linear dynamic range. Since real-world HPAs exhibit nonlinear saturation characteristics [12], [13], [14], [15], [16], it is impossible for a transmitter's HPA in the multiuser frequency-selective MIMO-OFDM system to operates in its linear dynamic range. This is because the multiuser MIMO-OFDM signals have very high PAPR. To operate the transmitter HPA in its linear dynamic range, output back-off (OBO) must be extremely severe, which means that the average transmission power must be very low. But such a low transmission power fails to meet the high link power budget requirement for successful transmission of high-order QAM based OFDM signals. Consequently, the multiuser frequency-selective MIMO-OFDM channel become nonlinear, which renders the powerful multiuser detection (MUD) for uplink and multiuser transmission for downlink ineffective.

It can be seen that for real-world multiuser MIMO-OFDM systems, the assumption of transmitter HPA operating in its linear dynamic range is invalid and the MIMO communication channel is nonlinear, owing to extremely high PAPR of OFDM signals. Explicitly, for the multiuser MIMO-OFDM systems with N subcarriers, the inverse fast Fourier transform (IFFT) operation at the transmitter scales up the PAPR of the transmitted signal by a factor of N. Since the cyclic prefix (CP) of each OFDM block reduces the spectrum efficiency, N must be large to maintain the effective throughput. Consequently, OFDM signals have the worst PAPR compared with other transmission techniques. Nevertheless, for the downlink, the problem can be tackled at the base station (BS) transmitter. This is because the BS can implement highly complicated PAPR reduction techniques [17], [18], [19], [20], [21], [22], [23] to mitigate high PAPR of transmitted OFDM signals as well as employ predistorters [24], [25], [26], [27], [28], [29], [30] to

Received 10 August 2025; revised 30 November 2025; accepted 9 January 2026. Date of publication 28 January 2026; date of current version 5 February 2026. This project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, under grant no. RG-3-135-40. The associate editor coordinating the review of this article and approving it for publication was H. Du. *(Corresponding author: Sheng Chen.)*

Sheng Chen is with the School of Electronics and Computer Science, University of Southampton, SO17 1BJ Southampton, U.K. (e-mail: sqc@ecs.soton.ac.uk).

Pengyu Wang and Mingkun Li are with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: wangpengyu@mail.tsinghua.edu.cn; lmk23@mails.tsinghua.edu.cn).

Emad F. Khalaf, Ali Morfeq, and Naif D. Alotaibi are with the Electrical and Computer Engineering Department, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia (e-mail: ekhalaf@kau.edu.sa; morfeq@kau.edu.sa; ndalotabi@kau.edu.sa).

Digital Object Identifier 10.1109/TCOMM.2026.3657448

compensate for nonlinear distortions of BS transmitter HPA. But for uplink, a mobile handset transmitter does not have the capacity to implement these high-complexity operations. This leaves the BS receiver to deal with the severely nonlinear MIMO-OFDM frequency-dispersive channel. Consequently, detecting multiple mobile users' transmitted data becomes an extremely challenging task at BS receiver.

A. Our Novelty and Contribution

In this paper machine learning is employed to tackle this challenging task. By utilizing the optimal modeling and inverse modeling capability of the complex-valued B-spline neural network (BSNN) for complex-valued nonlinear functions, we develop an effective and efficient machine-learning assisted design for this challenging multiuser nonlinear frequency-dispersive MIMO-OFDM uplink. Our original contributions are summarized as follows.

- We design a novel BSNN parametrized nonlinear multiuser MIMO-OFDM uplink, and develop a highly effective nonlinear identification framework to estimate the channel impulse response (CIR) matrix of the nonlinear multiuser frequency-selective MIMO-OFDM uplink and the nonlinear mapping of transmitter HPA as well as the inverse model of nonlinear transmitter HPA. Numerical results demonstrate that this BSNN based nonlinear identification scheme is highly accurate and converges extremely fast, requiring no more than 2 iterations. Also this BSNN based nonlinear identification scheme is computationally extremely efficient, as it involves two closed-form least squares (LS) estimations in each iteration.
- A powerful nonlinear MUD scheme is proposed using the BSNN based nonlinear identification results. Specifically, BS implements the per subcarrier MUD using the estimated multiuser MIMO-OFDM CIR matrix to remove the multiuser frequency-selective MIMO channel distortions, and this is followed by removing the nonlinear distortions of the transmitters' HPA using the estimated BSNN inverse model. Simulation study confirms the effectiveness of this nonlinear MUD scheme.

B. Related Works and Their Limitations

Our work is very different from [31]. The study [31] considers the frequency-nonselective MIMO uplink and does not employ OFDM. In practice, MIMO channels are frequency-selective, and OFDM transmission is typically adopted to tackle the frequency-selective MIMO channels. This work is also completely different from [32]. The paper [32] considers the single-carrier multiuser frequency-selective MIMO uplink. The BS receiver employs a bank of time-domain (TD) space-time equalizers to mitigate multiuser and frequency-selective channel interference, which imposes dramatically higher computational complexity, compared with our per-subcarrier MUD for multiuser MIMO-OFDM uplink. The nonlinear identification framework developed in this paper is very different from that of [32]. In [32], the single-carrier multiuser nonlinear frequency-selective MIMO channel is parametrized with three sets of parameters, and these three sets of parameters enter the model in a triple product form. The iterative alternating LS (ALS) estimator of [32] therefore requires the two-loop iterative procedure involving three steps of LS estimates. In this paper, we parametrize the nonlinear multiuser frequency-selective MIMO-OFDM uplink with two sets of parameters, i.e., with less total number of parameters. This enables us to design the single-loop iterative ALS procedure involving only two steps of LS estimates. In other words, our identification procedure imposes much lower computational complexity and converges much faster than the one proposed in [32].

Recently, deep learning has been extensively applied to MIMO communication systems. Many of the works in this field focus on deep neural networks (DNNs) based beamforming for downlink [10], [11], [33], [34]. In order for the BS to perform downlink beamforming, mobile users have to estimate the downlink channels and feed back the estimated channels to the BS. For uplink transmission, many studies have applied DNNs for MUD [35], [36], [37], [38], [39]. Typically, a linear frequency-nonselective MIMO channel is assumed, and the BS is assumed to have the perfect MIMO CIR matrix. With the perfect linear MIMO CIR matrix, various DNNs are trained to approximate the maximum likelihood or maximum a posteriori MIMO detection solutions [35], [36], [37], [38], [39]. There are several limitations of these DNN-based MIMO detectors, which prevent their applications to the practical multiuser frequency-selective MIMO-OFDM nonlinear uplink considered in this paper. First, training of such large-size DNN detectors requires huge amount of training data, which is often impractical. Second, standard channel estimation methods have difficulty to acquire the estimate of this realistic multiuser nonlinear frequency-selective MIMO channel. Third, to cope with this realistic multiuser nonlinear frequency-selective MIMO channel, the DNN detector would become prohibitively large, severely limiting its practicality. Alternatively, DNN detector can be trained directly to bypass channel estimation, at the cost of demanding even larger training data and longer training period.

Existing DNN detectors developed in the literature belong to the so-called black-box approach. A fundamental principle in data modeling is to incorporate available a priori information regarding the underlying data generating mechanism into the modeling process. Data-physics or grey-box models that are capable of incorporating prior knowledge typically outperform pure data-driven modeling [40], [41], [42]. The nonlinear multiuser MIMO detector developed in this paper is based on this grey-box modeling approach. Specifically, we explicitly utilize the physical knowledge that the multiuser MIMO-OFDM nonlinear uplink is an MIMO Hammerstein system consisting of the transmitters' nonlinear HPAs followed by frequency-selective MIMO CIR matrix. This knowledge allows us to design a powerful nonlinear estimator to simultaneously identify the MIMO CIR matrix and the BSNN model for the nonlinear transmitters' HPA together with the BSNN inversion for the nonlinear HPA at transmitters. With the aid of this physical knowledge, our nonlinear MIMO channel estimator is extremely efficient, as

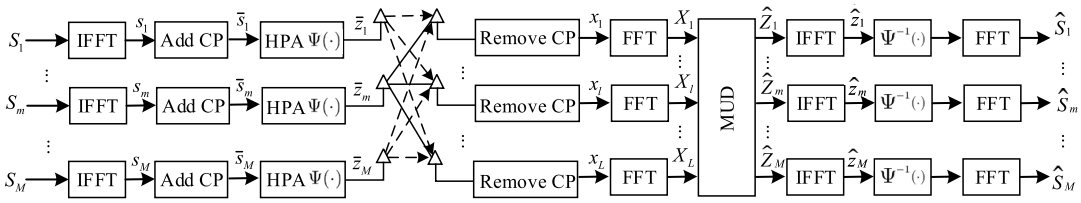


Fig. 1. Schematic of multiuser nonlinear frequency-selective MIMO-OFDM uplink.

it is an ALS procedure of no more than two iterations, with each iteration involving two closed-form linear LS estimations. By contrast, any DNN approach relies typically on gradient-descent nonlinear optimization to estimate its huge number of parameters, which requires huge number of training data and imposes huge computational complexity. Furthermore, in data detection stage, we can effective implement low-complexity linear MUD by utilizing the estimated MIMO-OFDM CIR matrix as well as to compensate for the transmitter HPAs' saturation distortions using the estimated BSNN inversion. This is in strikingly contrast to the high computational complexity of a DNN detector, which may not be able to meet the real-time constraint of data detection for the nonlinear multiuser MIMO-OFDM uplink considered in this paper.

It is also worth emphasizing that in this work, we adopt a more realistic nonlinear HPA model, and in our system, transmitter HPA actually operates into its saturation region owing to the extremely high PARA of OFDM signals. This is in contrast to most existing works, where an odd-order polynomial HPA model is typically adopted [43], [44], [45], [46]. Such a nonlinear HPA model is less effective in modeling the saturation characteristics of the HPA under the operation conditions of extremely-high-PARA OFDM signals.

II. Multiuser Nonlinear Frequency-Selective MIMO-OFDM Uplink

A. Signal Model

The uplink communication system considered is depicted in Fig. 1, where all the M single-antenna mobile users transmit to the BS using the same resource block of N subcarriers. The BS equipped with L receiving antennas, where $L \geq M$, is responsible for detecting all the M users' data. The dispersive channel's CIR linking the mth mobile user to the lth receive antenna of the BS can be expressed as

$$\boldsymbol{g}_{l,m} = \left[g_{l,m,0}\ g_{l,m,1} \cdots g_{l,m,n_H-1}\right]^{\mathrm{T}}, \tag{1}$$

where $1 \leq l \leq L$ and $1 \leq m \leq M$. Without loss of generality, the CIRs of all the links between the users and the BS have the same length of n_H.

The modulation scheme is assumed to be U-QAM. For national simplification, we omit the block index, and each OFDM transmission block of mobile user m consists of N data symbols on N subcarriers, which can be expressed as

$$\boldsymbol{S}_m = \left[S_{m,0}\ S_{m,1} \cdots S_{m,N-1}\right]^{\mathrm{T}},\ 1 \leq m \leq M. \tag{2}$$

The frequency-domain (FD) data symbols $S_{m,k} \in \mathbb{S}$ for $0 \leq k \leq N-1$ and $1 \leq m \leq M$, where the U-QAM constellation set is defined as

$$\mathbb{S} = \{d(2l-\sqrt{U}-1)+\mathrm{j}d(2q-\sqrt{U}-1), 1 \leq l, q \leq \sqrt{U}\}, \tag{3}$$

with $2d$ denoting the distance between adjacent constellation points. The N-point IFFT modulator at the transmitter of user m first converts $\boldsymbol{S}_m$ into the TD OFDM signal:

$$\boldsymbol{s}_m = \boldsymbol{F}^{\mathrm{H}}\boldsymbol{S}_m = \left[s_{m,0}\ s_{m,1} \cdots s_{m,N-1}\right]^{\mathrm{T}}, \tag{4}$$

where $\boldsymbol{F}$ is the fast Fourier transform (FFT) matrix, which satisfies $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{F} = \boldsymbol{F}\boldsymbol{F}^{\mathrm{H}} = \boldsymbol{I}_N$. Adding the CP of length N_{cp} to $\boldsymbol{s}_m$ results in

$$\bar{\boldsymbol{s}}_m = \left[s_{m,-N_{\mathrm{cp}}}\ s_{m,-N_{\mathrm{cp}}+1} \cdots s_{m,-1} \middle| \boldsymbol{s}_m^{\mathrm{T}}\right]^{\mathrm{T}}, \tag{5}$$

in which $s_{m,-k} = s_{m,N-k}$, $1 \leq k \leq N_{\mathrm{cp}}$, and $N_{\mathrm{cp}} \geq n_H$.

The TD signal block $\bar{\boldsymbol{s}}_m$ is amplified by the HPA to yield the actually transmitted signal block of the mth mobile user

$$\begin{aligned}\bar{\boldsymbol{z}}_m &= \left[z_{m,-N_{\mathrm{cp}}}\ z_{m,-N_{\mathrm{cp}}+1} \cdots z_{m,-1} \middle| z_{m,0}\ z_{m,1} \cdots z_{m,N-1}\right]^{\mathrm{T}} \\ &= \left[z_{m,-N_{\mathrm{cp}}}\ z_{m,-N_{\mathrm{cp}}+1} \cdots z_{m,-1} \middle| \boldsymbol{z}_m^{\mathrm{T}}\right]^{\mathrm{T}},\end{aligned} \tag{6}$$

where $z_{m,-k} = z_{m,N-k}$, $1 \leq k \leq N_{\mathrm{cp}}$, and

$$z_{m,k} = \Psi\left(s_{m,k}\right), -N_{\mathrm{cp}} \leq k \leq N-1 \tag{7}$$

In (7), $\Psi(\cdot)$ denotes the mapping of the nonlinear HPA at a mobile user's transmitter. Here we assume that all the M transmitters have the same type of HPA and therefore the transmitter index is dropped from $\Psi(\cdot)$. Given the input to the HPA $s_{m,k} = |s_{m,k}|e^{\mathrm{j}\angle s_{m,k}} = r_{m,k}e^{\mathrm{j}\angle s_{m,k}}$, the output of most real-world HPAs can be written as [15] and [16]

$$z_{m,k} = A(r_{m,k})e^{\mathrm{j}\left(\angle s_{m,k} + \Upsilon(r_{m,k})\right)}, \tag{8}$$

where the HPA's amplitude response is defined by

$$A(r) = \frac{g_a r}{\left(1 + \left(\frac{g_a r}{A_{\mathrm{sat}}}\right)^{2\beta_a}\right)^{\frac{1}{2\beta_a}}}, \tag{9}$$

the HPA's phase response is defined by

$$\Upsilon(r) = \frac{\alpha_\phi r^{q_1}}{1 + \left(\frac{r}{\beta_\phi}\right)^{q_2}}\ [\text{degree}], \tag{10}$$

and r is the magnitude of of the complex-valued input to the HPA. In other words, a practical nonlinear HPA is characterized by its amplitude response $A(r)$ with the parameters g_a, β_a and A_{sat}, and its phase response $\Upsilon(r)$ with the parameters α_ϕ, β_ϕ, q_1 and q_2 [15], [16]. In particular, A_{sat} represents the

HPA's saturation level. The HPA's operating point, typically specified by its average input power or its average output power, does not define the HPA's operating status. Rather, the OBO more accurately specifies the operating condition of the HPA, which is defined as

$$\text{OBO} = 10\log_{10}\frac{P_{\max}}{P_{\text{aop}}}\ [\text{dB}], \tag{11}$$

where $P_{\max}(=A_{\text{sat}}^2)$ and P_{aop} are the maximum output power and the average output power of the HPA output signal, respectively. The OBO specifies how many dBs the average output power is 'back off' from the maximum output power. For the fairness, every mobile user is allocated with the same transmission power, i.e., $\mathsf{E}\{|z_{m,k}|^2\}=\sigma_z^2$ for $1 \le m \le M$.

