B-Spline Neural Network Assisted Space-Time Equalization for Single-Carrier Multiuser Nonlinear Frequency-Selective MIMO Uplink

SHENG CHEN
School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK

Abstract—This paper designs a nonlinear space-time equalizer based on B-spline neural network (BSNN) for the singlecarrier high-throughput multiuser frequency-selective multipleinput multiple-output (MIMO) nonlinear uplink. Specifically, based on a BSNN parametrization of the nonlinear high power amplifiers (NHPAs) at mobile terminals' transmitters, a novel nonlinear identification scheme is developed to estimate the nonlinear dispersive MIMO uplink channel, which includes the BSNN models for the NHPAs at transmitters as well as the frequency-selective MIMO channel impulse response (CIR) matrix. Furthermore, the BSNN inverse models of the NHPAs are also estimated in closed-form. This allows the base station to implement nonlinear multiuser detection effectively using the space-time equalization (STE) based on the estimated frequencyselective MIMO CIR matrix and followed by compensating for the nonlinear distortion of the transmitters' NHPAs based on the estimated BSNN inverse models. Simulation results are utilized to demonstrate the superior bit error rate performance of our nonlinear STE approach for single-carrier high-throughput multiuser nonlinear frequency-selective MIMO uplink.

Keywords: Multiple-input multiple-output uplink, frequency-selective channel, nonlinear transmit high power amplifier, B-spline neural network, nonlinear inversion, nonlinear space-time equalization

Received: August 23, 2021. Revised: March 27, 2022. Accepted: April 26, 2022. Published: June 3, 2022.

1. Introduction

The spatial-domain non-orthogonal multiple access (NOMA) based on multi-input multi-output (MIMO) technology is a cornerstone for the fifth-generation (5G) mobile communication system, to support enhanced mobile broadband and massive machine-type communications with the limited frequency-time resources. This naturally leads to multiuser MIMO communication systems. In the literature, most existing MIMO system designs adopt the linear MIMO channel [1]–[10]. The linear MIMO channel assumption however is only valid when the transmitter high power amplifier (HPA) operates within its linear dynamic range. Practical HPAs on the other hand are often nonlinear, as they exhibit nonlinear saturation and phase distortion characteristics [11]–[15]. More specifically, the linear channel assumption critically depends on the transmitted signal's peak-to-average power ratio (PAPR) as well as the average transmission power. For the modulation constellations with unity PAPR, such as phase shift keying (PSK), the phase shift of the HPA's output is constant for all the symbol points. Consequently, the HPA does not cause amplitude distortion in this case, and the MIMO channel is linear. In order to meet high throughput requirement, however, multiuser MIMO systems typically utilize the high-throughput quadrature amplitude modulation (QAM) with multiple bits per symbol [16]. Since high-throughput QAM constellations have high PAPR, the nonlinear distortion of the transmitter HPA may become serious and the linear MIMO channel may no longer be valid. Note that high-throughput QAM transmission is achieved by imposing high average transmission power. Therefore, it is impossible to avoid the nonlinearity of transmitter HPA by using output back-off (OBO), because the OBO required would be too severe, which would be unable to meet the required link power budget.

This paper investigates the challenging single-carrier multiuser MIMO uplink with high-throughput QAM transmission, where transmitters are equipped with nonlinear HPAs (NHPAs). Note that for the single-carrier multiuser nonlinear MIMO downlink, where the base station (BS) transmits to multiple mobile users (MUs), effective solution for overcoming nonlinear distortions of NHPAs readily exists. Specifically, since the BS possesses sufficient computation capacity, it can implement digital predistorter [17]–[23] to pre-compensate for the nonlinear distortions of NHPA in addition to implement the multiuser transmit precoding to compensate for the MIMO channel interference. This leads to our recent design using a B-spline neural network (BSNN) based predistorter for single-carrier multiuser nonlinear MIMO downlink [24]. In uplink, by contrast, it is difficult for a mobile handset to implement the predistorter owning to its limited computation capacity. As a result, the BS receiver must first estimate the multiuser nonlinear MIMO channel and then performs the nonlinear multiuser detection (MUD), which is extremely challenging. In the literature, only few works [25]–[27] attempted to tackle this difficult task by employing the MIMO Volterra model to identify the frequency-selective nonlinear MIMO channel [25]–[27], which not only imposes impossibly heavy computational burden but also is impractical for implementing nonlinear MUD for the uplink with high-throughput QAM transmission. In [28], we proposed a nonlinear MUD scheme for single-carrier multiuser nonlinear MIMO uplink. However, our previous work [28] only considers the MIMO systems with frequency-nonselective or narrowband channels. In practice, MIMO channels are frequency-selective. Hence, practical single-carrier multiuser nonlinear MIMO uplink is much more complex than the case investigated in [28].