After the CP removal at the BS receiver, the MIMO channel-impaired sampled TD received signal blocks are obtained as

$$\boldsymbol{x}_l = \left[x_{l,0}\ x_{l,1}\cdots x_{l,N-1}\right]^{\text{T}}, 1 \le l \le L, \tag{12}$$

where $x_{l,k}$ is the k-th signal sample at the lth receive antenna which can be expressed as

$$\begin{aligned} x_{l,k} &= \sum_{m=1}^{M}\sum_{i=0}^{n_H-1} g_{l,m,i} z_{m,k-i} + \xi_{l,k} \\ &= \sum_{m=1}^{M} \boldsymbol{g}_{l,m}^{\text{T}} \boldsymbol{z}_{m,k}^{(C)} + \xi_{l,k}. \end{aligned} \tag{13}$$

In (13), $z_{m,k-i} = z_{m,N+k-i}$ for $k < i$ owing to the circular property of CP, $\xi_{l,k}$ denotes the complex-valued additive white Gaussian noise (AWGN) with $\mathsf{E}[|\xi_{l,k}|^2] = 2\sigma_\xi^2$, and $\boldsymbol{z}_{m,k}^{(C)} = \left[z_{m,k}\ z_{m,k-1}\cdots z_{m,k-n_H+1}\right]^{\text{T}}$. The TD receive signal vector at sample k of the BS's antenna array $\boldsymbol{x}_k^{(C)} = [x_{1,k}\ x_{2,k}\cdots x_{L,k}]^{\text{T}}$ on the other hand can be written by

$$\begin{aligned} \boldsymbol{x}_k^{(C)} &= \begin{bmatrix} \boldsymbol{g}_{1,1}^{\text{T}} & \boldsymbol{g}_{1,2}^{\text{T}} & \cdots & \boldsymbol{g}_{1,M}^{\text{T}} \\ \boldsymbol{g}_{2,1}^{\text{T}} & \boldsymbol{g}_{2,2}^{\text{T}} & \cdots & \boldsymbol{g}_{2,M}^{\text{T}} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{g}_{L,1}^{\text{T}} & \boldsymbol{g}_{L,2}^{\text{T}} & \cdots & \boldsymbol{g}_{L,M}^{\text{T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{z}_{1,k}^{(C)} \\ \boldsymbol{z}_{2,k}^{(C)} \\ \vdots \\ \boldsymbol{z}_{M,k}^{(C)} \end{bmatrix} + \boldsymbol{\xi}_k^{(C)} \\ &= \boldsymbol{G}\boldsymbol{z}_k^{(C)} + \boldsymbol{\xi}^{(C)}(k), \end{aligned} \tag{14}$$

where the MIMO-OFDM CIR matrix $\boldsymbol{G} \in \mathbb{C}^{L\times(n_H M)}$ and $\boldsymbol{\xi}_k^{(C)} = [\xi_{1,k}\ \xi_{2,k}\cdots\xi_{L,k}]^{\text{T}}$.

Remark 1: The HPA of (9) and (10) is much more accurate than the odd-order polynomial HPA model typically adopted in the literature [43], [44], [45], [46]. This is because the amplitude response (9) is particularly suitable for representing the saturated operation characteristics of the HPA in response to extremely-high-PAPR OFDM signals. The saturation output power of this HPA is $P_{\max} = A_{\text{sat}}^2$. Consider the typical values of the parameters for this nonlinear HPA as given in [15] and [16]

$$\begin{aligned} &A(\cdot):\ g_a = 19,\ \beta_a = 0.81,\ A_{\text{sat}} = 1.4; \\ &\Upsilon(\cdot):\ \alpha_\phi = -48000,\ \beta_\phi = 0.123,\ q_1 = 3.8,\ q_2 = 3.7. \end{aligned} \tag{15}$$

Fig. 2 plots this nonlinear HPA's amplitude and phase response. The HPA's operating point as specified by the average input magnitude and average output magnitude is indicated in the graph for the operation condition of OBO = 3 dB. Owing to extremely high PAPR of OFDM signals, the peak input magnitude is well reaching the HPA's saturation region with the true HPA's peak output magnitude saturated at A_{sat}.

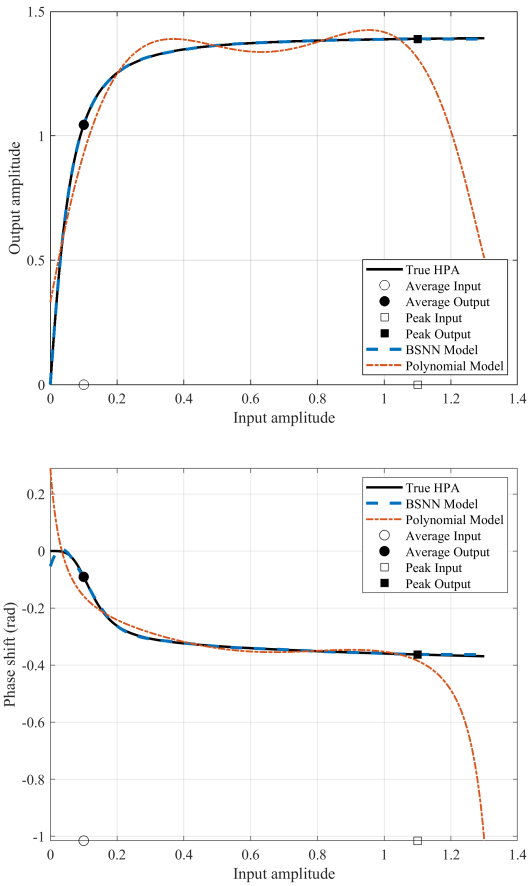


Fig. 2. The nonlinear HPA's amplitude and phase response with the parameters given by (15) in comparison with their real-valued BSNN and polynomial based estimates. The operating point, i.e., average input magnitude and average output magnitude, for OBO = 3 dB, is indicated in the graph.

If the PAPR of the input signal is not too extreme, say, the peak input magnitude is 4 times of the average input magnitude, then within this operating region, this nonlinear HPA can be accurately approximated by a polynomial model. However, for extremely-high-PAPR OFDM signals, the peak magnitude of the HPA input is well reaching into the HPA's output saturation region, causing the peak magnitude of the HPA output to saturate at A_{sat}. The polynomial HPA model [43], [44], [45], [46] has difficulty to accurately model this situation. By contrast, the BSNN HPA model developed in this paper is capable of accurately modeling this nonlinear saturation characteristics. To demonstrate this, we perform the LS fittings for the HPA's amplitude and phase response using the real-valued one dimensional (1D) BSNN model with 8 basis functions and polynomial degree 4 (See Section III) as well as polynomial model with polynomial degree 4. The training inputs are uniformly sampled from the HPA input range (0 to peak input), and desired outputs are the corresponding true HPA's output magnitudes and phases. The response of the estimated BSNN and polynomial based HPA models are

also plotted in Fig. 2. The results clearly demonstrate that the polynomial HPA model is incapable of representing this physical HPA accurately.

B. Multiuser Detection

The TD signal blocks $\boldsymbol{x}_l$ of the l-th receive antenna for $1 \le l \le L$ are first converted by the N-point FFT to yield the FD received signal blocks $\mathbf{X}_l = \boldsymbol{F}\boldsymbol{x}_l = \left[X_{l,0}\ X_{l,1} \cdots X_{l,N-1}\right]^{\mathrm{T}}$, $1 \le l \le L$. From (13), by utilizing the N-period cyclic property of $\bar{\boldsymbol{z}}_m$, it is well-known that

$$X_{l,n} = \sum_{m=1}^{M} G_{l,m,n} Z_{m,n} + \Xi_{l,n},\ 0 \le n \le N-1, \tag{16}$$

where the FD noise vector $\boldsymbol{\Xi}_l = [\Xi_{l,0}\ \Xi_{l,1} \cdots \Xi_{l,N-1}]^{\mathrm{T}} = \boldsymbol{F}[\xi_{l,0}\ \ \xi_{l,1} \cdots \xi_{l,N-1}]^{\mathrm{T}}$ with $\mathsf{E}\{|\Xi_{l,n}|^2\} = 2\sigma_\xi^2$, $\mathbf{Z}_m = \left[Z_{m,0} Z_{m,1} \cdots Z_{m,N-1}\right]^{\mathrm{T}} = \boldsymbol{F}\boldsymbol{z}_m$ with $\mathsf{E}\{|Z_{m,n}|^2\} = \sigma_z^2$, and the FD channel vector $\mathbf{G}_{l,m} = \left[G_{l,m,0}\ G_{l,m,1} \cdots G_{l,m,N-1}\right]^{\mathrm{T}}$ is the N-point FFT of $\boldsymbol{g}_{l,m}$ for $1 \le l \le L$ and $1 \le m \le M$.

It is well known that without the nonlinear HPA distortion at the M transmitters, the MUD can be done on the per subcarrier basis. Because of the nonlinear distortion introduced at the transmitters' HPA, however, the output of this MUD is not the estimate of the FD data symbols $\boldsymbol{S}_m$, $1 \le m \le M$. This can be be easily seen. Specifically, let us define $\underline{\mathbf{X}}_n = \left[X_{1,n}\ X_{2,n} \cdots X_{L,n}\right]^{\mathrm{T}}$, $\underline{\mathbf{Z}}_n = \left[Z_{1,n}\ Z_{2,n} \cdots Z_{M,n}\right]^{\mathrm{T}}$ and $\underline{\boldsymbol{\Xi}}_n = \left[\Xi_{1,n}\ \Xi_{2,n} \cdots \Xi_{L,n}\right]^{\mathrm{T}}$ for $0 \le n \le N-1$ as well as the nth subcarrier FD channel matrix

$$\underline{\mathbf{G}}_n = \begin{bmatrix} G_{1,1,n} & G_{1,2,n} & \cdots & G_{1,M,n} \\ G_{2,1,n} & G_{2,2,n} & \cdots & G_{2,M,n} \\ \vdots & \vdots & \cdots & \vdots \\ G_{L,1,n} & G_{L,2,n} & \cdots & G_{L,M,n} \end{bmatrix}. \tag{17}$$

Then (16) for $1 \le l \le L$ can be collected together as

$$\underline{\mathbf{X}}_n = \underline{\mathbf{G}}_n \underline{\mathbf{Z}}_n + \underline{\boldsymbol{\Xi}}_n,\ 0 \le n \le N-1. \tag{18}$$

Thus, the detection of $\underline{\mathbf{Z}}_n$ can be obtained as

$$\underline{\widehat{\mathbf{Z}}}_n = \underline{\mathbf{W}}_n \underline{\mathbf{X}}_n,\ 0 \le n \le N-1. \tag{19}$$

Here $\underline{\mathbf{W}}_n \in \mathbb{C}^{M \times L}$ denotes the linear MUD's weight matrix for the nth subcarrier. When the FD channel matrix at the nth subcarrier, $\underline{\mathbf{G}}_n$, is known at the BS receiver, the minimum mean square error (MMSE) solution for $\underline{\mathbf{W}}_n$ is given by:

$$\underline{\widehat{\mathbf{W}}}_n = \left(\underline{\mathbf{G}}_n^{\mathrm{H}} \underline{\mathbf{G}}_n + \frac{2\sigma_\xi^2}{\sigma_z^2} \boldsymbol{I}_M\right)^{-1} \underline{\mathbf{G}}_n^{\mathrm{H}},\ 0 \le n \le N-1. \tag{20}$$

Obviously, $\widehat{\mathbf{Z}}_m = \left[\widehat{Z}_{m,0}\ \widehat{Z}_{m,1} \cdots \widehat{Z}_{m,N-1}\right]^{\mathrm{T}}$ is not the estimate of the FD data block $\boldsymbol{S}_m$. Converting $\widehat{\mathbf{Z}}_m$ by the N-point IFFT yields the TD signal block given by

$$\widehat{\boldsymbol{z}}_m = \boldsymbol{F}^{\mathrm{H}} \widehat{\mathbf{Z}}_m = \left[\widehat{z}_{m,0}\ \widehat{z}_{m,1} \cdots \widehat{z}_{m,N-1}\right]^{\mathrm{T}}. \tag{21}$$

Observe from (7) that the actual TD transmitted signal block of mobile user m, $\boldsymbol{z}_m$, is the nonlinear HPA distorted TD signal block $\boldsymbol{s}_m$, i.e.,

$$\boldsymbol{z}_m = \Psi(\boldsymbol{s}_m) = \left[\Psi(s_{m,0})\ \Psi(s_{m,1}) \cdots \Psi(s_{m,N-1})\right]^{\mathrm{T}}. \tag{22}$$

Therefore, if the BS receiver also knows the inverse mapping $\Psi^{-1}(\cdot)$ of the transmitter's nonlinear HPA $\Psi(\cdot)$, it can estimate the mth mobile user's TD transmitted data block $\boldsymbol{s}_m$, where $1 \le m \le M$, according to

$$\begin{aligned} \widehat{\boldsymbol{s}}_m &= \Psi^{-1}\left(\widehat{\boldsymbol{z}}_m\right) \\ &= \left[\Psi^{-1}\left(\widehat{z}_{m,0}\right)\ \Psi^{-1}\left(\widehat{z}_{m,1}\right) \cdots \Psi^{-1}\left(\widehat{z}_{m,N-1}\right)\right]^{\mathrm{T}}. \end{aligned} \tag{23}$$

The estimate of the mth mobile user's FD transmitted data block $\boldsymbol{S}_m$ can then be obtained as the N-point FFT of $\widehat{\boldsymbol{s}}_m$:

$$\widehat{\boldsymbol{S}}_m = \boldsymbol{F}\widehat{\boldsymbol{s}}_m,\ 1 \le m \le M. \tag{24}$$

III. BSNN-Based Receiver Implementation

From the previous section, it can be seen that the BS receiver needs to know the multiuser MIMO-OFDM CIR matrix $\boldsymbol{G}$ and the inversion of the nonlinear HPA $\Psi(\cdot)$ at transmitters to detect the mobile users' data. Once $\boldsymbol{G}$ and $\Psi^{-1}(\cdot)$ are obtained, MUD can be implemented on the per subcarrier basis, and the nonlinear distortions to the transmitted data caused at transmitters can be removed at receiver. From the multiuser nonlinear frequency-selective MIMO-OFDM channel (14), it is clear that the multiuser MIMO-OFDM CIR matrix $\boldsymbol{G}$ is coupled with the nonlinear HPA $\Psi(\cdot)$, and it is necessary to jointly estimate $\boldsymbol{G}$ and $\Psi(\cdot)$ in order to acquire $\boldsymbol{G}$. This section develops an effective and efficient identification procedure for acquiring $\boldsymbol{G}$ and $\Psi(\cdot)$ as well as $\Psi^{-1}(\cdot)$.

However, the nonlinear MIMO channel (14) has an inherent permutation and scale ambiguity owing to the coupling of $\boldsymbol{G}$ and $\Psi(\cdot)$. To remove this ambiguity and therefore ensure a unique estimate of $\boldsymbol{G}$ and $\Psi(\cdot)$, we parametrize the multiuser nonlinear MIMO-OFDM channel as

$$\begin{aligned} \boldsymbol{G}\boldsymbol{z}_k^{(C)} = &\begin{bmatrix} \frac{1}{g_{1,1,0}}\boldsymbol{g}_{1,1}^{\mathrm{T}} & \frac{1}{g_{1,1,0}}\boldsymbol{g}_{1,2}^{\mathrm{T}} & \cdots & \frac{1}{g_{1,1,0}}\boldsymbol{g}_{1,M}^{\mathrm{T}} \\ \frac{1}{g_{1,1,0}}\boldsymbol{g}_{2,1}^{\mathrm{T}} & \frac{1}{g_{1,1,0}}\boldsymbol{g}_{2,2}^{\mathrm{T}} & \cdots & \frac{1}{g_{1,1,0}}\boldsymbol{g}_{2,M}^{\mathrm{T}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{g_{1,1,0}}\boldsymbol{g}_{L,1}^{\mathrm{T}} & \frac{1}{g_{1,1,0}}\boldsymbol{g}_{L,2}^{\mathrm{T}} & \cdots & \frac{1}{g_{1,1,0}}\boldsymbol{g}_{L,M}^{\mathrm{T}} \end{bmatrix} \\ &\times \begin{bmatrix} g_{1,1,0}\boldsymbol{z}_{1,k}^{(C)} \\ g_{1,1,0}\boldsymbol{z}_{2,k}^{(C)} \\ \vdots \\ g_{1,1,0}\boldsymbol{z}_{M,k}^{(C)} \end{bmatrix}. \end{aligned} \tag{25}$$

This parametrization corresponds to scale every CIR vector $\boldsymbol{g}_{l,m}$ by $g_{1,1,0}$ and multiply the nonlinear HPA's response $\Psi(\cdot)$ with $g_{1,1,0}$. For notational simplification, we still use $\boldsymbol{G}$ and $\Psi(\cdot)$ to represent the scaled MIMO-OFDM CIR matrix and the modified nonlinear HPA's response.

A. B-Spline Neural Network Parametrization of HPA

An effective estimator first requires an effective parametrization. The multiuser MIMO-OFDM CIR matrix $\boldsymbol{G}$ is naturally parametrized by its CIR coefficients, and we still need an effective parametrization of the nonlinear HPA mapping $\Psi(\cdot)$.

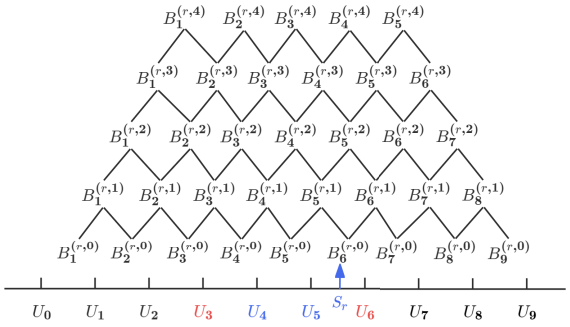


Fig. 3. Illustration of De Boor recursion or the structure of one-dimensional BSNN for $P_o = 4$, $N_{\rm r} = 5$, $U_{\min} = U_3$ and $U_{\max} = U_6$.

We choose a complex-valued BSNN to parametrize the nonlinear HPA (7). This choice is based on the computational efficiency and optimal modeling capability of the BSNN, which will be discussed in **Remarks 2 and 3**. Furthermore, the BSNN modeling well matches the underlying physics of HPA and QAM signaling. Obviously, the real-life HPA's nonlinearity $\Psi(\cdot)$ is a continuous and invertible mapping. This offers sound theoretical basis for using the BSNN to model $\Psi(\cdot)$ (and to model $\Psi^{-1}(\cdot)$). From the QAM signaling (3), the FD data symbol $S = S_R + {\rm j}S_I$ is upper and lower bounded by the constellation points, $\pm d\big(\sqrt{U}-1\big)(1+{\rm j})$, and the distributions of its real and imaginary parts, S_R and S_I, are identical and symmetric. The input to the HPA $\Psi(\cdot)$ is the TD signal $s = s_R + {\rm j}s_I$, which is the IFFF of the FD data symbols via (4). Therefore, physically, $U_{\min} < s_{\rm r} < U_{\max}$, for some known finite values $U_{\min}$ and $U_{\max}$, and we use $s_{\rm r} \in \mathbb{R}$ to represent either s_R or s_I. Clearly, the distributions of s_R and s_I are also identical and symmetric. The BSNN is particularly well suited for modeling in such a finite interval.

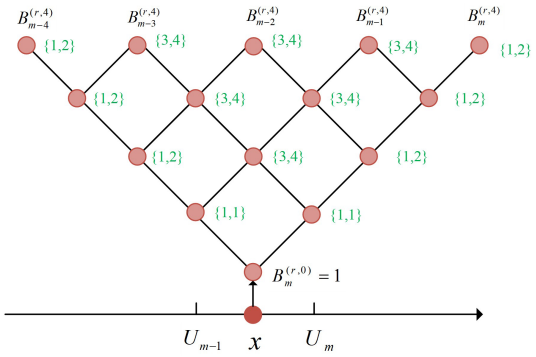
1) 1D BSNN Modeling: To model a nonlinearity in the 1D space of $s_{\rm r}$, we choose a 1D BSNN model with piecewise polynomial degree of P_o and $N_{\rm r}$ basis functions. Our reason for choosing the BSNN rather than other nonlinear models is because its optimal modeling capability [47], [48], [49] (also see Remark 3), that is, it is an optimal choice for our task. The 'receptive field' of this 1D BSNN is partitioned by the knot sequence specified by $(N_{\rm r}+P_o+1)$ fixed knot values, $\{U_0, U_1, \cdots, U_{N_{\rm r}+P_o}\}$, arranged in the order

$$\begin{aligned} U_0 < U_1 < \cdots < U_{P_o-2} < U_{P_o-1} = U_{\min} < U_{P_o} < \cdots < \\ U_{N_{\rm r}} < U_{N_{\rm r}+1} = U_{\max} < U_{N_{\rm r}+2} < \cdots < U_{N_{\rm r}+P_o}. \end{aligned} \tag{26}$$

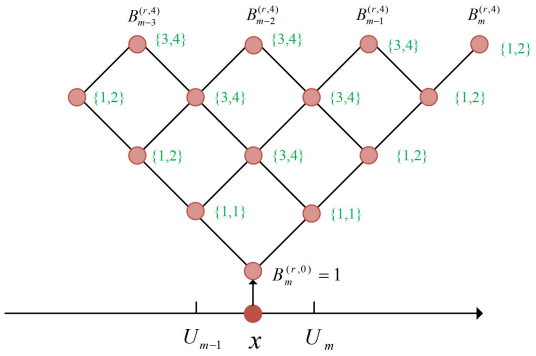
Outside the input region $\big[U_{\min},\ U_{\max}\big]$, there are $P_o - 1$ external knots and one boundary knot at each end. The number of so-called internal knots that cover the input region is given by $N_{\rm r}+1-P_o$. Based on $\{U_0, U_1, \cdots, U_{N_{\rm r}+P_o}\}$, the famous De Boor recursion [50] can be applied to determine the set of $N_{\rm r}$ B-spline basis functions.

Specifically, refer to the structure of BSNN illustrated in Fig. 3, which has P_o+1 layers. The layer 0 is on the receptive field with $N_{\rm r}+P_o$ nodes, to sense the input according to

$$B_l^{({\rm r},0)}(s_{\rm r}) = \begin{cases} 1, & \text{if } U_{l-1} \le s_{\rm r} < U_l, \\ 0, & \text{otherwise,} \end{cases} \quad 1 \le l \le N_{\rm r}+P_o. \tag{27}$$



(a) Upper-bound case when U_{m-1} and U_m are not boundary knots



(b) Case with $U_{m-1} = U_{\min}$ (Case with $U_m = U_{\max}$ is identical)

Fig. 4. Signal flow (complexity) in 1D BSNN with $P_o = 4$, where $\{a, b\}$ beside a node indicates that it requires a additions and b multiplications to compute the basis function value at this node.

The signal is propagated layer by layer for layers $p = 1, \cdots, P_o$ and nodes $l = 1, \cdots, N_{\rm r}+P_o-p$ according to

$$\begin{aligned} B_l^{({\rm r},p)}(s_{\rm r}) = \frac{s_{\rm r}-U_{l-1}}{U_{p+l-1}-U_{l-1}} B_l^{({\rm r},p-1)}(s_{\rm r}) \\ + \frac{U_{p+l}-s_{\rm r}}{U_{p+l}-U_l} B_{l+1}^{({\rm r},p-1)}(s_{\rm r}). \end{aligned} \tag{28}$$

To highlight our motivation of adopting the BSNN model, we elaborate the computational complexity and optimal optimal modeling property of BSNN. Since the polynomial model is the most common nonlinear model, we use it as the benchmark. Consider the task of modeling a real-valued continuous nonlinear function $y = f(x)$, $y, x \in \mathbb{R}$. We model this nonlinear function using the following 1D BSNN model:

$$\hat{y} = \sum\nolimits_{i=1}^{N_{\rm r}} a_i B_i(x), \tag{29}$$

where the basis functions $B_i(x) = B_i^{({\rm r},P_o)}(x)$ (see (28)), P_o is the polynomial degree, and a_i are the real-valued weights. As a comparison, we also use the 1D polynomial model with polynomial degree P_o, i.e.,

$$\widehat{y} = \sum\nolimits_{i=1}^{P_o} b_i x^i, \tag{30}$$

with the real-valued weights b_i.

Remark 2: The polynomial model (30) is well-known to have a computational complexity on the order of $\mathsf{O}\big((P_o+1)^2\big)$. On a casual look of Fig. 3 and De Boor recursion (27) and (28), it would seem that the BSNN model (29) has a much higher complexity. However, careful examination reveals that

the complexity of the BSNN model (29) depends only on the polynomial degree P_o, not on $N_{\rm r}$, and the computational complexity of this BSNN model is also $\mathsf{O}\big((P_o+1)^2\big)$, which is similar to that of the polynomial model. This is because given an input $x \in [U_{\min}, U_{\max}]$, there are only P_o+1 basis functions with nonzero values at most. Fig. 4 shows the complexity of generating the B-spline basis function set for $P_o = 4$, which indicates that the total requirements are 26 additions and 38 multiplications at most.

Also the 1D BSNN has significant computational advanatege over other neural network (NN) for 1D space modeling, particularly in training requirements. The 1D BSNN model (29) has a P_o-layer structure, and there are no any trainable parameter in its 'hidden nodes' of the (P_o-1) 'hidden layers'. The signal simply propagates through the network layer by layer to the 'output layer' via De Boor recursion (28) without involving any other network parameter, as illustrated in Fig. 4. Consider a typically NN with P_o layers. Each hidden node in its (P_o-1) hidden layers has number of trainable parameters and a node activation function. When the signal propagates through the NN layer by layer to the output layer, at each hidden node complex operations take place involving the node's parameters and its nonlinear activation function. From the training perspective, the BSNN model contains only N_r output-layer weights, and these parameters enter the model output linearly. The training of the BSNN model is a simply optimization with the closed-form solution. By contrast, a standard NN contains huge number of parameters, and most of them enter the model output nonlinearly. Training such a NN is a highly complex nonlinear and noncovex optimization, involving typically a long gradient-based iterative procedure.