In this paper, we develop a BSNN assisted space-time equalization (STE) scheme for this much more challenging

single-carrier multiuser nonlinear frequency-selective MIMO uplink with high-throughput QAM transmission and NHPAs. Our novel contribution is two-fold.

- By extending the work [28] originally derived for multiuser nonlinear narrowband MIMO channel, we develop a BSNN assisted identification scheme to identify the multiuser nonlinear frequency-selective MIMO channel. This includes the BSNN [29]–[36] modeling for the MUs' NHPAs and the estimate of the linear frequency-selective MIMO channel impulse response (CIR) matrix. By exploiting the results of this nonlinear MIMO channel identification, the BSNN inverse models of the transmitters' NHPAs are also identified.
- We implement the minimum mean square error (MMSE) space-time equalizer [3], [4] to combat the interference of the frequency-selective MIMO channel using the estimated linear frequency-selective MIMO CIR matrix. Then we compensate for the nonlinear distortions of the transmitters' NHPAs with the estimated BSNN inverse models for the NHPAs. An extensive simulation study is carried out to demonstrate the excellent bit error rate (BER) performance of our proposed nonlinear STE approach for multiuser nonlinear frequency-selective MIMO uplink with high-throughput QAM transmission and NHPAs.

The rest of this paper is structured as follows. The single-carrier multiuser nonlinear frequency-selective MIMO uplink system is introduced in Section II. This includes the NHPA model at each MU's transmitter and the frequency-selective MIMO channel model, as well as the nonlinear STE based MUD at the BS receiver that first uses a standard space-time equalizer to remove both multiuser interference and self-interference and then removes the nonlinear distortion of the transmitters' NHPAs by the nonlinear inversion of the NHPAs, assuming that both the frequency-selective MIMO CIR matrix and the inverse mappings of transmitters' NHPAs are known at the BS receiver. The proposed BSNN assisted nonlinear STE scheme is detailed in Section III. By utilizing a unique parametrization of the frequency-selective MIMO CIR matrix and the nonlinear transmitters as well as the effective BSNN modeling of the NHPAs, a new iterative alternating least squares (ALS) estimator is developed, which guarantees to attain the unbiased and accurate estimates of the frequency-selective MIMO CIR matrix and the BSNN parametrized NHPAs' models in a few iterations. Based on the nonlinear frequency-selective MIMO channel identification results, the closed-form BSNN inverse models for the NHPAs are also obtained. Section IV is devoted to simulation study, to investigate the effectiveness of our proposed BSNN assisted nonlinear STE scheme for single-carrier multiuser nonlinear frequency-selective MIMO uplink with high-throughput QAM transmission and NHPAs. Our conclusions are offered in Section V.

2. Nonlinear Frequency-selective Mimo Uplink

Fig. 1 depicts the system diagram of the spatial-domain NOMA based single-carrier multiuser nonlinear frequency-selective MIMO uplink, where M single-antenna MUs transmit to the BS equipped with L receive antennas using the same frequency-time resource block. Note that $L > M$.

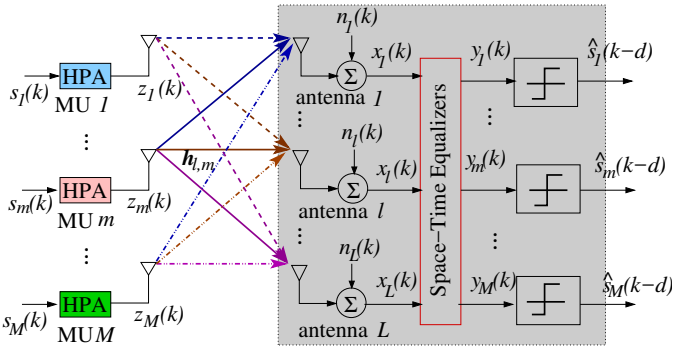

Fig. 1. Spatial-domain NOMA based single-carrier multiuser nonlinear frequency-selective MIMO uplink where the BS is equipped with the L-element antenna array to receive the data from M single-antenna MUs using the same single resource block.