Remark 3: The literature [47], [48], and [49] establish that among all the universe approximators with similar computational complexity for nonlinear continuous functions on the 1D space, the BSNN model (29) has the maximum robustness property. The previous applications involving noisy training input data [28], [51], [52], [53] have demonstrated that because of this property, the BSNN model outperforms other non-robust linear-combining-nonlinear-bases models, including the polynomial model, in terms of modeling accuracy. In other words, the BSNN model has the optimal modeling capability within the class of linear-combining-nonlinear-bases models having similar computational complexity. We use the analysis of [28] to demonstrate this property. Consider using the BSNN model (29) and the polynomial model (30) to model the nonlinear function $y = f(x)$ with a noisy training dataset. Assume that the estimation errors for models' weights are bounded by $\varepsilon_{\max}$. An upper bound of the modeling error by the BSNN model can be established to be

$$\begin{aligned}\big|y-\widehat{y}\big| &= \Big|\sum\nolimits_{i=1}^{N_{\rm r}} a_i B_i(x) - \sum\nolimits_{i=1}^{N_{\rm r}} \widehat{a}_i B_i(x)\Big| \\ &< \varepsilon_{\max}\Big|\sum\nolimits_{i=1}^{N_{\rm r}} B_i(x)\Big| = \varepsilon_{\max},\end{aligned} \tag{31}$$

while the modeling error of the polynomial model is upper bounded according to

$$\big|y-\widehat{y}\big| = \Big|\sum_{i=1}^{P_o} b_i x^i - \sum_{i=1}^{P_o} \widehat{b}_i x^i\Big| < \varepsilon_{\max}\Big|\sum\nolimits_{i=1}^{P_o} x^i\Big|. \tag{32}$$

Clearly, the BSNN model has much better modeling accuracy than the polynomial model. From (31), it is seen that the maximum numerical robustness or the optimal modeling capability of the BSNN model is due to the convexity of its model bases, i.e., $B_i(x) \geq 0$ for $1 \leq i \leq N_{\rm r}$ and $\sum_{i=1}^{N_{\rm r}} B_i(x) = 1$. By contrast, although the polynomial model is also a universe approximator with similar computational complexity, its bases are not convex. Therefore, it does not share this optimal modeling capability. A simple numerical example to verify this robustness analysis can be found in [28].

Another universe approximator with similar computational complexity is the radial basis function (RBF) network. But it does not have this optimal modeling capability because the RBF bases are not convex, namely, they sum up not to 1. On the other hand, the normalized RBF network shares this optimal modeling capability since the normalized RBF bases are all positive and sum up to one. However the normalized RBF network imposes much higher computational complexity than the BSNN. It is self-evident that the BSNN basis set can be interpreted as a complete set of mass probabilities, which is a highly desirable property in modeling. Many DNNs also 'normanlize' response of their output nodes to achieve similar probability like properties.

2) Complex-Valued BSNN Model of Nonlinear HPA: Since the nonlinear HPA is complex-valued, i.e., its input and output are both complex-valued, we adopt a complex-valued BSNN to model it. First, we construct the two sets of 1D B-spline basis functions, $B_r^{(R,P_o)}(s_R)$ for $1 \leq r \leq N_R$ and $B_i^{(I,P_o)}(s_I)$ for $1 \leq i \leq N_I$, on the real and imaginary parts of the complex-valued input $s = s_R + \mathrm{j}s_I$. Then we use the tensor product for these two sets of basis functions [54] to generate a new set of B-spline basis functions $B_{r,i}^{(P_o)}(s) = B_r^{(R,P_o)}(s_R) B_i^{(I,P_o)}(s_I)$, $1 \leq r \leq N_R$ and $1 \leq i \leq N_I$, defined on the complex-valued input $s \in \mathbb{C}$. This allows us to produce the complex-valued BSNN model for the nonlinear HPA[1]:

$$\begin{aligned} z = \Psi(s) &= \sum_{r=1}^{N_R}\sum_{i=1}^{N_I} B_{r,i}^{(P_o)}(s)\theta_{r,i} \\ &= \sum_{r=1}^{N_R}\sum_{i=1}^{N_I} B_r^{(R,P_o)}(s_R) B_i^{(I,P_o)}(s_I)\theta_{r,i}, \end{aligned} \tag{33}$$

with the complex-valued BSNN weights $\theta_{r,i} = \theta_{r,i_R} + \mathrm{j}\,\theta_{r,i_I}$ for $1 \leq r \leq N_R$ and $1 \leq i \leq N_I$. In this way, the nonlinear HPA $\Psi(\cdot)$ is parametrized by the complex-valued weight vector of the BSNN model (33), denoted by

$$\boldsymbol{\theta} = \big[\theta_{1,1}\ \theta_{1,2} \cdots \theta_{r,i} \cdots \theta_{N_R,N_I}\big]^{\rm T} \in \mathbb{C}^{N_B}, \tag{34}$$

where we have $N_B = N_R N_I$. By defining the BSNN basis vector of (33) for given input s as

$$\boldsymbol{b}(s) = \big[B_{1,1}^{(P_o)}(s)\ B_{1,2}^{(P_o)}(s) \cdots B_{r,i}^{(P_o)}(s) \cdots B_{N_R,N_I}^{(P_o)}(s)\big]^{\rm T}, \tag{35}$$

[1]Directly using a 1D BSNN to model both HPA's amplitude response $A(|s|)$ and phase response $\Upsilon(|s|)$ requires a much smaller model with only $\sqrt{N_R N_I}$ weights. However, the parameters of this HPA model enters the nonlinear channel (concatenated MIMO CIR matrix and nonlinear HPA) in a highly nonlinear manner, and the identification of such a nonlinear channel model has to rely on high-complexity, slow-convergence gradient based iterative procedure.

TABLE I
COMPLEXITY COMPARISON OF THE BSNN MODEL (33) AND THE POLYNOMIAL MODEL (37) FOR $P_o = 4$

Computation	Multiplications	Additions
BSNN model upper bound:		
Two sets of 1D basis functions	2×38	2×26
Output of (33)	3×25	2×24
Total	151	100
BSNN model lower bound:		
Two sets of 1D basis functions	2×36	2×25
Output of (33)	3×16	2×15
Total	120	80
Polynomial model:		
Two sets of 1D basis functions	2×4	0
Output of (37)	3×25	2×24
Total	83	48

where $\boldsymbol{b}(s) \in \mathbb{R}^{N_B}$, the BSNN parametrized nonlinear HPA model (33) can be expressed concisely as

$$z = \boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{b}(s). \tag{36}$$

It can be seen that identifying the HPA's nonlinearity $\Psi(\cdot)$ has been turned into estimating the parameter vector $\boldsymbol{\theta}$ of the linear-in-the-parameter BSNN model.

Remark 4: The BSNN is well suited for this identification task, and the structure parameters of the BSNN model can easily be determined. First, since it is sufficient to choose the polynomial degree $P_o = 3$ or 4 for modeling most nonlinearities encountered in the real-world, we can set $P_o = 3$ or 4 for our application. Second, for universal approximation over the finite and known interval $[U_{\min},\ U_{\max}]$, choosing the number of B-spline basis functions $N_{\mathrm{r}} = 10$ to 14 is sufficient for our BSNN model to achieve accurate modeling. Third, the knot sequence (26), which defines the receptive field, can easily be predetermined from the known HPA input signal's range $U_{\min} < s_{\mathrm{r}} < U_{\max}$. Specifically, the two boundary knots are known to be $U_{P_o-1} = U_{\min}$ and $U_{N_{\mathrm{r}}+1} = U_{\max}$, and we can uniformly distribute the $N_{\mathrm{r}}+1-P_o$ internal knots to cover the input region $[U_{\min},\ U_{\max}]$. Since there exists no data outside the input region, the positions of the external knots are not particularly crucial, and we may empirically set them to achieve good extrapolating capability of the BSNN. Furthermore, the knot sequence must be symmetric, because for real-world QAM signals, the distribution of the HPA input signal s_{r} is naturally symmetric.

3) Complex-Valued Polynomial Model of Nonlinear HPA: The polynomial HPA model is typically adopted in the literature [43], [44], [45], [46], which can be expressed as:

$$z = \sum_{r=0}^{P_o} \sum_{i=0}^{P_o} \left(s_R^r s_I^i\right) \theta_{r,i}, \tag{37}$$

which has $(P_o + 1)^2$ complex-valued parameters, where P_o is the polynomial degree. Table I compares the computation complexity of the BSNN model (33) with those of the polynomial model (37), given the polynomial degree $P_o = 4$. It can be seen that the complexity of the BSNN model is no more than twice of the polynomial model, and its complexity is still on the order of $\mathsf{O}\big((P_o + 1)^2\big)$.

B. Identifying Multiuser Nonlinear MIMO-OFDM Uplink

Give K training data samples, $\left\{\boldsymbol{s}_k^{(C)}, \boldsymbol{x}_k^{(C)}\right\}_{k=1}^{K}$, where the training input data are given by

$$\boldsymbol{s}_k^{(C)} = \left[\left(\boldsymbol{s}_{1,k}^{(C)}\right)^{\mathrm{T}} \left(\boldsymbol{s}_{2,k}^{(C)}\right)^{\mathrm{T}} \cdots \left(\boldsymbol{s}_{M,k}^{(C)}\right)^{\mathrm{T}}\right]^{\mathrm{T}}, \tag{38}$$

with $\boldsymbol{s}_{m,k}^{(C)} = \left[s_{m,k}\ s_{m,k-1} \cdots s_{m,k-n_H+1}\right]^{\mathrm{T}}$ for $1 \le m \le M$. Our task is to identify the multiuser MIMO-OFDM CIR matrix $\boldsymbol{G}$ and the transmitter nonlinear HPA model's parameter vector $\boldsymbol{\theta}$. The model outputs $\widehat{x}_{l,k}$ of the concatenated multiuser frequency-selective MIMO-OFDM channel and BSNN parametrized HPA model can be calculated as

$$\begin{aligned} \widehat{x}_{l,k} &= \sum_{m=1}^{M} \sum_{q=0}^{n_H-1} g_{l,m,q} \widehat{z}_{m,k-q} \\ &= \sum_{m=1}^{M} \sum_{q=0}^{n_H-1} g_{l,m,q} \boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{b}\big(s_{m,k-q}\big), \end{aligned} \tag{39}$$

for $1 \le l \le L$ and $1 \le k \le K$. Our task is to acquire $\boldsymbol{G}$ and $\boldsymbol{\theta}$ by minimizing the sum of squared errors over the training dataset given by

$$\mathrm{SSR}(\boldsymbol{G}, \boldsymbol{\theta}) = \sum_{l=1}^{L} \sum_{k=1}^{K} \left|x_{l,k} - \widehat{x}_{l,k}\right|^2. \tag{40}$$

This is a nonlinear optimization problem, since $\boldsymbol{G}$ and $\boldsymbol{\theta}$ to be identified enter the overall multiuser MIMO-OFDM channel model in a product form. We design an accurate and efficient iterative ALS algorithm to jointly estimate $\boldsymbol{G}$ and $\boldsymbol{\theta}$ for this MIMO-OFDM nonlinear channel.

The beauty or essence of our ALS design is to form the two 'linear' regression models for $\boldsymbol{G}$ and $\boldsymbol{\theta}$ alternatively, and each 'linear' model admits a closed-form LS solution. The 'linear' regression model for $\boldsymbol{G}$ can be formed as

$$\boldsymbol{X} = \boldsymbol{G}\boldsymbol{Q} + \boldsymbol{\Xi}. \tag{41}$$

In (41), the desired output matrix $\boldsymbol{X} \in \mathbb{C}^{L \times K}$ is given by

$$\boldsymbol{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,K} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,K} \\ \vdots & \vdots & \cdots & \vdots \\ x_{L,1} & x_{L,2} & \cdots & x_{L,K} \end{bmatrix} = \begin{bmatrix} \widetilde{\boldsymbol{x}}_1^{\mathrm{T}} \\ \widetilde{\boldsymbol{x}}_2^{\mathrm{T}} \\ \vdots \\ \widetilde{\boldsymbol{x}}_L^{\mathrm{T}} \end{bmatrix}, \tag{42}$$

where $\widetilde{\boldsymbol{x}}_l^{\mathrm{T}} = \left[x_{l,1}\ x_{l,2} \cdots x_{l,K}\right]$ for $1 \le l \le L$, and $\boldsymbol{\Xi} \in \mathbb{C}^{L \times K}$ is the corresponding MIMO-OFDM channel AWGN matrix, while $\boldsymbol{Q} \in \mathbb{C}^{(Mn_H) \times K}$ is the associated 'regression' matrix, which depends on $\boldsymbol{\theta}$ and can be expressed as

$$\boldsymbol{Q} = \begin{bmatrix} \widehat{\boldsymbol{z}}_{1,1} & \widehat{\boldsymbol{z}}_{1,2} & \cdots & \widehat{\boldsymbol{z}}_{1,K} \\ \widehat{\boldsymbol{z}}_{2,1} & \widehat{\boldsymbol{z}}_{2,2} & \cdots & \widehat{\boldsymbol{z}}_{2,K} \\ \vdots & \vdots & \cdots & \vdots \\ \widehat{\boldsymbol{z}}_{M,1} & \widehat{\boldsymbol{z}}_{M,2} & \cdots & \widehat{\boldsymbol{z}}_{M,K} \end{bmatrix}. \tag{43}$$

In (43), $\widehat{\boldsymbol{z}}_{m,k} = \left[\widehat{z}_{m,k}\ \widehat{z}_{m,k-1} \cdots \widehat{z}_{m,k-n_H+1}\right]^{\mathrm{T}}$ with

$$\widehat{z}_{m,k-q} = \boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{b}\big(s_{m,k-q}\big),\ 0 \le q \le n_H - 1, 1 \le m \le M. \tag{44}$$

Clearly, given $\boldsymbol{\theta}$, the MIMO channel matrix $\boldsymbol{G}$ can be estimated with the LS algorithm based on the 'linear' regression model (41).

The 'linear' regression model for $\boldsymbol{\theta}$ can be expressed as

$$\widetilde{\boldsymbol{x}}_l = \boldsymbol{P}_l\boldsymbol{\theta} + \boldsymbol{\xi}_l, 1 \le l \le L, \tag{45}$$

where $\boldsymbol{\xi}_l \in \mathbb{C}^K$ is the associated channel AWGN vector, and $\boldsymbol{P}_l \in \mathbb{C}^{K\times N_B}$ is the corresponding 'regression' matrix, which depends on $\boldsymbol{G}$ and can be written as

$$\boldsymbol{P}_l = \begin{bmatrix} \phi_{1,1}^{(l)}(1) & \phi_{1,2}^{(l)}(1) & \cdots & \phi_{N_R,N_I}^{(l)}(1) \\ \phi_{1,1}^{(l)}(2) & \phi_{1,2}^{(l)}(2) & \cdots & \phi_{N_R,N_I}^{(l)}(2) \\ \vdots & \vdots & \cdots & \vdots \\ \phi_{1,1}^{(l)}(K) & \phi_{1,2}^{(l)}(K) & \cdots & \phi_{N_R,N_I}^{(l)}(K) \end{bmatrix}, \tag{46}$$

with

$$\phi_{r,i}^{(l)}(k) = \sum_{m=1}^{M}\sum_{q=0}^{n_H-1} g_{l,m,q} B_{r,i}^{(P_o)}\left(s_{m,k-q}\right), \quad \text{for } 1 \le r \le N_R \text{and } 1 \le i \le N_I. \tag{47}$$

We can aggregate $\widetilde{\boldsymbol{x}}_l$ for $1 \le l \le L$ and define

$$\boldsymbol{x}_{\text{ave}} = \sum_{l=1}^{L}\widetilde{\boldsymbol{x}}_l, \ \boldsymbol{P}_{\text{ave}} = \sum_{l=1}^{L}\boldsymbol{P}_l, \ \boldsymbol{\xi}_{\text{ave}} = \sum_{l=1}^{L}\boldsymbol{\xi}_l, \tag{48}$$

to obtain the aggregated 'linear' regression model for $\boldsymbol{\theta}$:

$$\boldsymbol{x}_{\text{ave}} = \boldsymbol{P}_{\text{ave}}\boldsymbol{\theta} + \boldsymbol{\xi}_{\text{ave}}. \tag{49}$$

Therefore, given $\boldsymbol{G}$, the nonlinear HPA's parameter vector $\boldsymbol{\theta}$ can be obtained by the LS algorithm based on (49).

We now present our iterative ALS procedure for jointly estimating the multiuser MIMO-OFDM CIR matrix $\boldsymbol{G}$ and the nonlinear HPA model's parameter vector $\boldsymbol{\theta}$.

Initialization. Our ALS procedure is initialized with the initial LS estimate of $\boldsymbol{G}$. Specifically, 'ignoring' the transmitter nonlinear HPA, we have the approximate regression model

$$\boldsymbol{X} \approx \boldsymbol{G}\boldsymbol{S} + \boldsymbol{\Xi}. \tag{50}$$

In (50), the regression matrix $\boldsymbol{S} \in \mathbb{C}^{(n_H M)\times K}$ is only depends on the training input data, and is defined as

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{s}_{1,1} & \boldsymbol{s}_{1,2} & \cdots & \boldsymbol{s}_{1,K} \\ \boldsymbol{s}_{2,1} & \boldsymbol{s}_{2,2} & \cdots & \boldsymbol{s}_{2,K} \\ \vdots & \vdots & \cdots & \vdots \\ \boldsymbol{s}_{M,1} & \boldsymbol{s}_{M,2} & \cdots & \boldsymbol{s}_{M,K} \end{bmatrix}, \tag{51}$$

in which $\boldsymbol{s}_{m,k} = \left[s_{m,k} \ s_{m,k-1} \cdots s_{m,k-n_H+1}\right]^{\text{T}}$. From (50), the closed-form initial LS estimate of $\boldsymbol{G}$, denoted as $\widehat{\boldsymbol{G}}^{[0]}$, is readily obtained as

$$\widehat{\boldsymbol{G}}^{[0]} = \boldsymbol{X}\boldsymbol{S}^{\text{H}}\left(\boldsymbol{S}\boldsymbol{S}^{\text{H}}\right)^{-1}. \tag{52}$$

Then we 'normalize' $\widehat{\boldsymbol{G}}^{[0]}$ according to

$$\widehat{\boldsymbol{G}}^{[0]} = \frac{1}{\widehat{g}_{1,1,0}^{[0]}}\widehat{\boldsymbol{G}}^{[0]}. \tag{53}$$

Iterative ALS procedure. For $1 \le I \le I_{\max}$, where $I_{\max}$ is the maximum number of iterations, alternatively carry out the following two LS estimations.

i) *Estimating the HPA parameter vector*. Fix the CIR matrix $\boldsymbol{G}$ in $\boldsymbol{P}_{\text{ave}}$ to $\widehat{\boldsymbol{G}}^{[I-1]}$ to yield the regression matrix $\boldsymbol{P}_{\text{ave}}^{[I]}$, and compute the regularized LS estimate of $\boldsymbol{\theta}$ in closed-form:

$$\boldsymbol{\theta}^{[I]} = \left(\left(\boldsymbol{P}_{\text{ave}}^{[I]}\right)^{\text{H}}\boldsymbol{P}_{\text{ave}}^{[I]} + \lambda\boldsymbol{I}_{N_B}\right)^{-1}\left(\boldsymbol{P}_{\text{ave}}^{[I]}\right)^{\text{H}}\boldsymbol{x}_{\text{ave}}, \tag{54}$$

where the regularization parameter λ is set to a very small positive value, e.g., $\lambda = 10^{-6}$.

ii) *Estimating the MIMO-OFDM CIR matrix*. Set the BSNN based HPA model parameter $\boldsymbol{\theta}$ in $\boldsymbol{Q}$ to the closed-form solution $\boldsymbol{\theta}^{[I]}$ of (54) to yield the regression matrix $\boldsymbol{Q}^{[I]}$, calculate the closed-form LS estimate of $\boldsymbol{G}$:

$$\widehat{\boldsymbol{G}}^{[I]} = \boldsymbol{X}\left(\boldsymbol{Q}^{[I]}\right)^{\text{H}}\left(\boldsymbol{Q}^{[I]}\left(\boldsymbol{Q}^{[I]}\right)^{\text{H}}\right)^{-1}, \tag{55}$$

and normalize this LS estimate according to

$$\widehat{\boldsymbol{G}}^{[I]} = \frac{1}{\widehat{g}_{1,1,0}^{[I]}}\widehat{\boldsymbol{G}}^{[I]}. \tag{56}$$

Algorithm 1 ALS Estimator for Multiuser Nonlinear MIMO-OFDM Uplink

Require: Training data $\left\{\boldsymbol{s}_k^{(C)}, \boldsymbol{x}_k^{(C)}\right\}_{k=1}^{K}$, and maximum number of iterations $I_{\max}$

Ensure: MIMO channel matrix $\widehat{\boldsymbol{G}}^{[I_{\max}]}$, and HPA's mapping $\boldsymbol{\theta}^{[I_{\max}]}$

1: Construct desired output matrix $\boldsymbol{X}$ (42) and regression matrix $\boldsymbol{S}$ (51)
2: Compute LS estimate $\widehat{\boldsymbol{G}}^{[0]}$ according to (52) and (53)
3: **for** $1 \le I \le I_{\max}$ **do**
4: Set $\boldsymbol{G}$ in $\boldsymbol{P}_{\text{ave}}$ to $\widehat{\boldsymbol{G}}^{[I-1]}$ to yield $\boldsymbol{P}_{\text{ave}}^{[I]}$
5: Compute regularized LS estimate $\boldsymbol{\theta}^{[I]}$ of (54)
6: Set $\boldsymbol{\theta} = \boldsymbol{\theta}^{[I]}$ in $\boldsymbol{Q}$ to obtain regression matrix $\boldsymbol{Q}^{[I]}$
7: Compute LS estimate $\widehat{\boldsymbol{G}}^{[I]}$ according to (55) and (56)
8: **end for**
9: **return** $\widehat{\boldsymbol{G}}^{[I_{\max}]}$ and $\boldsymbol{\theta}^{[I_{\max}]}$

Algorithm 1 summarizes this ALS estimator for the multiuser nonlinear frequency-selective MIMO-OFDM channel.

Remark 5: This ALS procedure is very different from the ALS estimator of [32]. Since the work [32] parametrizes the **single-carrier** based multiuser nonlinear frequency-selective MIMO channel with **three sets** of parameters and these three sets of parameters enter the concatenated model in a triple product form, the ALS estimator of [32] requires the **two-loop** iterative procedure involving **three steps** of LS estimates. In this paper, we parametrize the **OFDM** based multiuser nonlinear frequency-selective MIMO channel with only **two sets** of parameters, yielding less total number of parameters. This also enables us to design the **single-loop** iterative procedure involving only **two steps** of LS estimates (see Algorithm 1). In other words, our ALS estimator imposes much less computational complexity than the one proposed in [32] (although two estimators deal with different systems).

We will demonstrate that our iterative ALS estimator converges extremely fast and attains highly accurate estimates of

$\boldsymbol{G}$ and $\boldsymbol{\theta}$ with no more than two iterations in the simulation study. Therefore, our ALS estimator involves no more than $(1+2)=3$ closed-form LS estimates for $\boldsymbol{G}$ and 2 closed-form LS estimates for $\boldsymbol{\theta}$. This complexity is the lowest compared with any other NN based nonlinear estimator for identifying the multiuser nonlinear frequency-selective MIMO uplink channel considered in this paper. Our ALS estimator also imposes the lowest training pilots, since number of parameters for any other NN based nonlinear estimator is typically several orders of magnitude more than our estimator.

C. Identifying Nonlinear HPA's Inverse Mapping

Define the true inverse mapping of the nonlinear HPA as

$$s = \Psi^{-1}(z) = \Phi(z). \tag{57}$$

Based on the optimal modeling capability as discussed in Remark 3, we adopt another complex-valued BSNN to parametrize $\Phi(\cdot)$. Specifically, we introduce the two knots sequences for the real and imaginary parts of the input $z = z_R + \mathrm{j}z_I$, similar to (26), over which the two sets of 1D B-spline basis functions, $\bar{B}_r^{(R,P_o)}(z_R)$ for $1 \le r \le N_R$ and $\bar{B}_i^{(I,P_o)}(z_I)$ for $1 \le i \le N_Y$, are calculated using De Boor recursion (27) and (28). Then similar to (33), the inverting BSNN model of the transmitter nonlinear HPA is readily formed as

$$\begin{aligned}\widehat{s} = \widehat{\Phi}(z) &= \sum_{r=1}^{N_R}\sum_{i=1}^{N_I} \bar{B}_{r,i}^{(P_o)}(z)\alpha_{r,i} \\ &= \sum_{r=1}^{N_R}\sum_{i=1}^{N_I} \bar{B}_r^{(R,P_o)}(z_R)\bar{B}_i^{(I,P_o)}(z_I)\alpha_{r,i} = \boldsymbol{\alpha}^{\mathrm{T}}\bar{\boldsymbol{b}}(z),\end{aligned} \tag{58}$$

where the parameter vector of this inverting model is given by

$$\boldsymbol{\alpha} = \left[\alpha_{1,1}\ \alpha_{1,2}\cdots\alpha_{r,i}\cdots\alpha_{N_R,N_I}\right]^{\mathrm{T}} \in \mathbb{C}^{N_B}, \tag{59}$$

and $\bar{\boldsymbol{b}}(z) = \left[\bar{B}_{1,1}^{(P_o)}(z)\ \bar{B}_{1,2}^{(P_o)}(z)\cdots\bar{B}_{r,i}^{(P_o)}(z)\cdots\bar{B}_{N_R,N_I}^{(P_o)}(z)\right]^{\mathrm{T}}$ is the basis vector of the inverting BSNN model for input z.

Standard estimation procedure would require the training dataset $\{z_{m,k}, s_{m,k}\}_{k=1}^{K}$ for $1 \le m \le M$. However, the input $z_{m,k}$ of this training set is unavailable, since it is unobserved at the receiver. To get around this difficulty, we calculate $\widehat{\bar{z}}_{m,k} = \widehat{\Psi}(s_{m,k})$ based on the estimated HPA's nonlinearity $\widehat{\Psi}(\cdot)$. This allows us to construct the substituting training dataset of $\{\widehat{\bar{z}}_{m,k}, s_{m,k}\}_{k=1}^{K}$, $1 \le m \le M$. Next, we average the regression matrices

$$\widetilde{\boldsymbol{B}}_m = \begin{bmatrix} \bar{B}_{1,1}^{(P_o)}(\widehat{\bar{z}}_{m,1}) & \cdots & \bar{B}_{N_R,N_I}^{(P_o)}(\widehat{\bar{z}}_{m,1}) \\ \bar{B}_{1,1}^{(P_o)}(\widehat{\bar{z}}_{m,2}) & \cdots & \bar{B}_{N_R,N_I}^{(P_o)}(\widehat{\bar{z}}_{m,2}) \\ \vdots & \cdots & \vdots \\ \bar{B}_{1,1}^{(P_o)}(\widehat{\bar{z}}_{m,K}) & \cdots & \bar{B}_{N_R,N_I}^{(P_o)}(\widehat{\bar{z}}_{m,K}) \end{bmatrix}, \tag{60}$$

for $1 \le m \le M$ to generate the aggregated regression matrix of $\widetilde{\boldsymbol{B}}_{\mathrm{ave}} = \sum_{m=1}^{M} \widetilde{\boldsymbol{B}}_m \in \mathbb{R}^{K \times N_B}$. Similarly, we form the aggregated desired output vector $\widetilde{\boldsymbol{s}}_{\mathrm{ave}} = \sum_{m=1}^{M} \widetilde{\boldsymbol{s}}_m$, where

$$\widetilde{\boldsymbol{s}}_m = \left[s_{m,1}\ s_{m,2}\cdots s_{m,K}\right]^{\mathrm{T}} \in \mathbb{C}^{K}. \tag{61}$$

Then the closed-form LS estimate of $\boldsymbol{\alpha}$ is readily given by

$$\widehat{\boldsymbol{\alpha}} = \left(\widetilde{\boldsymbol{B}}_{\mathrm{ave}}^{\mathrm{T}}\widetilde{\boldsymbol{B}}_{\mathrm{ave}}\right)^{-1}\widetilde{\boldsymbol{B}}_{\mathrm{ave}}^{\mathrm{T}}\widetilde{\boldsymbol{s}}_{\mathrm{ave}}. \tag{62}$$

D. Polynomial Model Based Multiuser MIMO-OFDM Nonlinear Receiver

Since the polynomial HPA model is widely adopted in the literature [43], [44], [45], [46], we also implement a polynomial model based multiuser MIMO-OFDM nonlinear receiver as the benchmark by replacing the BSNN based HPA model of Subsection III-A with the polynomial based HPA model (37). The identification algorithm of Subsection III-B is equally applicable to identify this polynomial based multiuser MIMO-OFDM nonlinear channel model. As mentioned in Remark 2, the computational complexity of the BSNN HPA model (33) is similar to that of the polynomial HPA model (37) [51]. Therefore, the complexity of the identification algorithm for the polynomial based multiuser MIMO-OFDM nonlinear channel model is similar to that of the identification algorithm for the BSNN based multiuser MIMO-OFDM nonlinear channel model.

Likewise, we replace the BSNN based HPA inverse model (58) of Subsection III-C by the following polynomial based HPA inverse model:

$$\widehat{s} = \sum_{r=0}^{P_o}\sum_{i=0}^{P_o} \left(z_R^r z_I^i\right)\alpha_{r,i}. \tag{63}$$

Obviously, the identification algorithm of Subsection III-C remains applicable. Clearly, the computational complexity of the proposed BSNN based multiuser MIMO-OFDM nonlinear receiver is similar to the complexity of this polynomial based multiuser MIMO-OFDM nonlinear receiver.

The reason we do not consider any other NN or DNN based approach as benchmark is that no existing NN or DNN based approach in the literature can accurately estimate the multiuser frequency-selective MIMO-OFDM nonlinear channel at low pilot and computational overheads as our BSNN based design. Any other NN or DNN based design would required several orders of magnitude more pilot and computational overheads, which would preclude it from practical implementation. Moreover, the nonlinear channel estimate acquired by such a NN or DNN estimator cannot be utilized for implementing low-complex MUD of Subsection II-B. Instead, it can only be utilized to train a DNN-based MUD, which is computationally prohibitive for the multiuser frequency-selective MIMO-OFDM nonlinear uplink considered in this paper.

Remark 6: The performance of our BSNN based multiuser MIMO-OFDM nonlinear receiver is much better than the polynomial based benchmark for two reasons. Firstly, as discussed in Remark 1, the BSNN HPA model is much more accurate than the polynomial HPA model. Secondly and more importantly, the training input $\widehat{\bar{z}}_{m,k}$ for the LS estimator of the inverse HPA model is noisy, since it is calculated using the estimated HPA $\widehat{\Psi}(\cdot)$. From basic estimation theory, the estimate is biased when the training input is noisy. As elaborated in Remark 3, the BSNN model has the maximum robustness property, which is useful in minimizing this bias.

TABLE II
KNOT SEQUENCES FOR THE BSNN BASED NONLINEAR RECEIVER

Symmetric and identical knot sequence for s_R and s_I																
-0.65,	-0.6,	-0.55,	**-0.5**,	-0.4,	-0.3,	-0.2,	-0.1,	0,	0.1	0.2,	0.3,	0.4,	**0.5**,	0.55,	0.6,	0.65
Symmetric and identical knot sequence for z_R and z_I																
-5.0,	-4.0,	-3.0,	**-2.5**,	-2.0,	-1.5,	-1.0,	-0.5	0,	0.5	1.0,	1.5,	2.0,	**2.5**,	3.0,	4.0	5.0

E. Complexity of BSNN Based MUD

We quantify the complexity of the MUD of Subsection II-B with the BSNN HPA inversion. To detect the M users' data streams, each containing N symbols, the process involves N linear MUD per subcarrier (18), followed by MN-point IFFT, MN BSNN inversion (23), and MN-point FFT. M-users per subcarrier MUD complexity is $\mathsf{O}(ML)$, where L is the number of BS antennas, and the total complexity of the linear MUD part is $\mathsf{O}(MNL)$. Each N-point IFFT/FFT has the complexity of $\mathsf{O}(\log_2(N)N)$, and the total complexity of IFFF and FFT operation is $\mathsf{O}(MN2\log_2(N))$. The complexity of a BSNN inversion is $\mathsf{O}((P_o+1)^2)$, where P_o is the polynomial degree, and the total complexity to compensate HPA nonlinear distortion is $\mathsf{O}(MN(P_o+1)^2)$. The complexity of detecting MN data symbols by this nonlinear MUD is therefore

$$\mathsf{C}_{\text{total}} = \mathsf{O}\left(MN\left(L + 2\log_2(N) + (P_o+1)^2\right)\right), \tag{64}$$

which corresponds to the complexity per symbol of

$$\mathsf{C}_{\text{per symbol}} = \mathsf{O}\left(L + 2\log_2(N) + (P_o+1)^2\right). \tag{65}$$

IV. SIMULATION RESULTS

A. Experiment Setup

We simulate a multiuser MIMO-OFDM uplink system,[2] which consists of a BS equipped with $L=5$ receive antennas to serve $M=3$ single-antenna users. The modulation scheme is 64-QAM. Each mobile user's transmitter is equipped with the HPA specified by (9) and (10), whose parameters are given in (15). We define the signal-to-noise ratio (SNR) of this MIMO-OFDM uplink of M users and L receive antennas as

$$\text{SNR} = M \cdot \sigma_z^2 / L \cdot 2\sigma_\xi^2. \tag{66}$$

The length of each link's CIR is set to $n_H = 3$, and therefore the CP length is chosen to be $N_{\text{cp}} = n_H = 3$. We set the number of subcarriers to $N = 1024$

The structures of the BSNN based HPA model and HPA inversion model are determined according to the guidelines of Remark 4. Specifically, we choose $N_R = N_I = 12$ as the number of 1D B-spline basis functions and set the polynomial degree to $P_o = 4$ for both the BSNN based HPA model and HPA inverse model. The two knot sequences for the BSNN based HPA and inverse HPA models are also empirically determined with the assistance of Remark 4. Fig. 5 plots the distributions of the HPA input s and output $z = \Psi(s)$. It can be seen from Fig. 5 (a) that the real and imaginary parts of the HPA input s is bounded in magnitude by 0.1 for this set of data and the HPA operating status. To take care of extremely high PAPR signals, whose peak signals' amplitudes can well reach into the HPA's saturation region as illustrated in Fig. 1, we set the two boundary knots to -0.5 and 0.5 for the BSNN based HPA model. From Fig. 5 (b), it is seen that the real and imaginary parts of the HPA input z is bounded by 1.2, which meets the physic limit of the HPA's saturation output magnitude $A_{\text{sat}} = 1.4$. Considering that the input to the HPA inversion model is the estimated HPA output $\widehat{z} = \widehat{\Psi}(s)$, which contains estimation errors, where $\widehat{\Psi}(\cdot)$ is the estimated BSNN HPA model, we set the two boundary knots to -2.5 and 2.5 for the BSNN based inverse model. The resulting two knot sequences for the BSNN based HPA and inverse HPA models are listed in Table II. For a fair comparison, the polynomial HPA model (37) and polynomial inversion HPA model (63) have the polynomial degree of $P_o = 4$. The training set consists of $K = N$ data samples, and we set the number of iterations for Algorithm 1 to $I_{\max} = 2$.

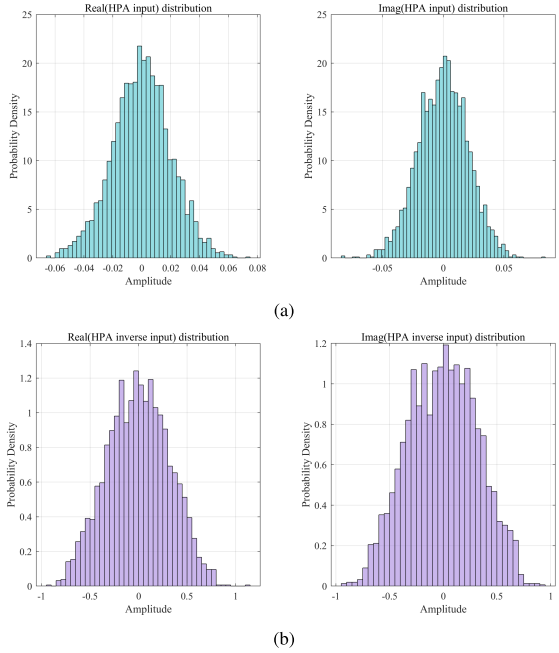


Fig. 5. Distributions of (a) the real and imaginary parts of HPA input, and (b) the real and imaginary parts of HPA output.

B. Nonlinear MIMO Channel Estimation Results

We repeat the experiments for $N_{\text{run}} = 100$ MIMO CIR matrix realizations. In each realization, the true MIMO CIR matrix $\boldsymbol{G}$ is randomly generated according to $g_{l,m,q} \sim \mathcal{CN}(0,1)$. The estimation results are averaged over N_{run} runs. Two OBO values of 3 dB and 5 dB are used in the simulations.

1) Evaluation Metrics: The test performance of an estimator for the MIMO-OFDM channel estimate $\widehat{\boldsymbol{G}}$ can be

[2]The simulation code is available on request to the author Dr Pengyu Wang.

quantified by the normalized root mean square error (NRMSE) metric, which is calculated according to

$$\text{NRMSE}_{\text{channel}} = \frac{\|\boldsymbol{G} - \widehat{\boldsymbol{G}}\|_F}{\|\boldsymbol{G}\|_F} \text{ [dB]}, \tag{67}$$

where $\boldsymbol{G}$ denotes the true MIMO-OFDM CIR matrix.

The NRMSE of the HPA fitting errors over the training input set, $\{s_{m,k}\}_{k=1}^{K}$ for $1 \le m \le M$, is calculated according to

$$\begin{aligned}\text{NRMSE}_{\text{HPA}} &= \frac{1}{K}\sum_{k=1}^{K}\sqrt{\frac{\sum_{m=1}^{M}|z_{m,k}-\widehat{z}_{m,k}|^2}{\sum_{m=1}^{M}|z_{m,k}|^2}} \\ &= \frac{1}{K}\sum_{k=1}^{K}\sqrt{\frac{\sum_{m=1}^{M}\left|\Psi(s_{m,k})-\widehat{\Psi}(s_{m,k})\right|^2}{\sum_{m=1}^{M}|\Psi(s_{m,k})|^2}} \text{ [dB]},\end{aligned} \tag{68}$$

where $z_{m,k} = \Psi(s_{m,k})$ is the true HPA output and $\widehat{z}_{m,k} = \widehat{\Psi}(s_{m,k})$ is the output of the estimated HPA, for given input $s_{m,k}$. This metric measures the training performance of a HPA estimator $\widehat{\Psi}(\cdot)$. To evaluate the test or generalization performance, we also compare the estimated HPA response $\widehat{\Psi}(\cdot)$ with the true HPA response $\Psi(\cdot)$.

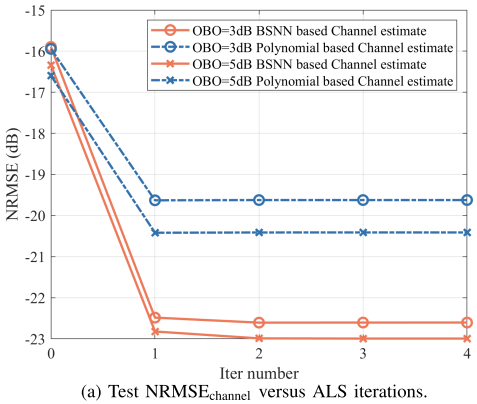
We define the NRMSE of the fitting errors between the ideal concatenated response of the true HPA inversion and the true HPA, i.e., $s = \Phi\big(\Psi(s)\big)$, and the concatenated response of the estimate HPA inversion and the true HPA, i.e., $\widehat{\Phi}\big(\Psi(s)\big)$, over the training input data $\{s_{m,k}\}_{k=1}^{K}$ for $1 \le m \le M$ as:

$$\text{NRMSE}_{\text{IHPA}} = \frac{1}{K}\sum_{k=1}^{K}\sqrt{\frac{\sum_{m=1}^{M}\left|s_{m,k}-\widehat{\Phi}\big(\Psi(s_{m,k})\big)\right|^2}{\sum_{m=1}^{M}|s_{m,k}|^2}} \text{ [dB]}. \tag{69}$$

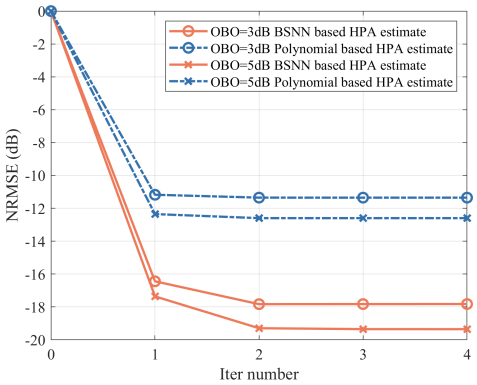
This metric can be used to evaluate the training performance of a HPA inverse estimator $\widehat{\Phi}(\cdot)$. To evaluate the test or generalization performance of the HPA inverse estimator $\widehat{\Phi}(\cdot)$, we also plot the concatenated response of the estimate HPA inversion and the true HPA $\widehat{\Phi}\big(\Psi(s)\big)$ in comparison with the ideal concatenated response, i.e., s.

2) Convergence of ALS Algorithm: To demonstrate the rapid convergence performance of the proposed ALS algorithm for jointly estimating the HPA model parameters $\boldsymbol{\theta}$ and the system's MIMO CIR matrix, Fig. 6 depicts the test NRMSE of the MIMO-OFDM channel estimate $\widehat{\boldsymbol{G}}$ and the training NRMSE of the HPA fitting errors as the function of the number of ALS iterations for the both BSNN and polynomial based models, given SNR = 20 dB. It can be seen from Fig. 6 that the convergence of the ALS estimator is achieved within $I_{\max} = 2$ iterations. Therefore, in all the simulations, we set $I_{\max} = 2$.

3) Estimation of MIMO CIR Matrix: The achievable test NRMSE performance of the MIMO-OFDM channel estimates acquired by the BSNN and polynomial assisted models are compared in Fig. 7. The results of Fig. 7 convincingly demonstrate that the estimation accuracy of the MIMO CIR matrix based on the BSNN model is considerably better than that based on the polynomial model. As expected, the higher the OBO is, the better the estimation accuracy can be achieved.



(a) Test $\text{NRMSE}_{\text{channel}}$ versus ALS iterations.



(b) Training $\text{NRMSE}_{\text{HPA}}$ versus ALS iterations.

Fig. 6. Convergence of the proposed ALS estimator for the both BSNN and polynomial based models: (a) the test NRMSE of the MIMO-OFDM channel estimate, and (b) the training NRMSE of the HPA fitting errors, as the function of ALS iterations, given two OBO values and SNR = 20 dB.

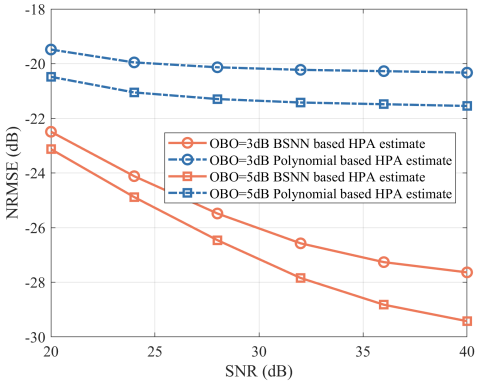


Fig. 7. The $\text{NRMSE}_{\text{channel}}$ test performance as the function of the SNR for the both BSNN and polynomial based models, given two OBO values.

4) Estimation of HPA: The training NRMSE performance of the HPA fitting errors by the BSNN and polynomial based models are compared in Fig. 8, which demonstrates that the estimation accuracy of the nonlinear HPA by the BSNN model is considerably better than that by the polynomial model. The results of Fig. 6 and Fig. 8 are consistent with the well known optimal modeling capability of the BSNN model as elaborated in Remark 3. Fig. 8 indicates that the SNR has negligible impact on the polynomial model estimation accuracy. This is because the modeling error between the true HPA and the

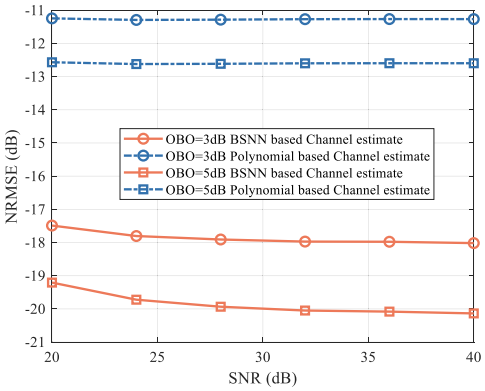


Fig. 8. The $\text{NRMSE}_{\text{HPA}}$ training performance as the function of the SNR for the both BSNN and polynomial based models, given two OBO values.

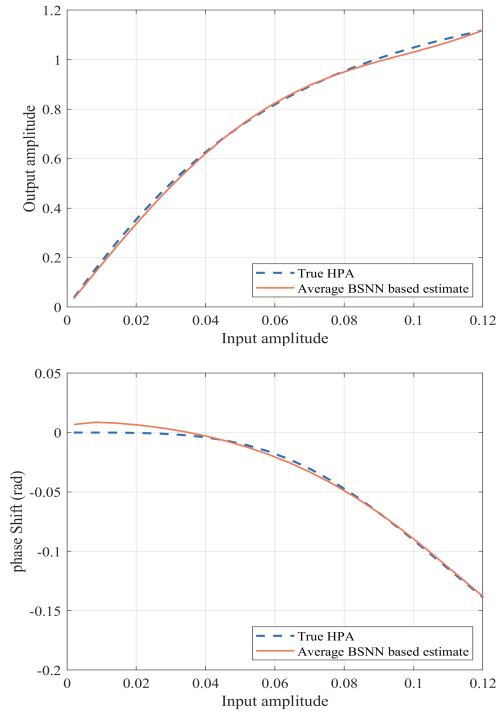


Fig. 9. The BSNN estimated HPA nonlinear response $\widehat{\Psi}(\cdot)$ in comparison with the true HPA's nonlinear response $\Psi(\cdot)$. The OBO is 3 dB and SNR is 20 dB.

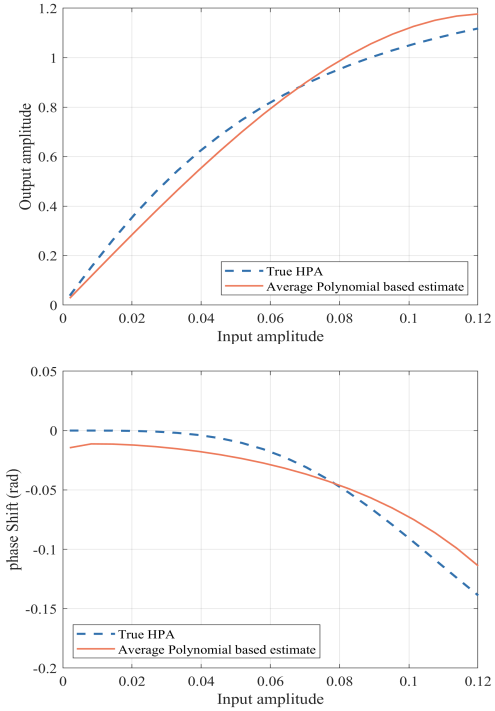


Fig. 10. The polynomial estimated HPA nonlinear response $\widehat{\Psi}(\cdot)$ in comparison with the true HPA's nonlinear response $\Psi(\cdot)$. The OBO is 3 dB and SNR is 20 dB.

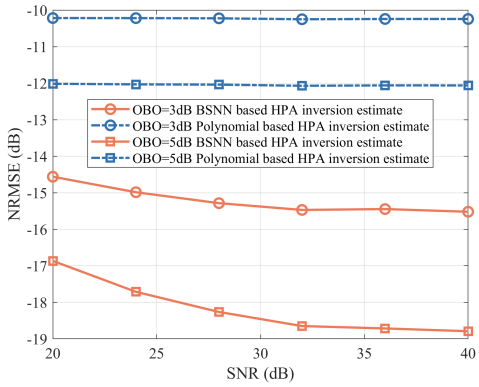


Fig. 11. The $\text{NRMSE}_{\text{IHPA}}$ training performance as the function of the SNR for the both BSNN and polynomial based inverse models, given two OBO values.

polynomial HPA model is considerably bigger than the noise in the received signal, and it dominates the estimation accuracy.

Figs. 9 and 10 compare the true HPA's nonlinear response with the BSNN estimated HPA nonlinear response and the polynomial estimated HPA nonlinear response, respectively, given OBO = 3 dB and SNR = 20 dB. For space economy, we omit the case of OBO = 5 dB, which show the similar trends to Figs. 9 and 10. It is clear that the BSNN based HPA model is much more accurate than the polynomial based HPA model.

5) Estimation of Inverse HPA: The achievable training NRMSE performance of the estimated inverse HPA fitting errors by the BSNN and polynomial based inverse models are compared in Fig. 11, where it is self-evidence that the estimation accuracy of the BSNN based inverse model is considerably better than that of the polynomial based inverse model. This is again consistent with the well known optimal modeling capability of the BSNN model that is also elaborated in Remark 3 and Remark 6. The results also indicate that the SNR has almost no impact on the training estimation accuracy of the polynomial HPA inverse model for the same reason as the polynomial HPA model.

Figs. 12 and 13 compare the concatenated response of the true HPA and the estimated HPA inversions with the ideal concatenated response for both the BSNN inversion model and the polynomial inversion model, respectively, given OBO = 3 dB and SNR = 20 dB. Again the results for OBO = 5 dB are omitted since they show the similar trends to Figs. 12 and 13. The results clearly demonstrate that the BSNN inversion model is much closer to the true HPA inversion than the polynomial inversion model.

The estimation results presented here clearly demonstrate that the proposed BSNN based receiver is highly accurate and

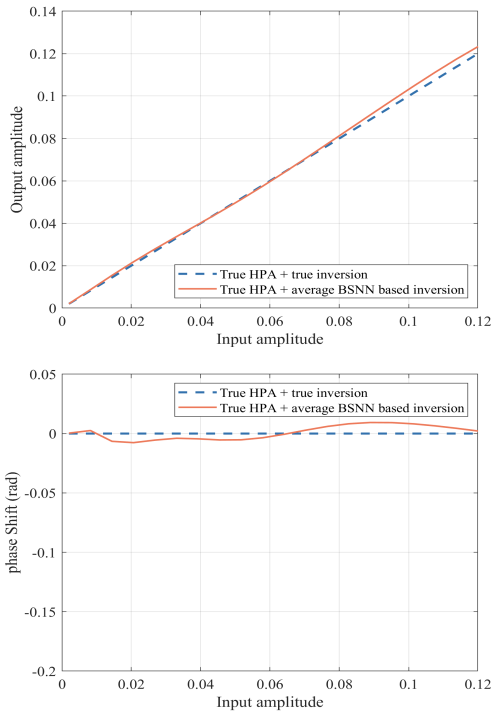


Fig. 12. Comparison of the estimated concatenated response $\widehat{\Psi}^{-1}(\Psi(\cdot))$ with the ideal concatenated response $\Psi^{-1}(\Psi(\cdot))$, where $\Psi(\cdot)$ is the true HPA, $\widehat{\Psi}^{-1}(\cdot)$ is the BSNN HPA inversion, and $\Psi^{-1}(\cdot)$ is the true HPA inversion. The OBO is 3 dB and SNR is 20 dB.

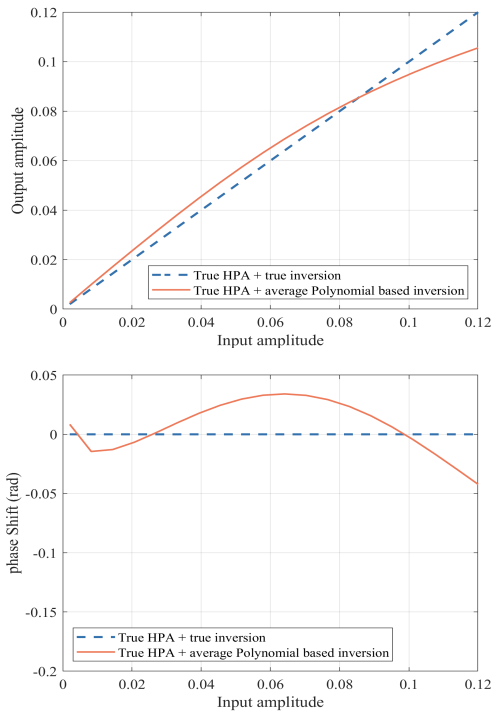


Fig. 13. Comparison of the estimated concatenated response $\widehat{\Psi}^{-1}(\Psi(\cdot))$ with the ideal concatenated response $\Psi^{-1}(\Psi(\cdot))$, where $\Psi(\cdot)$ is the true HPA, $\widehat{\Psi}^{-1}(\cdot)$ is the polynomial HPA inversion, and $\Psi^{-1}(\cdot)$ is the true HPA inversion. The OBO is 3 dB and SNR is 20 dB.

very efficient for joint identification of the multiuser MIMO-OFDM channel and the transmitters' nonlinear HPA mapping as well as for estimating the inverse nonlinear HPA model.

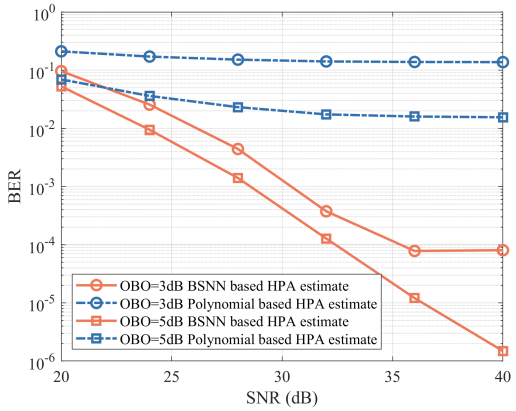


Fig. 14. Average BER performance comparison between the proposed BSNN based MUD approach and the traditional polynomial based MUD approach for the simulated multiuser frequency-selective MIMO-OFDM uplink, given two values of OBO.

C. Multiuser Detection Results

The achievable user BER is the ultimate performance metric for the considered multiuser MIMO-OFDM nonlinear uplink. Given a specific CIR realization, different users will attain different levels of BER. However, since we evaluate the performance over 100 random channel realizations, all the users will achieve similar average BER performance. Therefore, we can consider the average BER over all the three users. Fig. 14 compares the average BER of the BSNN based MUD approach with that of the polynomial based MUD approach. The results of Fig. 14 convincingly demonstrate that our BSNN assisted MUD approach attains much better BER performance than the traditional polynomial assisted MUD approach. Owing to inferior nonlinear channel estimation accuracy and inaccurate HPA inversion model, the polynomial based MUD exhibits very high BER floor. Also observe that for OBO $= 3$ dB, the BSNN based MUD exhibits an BER floor below 10^{-4} for SNR > 36 dB, which is due to the estimation residual error.

V. Conclusion

Owing to the extremely high PAPR of its transmission signals, the multiuser frequency-selective MIMO-OFDM uplink is severely nonlinear. This imposes a considerable challenge for the BS receiver. To overcome this difficulty, we have proposed an original BSNN assisted MUD scheme for this multiuser nonliner frequency-selective MIMO-OFDM uplink. Specifically, by exploiting the optimal modeling capability of BSNN, we have designed a highly effective and efficient BSNN based approach to simultaneously identify the multiuser MIMO-OFDM channel and the nonlinear HPA at transmitters together with the construction of the BSNN inversion model for the transmitters' nonlinear HPA. Based on the joint identification results, the BS can effectively implement low-complexity MUD to mitigate the multiuser and MIMO channel interference and compensate for the nonlinear distortions caused at the transmitters' HPA. Simulation results have confirmed that our BSNN assisted scheme significantly outperforms the traditional polynomial based scheme, and it

offers a state-of-the-art receiver design for multiuser nonliner frequency-selective MIMO-OFDM uplink.

ACKNOWLEDGMENT

The authors, therefore, acknowledge with thanks DSR technical and financial support.

REFERENCES

[1] L. Hanzo et al., *OFDM and MC-CDMA for Broadband Multi-User Communications, WLANs, and Broadcasting*. Chichester, U.K.: Wiley, 2003.

[2] L. Hanzo et al., *Quadrature Amplitude Modulation: From Basics to Adaptive Trellis-Coded, Turbo-Equalised and Space-Time Coded OFDM, CDMA and MC-CDMA Systems*. Chichester, U.K.: Wiley, 2004.

[3] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2294–2323, 3rd Quart., 2018.

[4] M. Y. Alias, A. K. Samingan, S. Chen, and L. Hanzo, "Multiple antenna aided OFDM employing minimum bit error rate multiuser detection," *Electron. Lett.*, vol. 39, no. 24, pp. 1769–1770, Nov. 2003.

[5] S. Chen, L. Hanzo, and A. Livingstone, "MBER space-time decision feedback equalization assisted multiuser detection for multiple antenna aided SDMA systems," *IEEE Trans. Signal Process.*, vol. 54, no. 8, pp. 3090–3098, Aug. 2006.

[6] P. Zhang, S. Chen, and L. Hanzo, "Two-tier channel estimation aided near-capacity MIMO transceivers relying on norm-based joint transmit and receive antenna selection," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 122–137, Jan. 2015.

[7] M. I. Kadir, S. Sugiura, S. Chen, and L. Hanzo, "Unified MIMO-multicarrier designs: A space–time shift keying approach," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 550–579, 2nd Quart., 2015.

[8] S. Gong, S. Wang, S. Chen, C. Xing, and L. Hanzo, "Robust energy efficiency optimization for amplify-and-forward MIMO relaying systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 9, pp. 4326–4343, Sep. 2019.

[9] A. Liao, Z. Gao, H. Wang, S. Chen, M.-S. Alouini, and H. Yin, "Closed-loop sparse channel estimation for wideband millimeter-wave full-dimensional MIMO systems," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8329–8345, Dec. 2019.

[10] M. Wu, Z. Gao, Z. Wang, D. Niyato, G. K. Karagiannidis, and S. Chen, "Deep joint semantic coding and beamforming for near-space airship-borne massive MIMO network," *IEEE J. Sel. Areas Commun.*, vol. 43, no. 1, pp. 260–278, Jan. 2025.

[11] Y. Zeng et al., "CSI-GPT: Integrating generative pre-trained transformer with federated-tuning to acquire downlink massive MIMO channels," *IEEE Trans. Veh. Technol.*, vol. 74, no. 3, pp. 5187–5192, Mar. 2025.

[12] A. A. M. Saleh, "Frequency-independent and frequency-dependent nonlinear models of TWT amplifiers," *IEEE Trans. Commun.*, vol. COM-29, no. 11, pp. 1715–1720, Nov. 1981.

[13] M. Honkanen and S.-G. Haggman, "New aspects on nonlinear power amplifier modeling in radio communication system simulations," in *Proc. 8th Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, vol. 3, Sep. 1997, pp. 844–848.

[14] C. J. Clark, G. Chrisikos, M. S. Muha, A. A. Moulthrop, and C. P. Silva, "Time-domain envelope measurement technique with application to wideband power amplifier modeling," *IEEE Trans. Microw. Theory Techn.*, vol. 46, no. 12, pp. 2531–2540, Dec. 1998.

[15] C.-S. Choi et al., *RF Impairment Models 60 GHz Band SYS/PHY Simulation*, document IEEE 802.15-06-0477-01-003c, Nov. 2006. [Online]. Available: https://mentor.ieee.org/802.15/dcn/06/15-06-0477-01-003c-RF-impairment-models-60ghz-band-sysphy-simulation.pdf

[16] V. Erceg et al., *60 GHz Impairments Modeling*, document IEEE 802.11-09/1213r1, Nov. 2009.

[17] T. Jiang, Y. Yang, and Y.-H. Song, "Exponential companding technique for PAPR reduction in OFDM systems," *IEEE Trans. Broadcast.*, vol. 51, no. 2, pp. 244–248, Jun. 2005.

[18] S. Y. Le Goff, S. S. Al-Samahi, B. K. Khoo, C. C. Tsimenidis, and B. S. Sharif, "Selected mapping without side information for PAPR reduction in OFDM," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3320–3325, Jul. 2009.

[19] Y.-C. Wang and Z.-Q. Luo, "Optimized iterative clipping and filtering for PAPR reduction of OFDM signals," *IEEE Trans. Commun.*, vol. 59, no. 1, pp. 33–37, Jan. 2011.

[20] M. Kim, W. Lee, and D.-H. Cho, "A novel PAPR reduction scheme for OFDM system based on deep learning," *IEEE Commun. Lett.*, vol. 22, no. 3, pp. 510–513, Mar. 2018.

[21] S. Prasad and R. Jayabalan, "PAPR reduction in OFDM systems using modified SLM with different phase sequences," *Wireless Pers. Commun.*, vol. 110, no. 2, pp. 913–929, Jan. 2020.

[22] S. Elaage, A. Hmamou, M. EL Ghzaoui, and N. Mran, "PAPR reduction techniques optimization-based OFDM signal for wireless communication systems," *Telematics Informat. Rep.*, vol. 14, Jun. 2024, Art. no. 100137.

[23] B. S. D. C. da Silva, V. D. P. Souto, R. D. Souza, and L. L. Mendes, "A survey of PAPR techniques based on machine learning," *Sensors*, vol. 24, no. 6, pp. 1–35, Mar. 2024.

[24] M.-C. Chiu, C.-H. Zeng, and M.-C. Liu, "Predistorter based on frequency domain estimation for compensation of nonlinear distortion in OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 57, no. 2, pp. 882–892, Mar. 2008.

[25] V. P. G. Jimenez, Y. Jabrane, A. G. Armada, B. Ait Es Said, and A. Ait Ouahman, "High power amplifier pre-distorter based on neural-fuzzy systems for OFDM signals," *IEEE Trans. Broadcast.*, vol. 57, no. 1, pp. 149–158, Mar. 2011.

[26] S. Chen, "An efficient predistorter design for compensating nonlinear memory high power amplifiers," *IEEE Trans. Broadcast.*, vol. 57, no. 4, pp. 856–865, Dec. 2011.

[27] S. Chen, X. Hong, Y. Gong, and C. J. Harris, "Digital predistorter design using B-spline neural network and inverse of de boor algorithm," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 60, no. 6, pp. 1584–1594, Jun. 2013.

[28] S. Chen, S. X. Ng, E. Khalaf, A. Morfeq, and N. Alotaibi, "Particle swarm optimization assisted B-spline neural network based predistorter design to enable transmit precoding for nonlinear MIMO downlink," *Neurocomputing*, vol. 458, pp. 336–348, Oct. 2021.

[29] X. Lu et al., "A low-computational-complexity digital predistortion model for wideband power amplifier," *Sensors*, vol. 24, no. 21, pp. 1–13, Oct. 2024.

[30] L. Silva, A. Rathnayake, H. Rezaei, and N. Rajatheva, "Transformer neural network-based behavioral modeling and predistortion for wideband power amplifiers," in *Proc. IEEE 35th Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2024, pp. 1–6.

[31] S. Chen, S. X. Ng, E. F. Khalaf, A. Morfeq, and N. D. Alotaibi, "Multiuser detection for nonlinear MIMO uplink," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 207–219, Jan. 2020.

[32] S. Chen, "B-spline neural network assisted space-time equalization for single-carrier multiuser nonlinear frequency-selective MIMO uplink," *WSEAS Trans. Commun.*, vol. 21, pp. 155–169, Jun. 2022.

[33] E. Yu et al., "Deep learning assisted multiuser MIMO load modulated systems for enhanced downlink mmWave communications," *IEEE Trans. Wireless Commun.*, vol. 23, no. 7, pp. 6750–6764, Jul. 2024.

[34] K. Wang et al., "Knowledge and data dual-driven channel estimation and feedback for ultra-massive MIMO systems under hybrid field beam squint effect," *IEEE Trans. Wireless Commun.*, vol. 23, no. 9, pp. 11240–11259, Sep. 2024.

[35] L. V. Nguyen, N. T. Nguyen, N. H. Tran, M. Juntti, A. L. Swindlehurst, and D. H. N. Nguyen, "Leveraging deep neural networks for massive MIMO data detection," *IEEE Wireless Commun.*, vol. 30, no. 1, pp. 174–180, Feb. 2023.

[36] Q. Hu, F. Gao, H. Zhang, G. Y. Li, and Z. Xu, "Understanding deep MIMO detection," *IEEE Trans. Wireless Commun.*, vol. 22, no. 12, pp. 9626–9639, Dec. 2023.

[37] L. Bai, Q. Zeng, R. Han, J. Choi, and W. Zhang, "Deep learning-based low complexity MIMO detection via partial MAP," *IEEE Trans. Wireless Commun.*, vol. 24, no. 3, pp. 2126–2139, Mar. 2025.

[38] Burera, S. Ahmed, and S. Kim, "Transformer learning-based efficient MIMO detection method," *Phys. Commun.*, vol. 70, Jun. 2025, Art. no. 102637.

[39] H. Ye and L. Liang, "On purely data-driven massive MIMO detectors," *IEEE Trans. Signal Process.*, vol. 73, pp. 3079–3093, 2025.

[40] X. Hong and S. Chen, "A new RBF neural network with boundary value constraints," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 39, no. 1, pp. 298–303, Feb. 2009.

[41] S. Chen, X. Hong, and C. J. Harris, "Grey-box radial basis function modelling," *Neurocomputing*, vol. 74, no. 10, pp. 1564–1571, May 2011.

[42] Y. Meng, E. Rigall, X. Chen, F. Gao, J. Dong, and S. Chen, "Physics-guided generative adversarial networks for sea subsurface temperature prediction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 7, pp. 3357–3370, Jul. 2023.

[43] T. Schenk, *RF Imperfections in High-Rate Wireless Systems: Impact and Digital Compensation*. Cham, Switzerland: Springer, 2008.
[44] C. A. R. Fernandes, J. C. M. Mota, and G. Favier, "Analysis and power diversity-based cancellation of nonlinear distortions in OFDM systems," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3520–3531, Jul. 2012.
[45] Z. Huang, D. He, J. Chen, Z. Wang, and S. Chen, "Autoencoder with fitting network for terahertz wireless communications: A deep learning approach," *China Commun.*, vol. 19, no. 3, pp. 172–180, Mar. 2022.
[46] Y. Zhou, W. Li, and Y. Chen, "Nonlinear modeling of MIMO antenna array power amplifiers based on time-delay neural network," in *Proc. IEEE 15th Int. Conf. ASIC (ASICON)*, Oct. 2023, pp. 1–4.
[47] J. M. Pena, "B-spline and optimal stability," *Math. Comput.*, vol. 66, no. 220, pp. 1555–1560, Oct. 1997.
[48] T. Lyche and J. M. Peña, "Optimally stable multivariate bases," *Adv. Comput. Math.*, vol. 20, nos. 1–3, pp. 149–159, Jan. 2004.
[49] E. Mainar and J. M. Pena, "Optimal stability of bivariate tensor product B-bases," *J. Numer. Anal., Ind. Appl. Math.*, vol. 6, nos. 3–4, pp. 95–104, 2011.
[50] C. De Boor, *A Practical Guide to Splines*. New York, NY, USA: Spring Verlag, 1978.
[51] S. Chen, X. Hong, E. F. Khalaf, F. E. Alsaadi, and C. J. Harris, "Comparative performance of complex-valued B-spline and polynomial models applied to iterative frequency-domain decision feedback equalization of Hammerstein channels," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 12, pp. 2872–2884, Dec. 2017.
[52] S. Chen, X. Hong, J. Gao, and C. J. Harris, "Complex-valued B-spline neural networks for modeling and inverting Hammerstein systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 9, pp. 1673–1685, Sep. 2014.
[53] S. Chen, X. Hong, E. F. Khalaf, A. Morfeq, N. D. Alotaibi, and C. J. Harris, "Single-carrier frequency-domain equalization with hybrid decision feedback equalizer for Hammerstein channels containing nonlinear transmit amplifier," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 3341–3354, May 2017.
[54] C. J. Harris, X. Hong, and Q. Gan, *Adaptive Modelling, Estimation and Fusion from Data: A Neurofuzzy Approach*. Berlin, Germany: Springer-Verlag, 2002.

Sheng Chen (Life Fellow, IEEE) received the B.Eng. degree from East China Petroleum Institute, Dongying, China, in 1982, the Ph.D. degree from the City, University of London, in 1986, both in control engineering, and the D.Sc. degree from the University of Southampton, Southampton, U.K., in 2005.

From 1986 to 1999, he held research and academic appointments at The University of Sheffield, U.K., The University of Edinburgh, U.K., and the University of Portsmouth, U.K. Since 1999, he has been with the School of Electronics and Computer Science, University of Southampton, where he holds the post of Professor of Intelligent Systems and Signal Processing. In 2023, he was appointed as a Distinguished Professor with the Faculty of Information Science and Engineering, Ocean University of China, Qingdao, China. He has published more than 700 research articles. He has more than 22 000 Web of Science citations with an H-index of 65 and more than 42 000 Google Scholar citations with an H-index of 87. His research interests include adaptive signal processing, wireless communications, modeling and identification of nonlinear systems, neural networks, and machine learning. He is a fellow of the United Kingdom Royal Academy of Engineering, Asia-Pacific Artificial Intelligence Association, and IET. He is an IEEE ComSoc Signal Processing and Computing for Communications (SPCC) 2025 Technical Recognition Award recipient. He was one of the original 200 ISI highly cited researchers in engineering in March 2004.

Pengyu Wang (Member, IEEE) received the Ph.D. degree from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2023. He is currently a Post-Doctoral Researcher with the Department of Electronic Engineering, Tsinghua University, Beijing, China. His current research interests include artificial intelligence, cognitive radio, and machine learning in communication.

Mingkun Li received the B.S. degree in space science and technology from Xidian University, Xi'an, China, in 2023. He is currently pursuing the M.S. degree in electronic information engineering with Tsinghua University, Beijing, China. His current research interests include cognitive radio and machine learning in communication.

Emad F. Khalaf received the integrated B.Eng. and M.Eng. degree in computer engineering and the Ph.D. degree in computer networks from Wrocław University of Science and Technology, Poland, in 1992 and 2002, respectively.

From 2003 to 2011, he was an Assistant Professor with the Computer Engineering Department, Faculty of Engineering, Philadelphia University. Since 2012, he has been an Assistant Professor with the Electrical and Computer Engineering Department, Faculty of Engineering, King Abdulaziz University, Jeddah, Saudi Arabia. His research interests include network security, speech classification and recognition, and image and signal processing.

Ali Morfeq received the M.Sc. degree from Oregon State University, Corvallis, OR, USA, in 1985, and the Ph.D. degree from the University of Colorado Boulder, Boulder, CO, USA, in 1990.

Since 2013, he has been the Vice Dean with the Faculty of Engineering, King Abdulaziz University (KAU), where he is currently an Associate Professor. He also served at KAU in several administrative positions and as the Head of many academic and administrative committees. In addition, he was a Consultant to many firms, including big industrial companies and several hospitals. His work involved the automation of office environments and the supporting databases. He had, and still has, many graduate students working in different areas, such as decision support systems, data marts, and data warehouses.

Naif D. Alotaibi received the B.Sc. degree in computer engineering from King Abdulaziz University (KAU), Jeddah, Saudi Arabia, in 2000, and the master's and Ph.D. degrees in information technology from the Queensland University of Technology, Brisbane, QLD, Australia, in 2006 and 2012, respectively.

In 2013, he joined the Electrical and Computer Engineering Department, Faculty of Engineering, KAU, as an Assistant Professor, where he is currently an Associate Professor. His research interests include IT infrastructure, computer network quality, network QoS, and quality measurement.