2.1 Channel and transmitter models

Since we consider the wideband or frequency-selective channel, the CIR from the mth mobile to the lth antenna of the BS can be expressed by

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

for $1 \le l \le L$ and $1 \le m \le M$, where for notational simplicity, all the $L \cdot M$ CIRs are assumed to have the same CIR length of n_H. The kth data symbol transmitted by the mth MU is denoted by $s_m(k) = |s_m(k)| \cdot \exp\left(\mathrm{j}\angle s_m(k)\right)$, where $\mathrm{j} = \sqrt{-1}$, $|s_m(k)|$ denotes the amplitude of $s_m(k)$ and $\angle s_m(k)$ is the phase of $s_m(k)$. As we use the U-QAM constellation with $\log_2(U)$ bits per symbol, to enhance the achievable throughput, $s_m(k)$ takes the value from the constellation set:

$$\begin{aligned}\mathbb{S} = \Big\{ & d_{\mathrm{S}}(2l - \sqrt{U} - 1) + \mathrm{j}\, d_{\mathrm{S}}(2q - \sqrt{U} - 1), \\ & 1 \le l, q \le \sqrt{U} \Big\}.\end{aligned} \tag{2}$$

The minimum distance between the symbol points of $\mathbb{S}$ is $2d_{\mathrm{S}}$.

Without loss of generality, the HPAs at all the MUs' transmitters are assumed to be the same type. Hence under the same given operation condition, they exhibit the same nonlinear characteristics. We employ a common and practical HPA, the solid state power amplifier [14], [15]. For this type of NHPA, the transmitted signal of the mth MU, $1 \le m \le M$, can be expressed as

$$\begin{aligned} z_m(k) =& \Psi\left(s_m(k)\right) \\ =& A\big(|s_m(k)|\big) \cdot \exp\Big(\mathrm{j}\big(\angle s_m(k) + \Upsilon(|s_m(k)|)\big)\Big), \end{aligned} \tag{3}$$

where $\Psi(\cdot)$ represents the NHPA at a MU's transmitter, $A(\cdot)$ is its amplitude response and $\Upsilon(\cdot)$ is its phase response. The output $\Psi(s)$ of this type of NHPA is specified by its amplitude response $A(r)$ and phase response $\Upsilon(r)$, where $r = |s|$ denotes the amplitude of the input s to the NHPA. Note that the distortion caused by this type of NHPA depends only on the

amplitude of the NHPA input. According to [14], [15], the amplitude response $A(r)$ is given by

$$A(r) = \frac{g_a r}{\left(1 + \left(\frac{g_a r}{A_{\rm sat}}\right)^{2\beta_a}\right)^{\frac{1}{2\beta_a}}}, \tag{4}$$

while the phase response $\Upsilon(r)$ is defined by

$$\Upsilon(r) = \frac{\alpha_\phi r^{q_1}}{1 + \left(\frac{r}{\beta_\phi}\right)^{q_2}} \ \text{[degree]}, \tag{5}$$

where the parameters g_a, β_a and $A_{\rm sat}$ are known as the small signal's gain, the smoothness factor and the saturation level, respectively, while the parameters α_ϕ, β_ϕ, q_1 and q_2 specify the NHPA's phase response [14], [15]. If we define the maximum output power of the NHPA as $P_{\max} = A_{\rm sat}^2$. Further denote the average output amplitude of the NHPA as $A_{\rm ave}$, which means that the the average output power of the NHPA output signal is $P_{\rm ave} = A_{\rm ave}^2$. Then the operating status of the HPA is defined by the ratio of the maximum output power $P_{\max}$ of the NHPA to the average output power $P_{\rm ave}$ of the NHPA output signal, called the OBO, which is given by

$$\text{OBO} = 10 \cdot \log_{10} \frac{P_{\max}}{P_{\rm ave}} \ \text{[dB]}. \tag{6}$$

Note that the maximum output power $P_{\max}$ is the NHPA's saturated output power. The smaller the OBO is the deeper the NHPA is operating into its saturation region and hence causing more severe nonlinear distortion.

Recall the CIR (1) and denote the mth MU's transmitted signal vector by $\boldsymbol{z}_m^{(h)}(k) = \left[z_m(k)\ z_m(k-1)\cdots z_m(k-n_H+1)\right]^{\rm T}$ with $z_m(k-i) = \Psi(s_m(k-i))$ for $0 \le i \le n_H - 1$. Then the received signal sample $x_l(k)$ at the BS's lth receiver antenna can be expressed by

$$\begin{aligned} x_l(k) &= \sum_{m=1}^{M} \boldsymbol{h}_{l,m}^{\rm T} \boldsymbol{z}_m^{(h)}(k) + n_l(k) \\ &= \sum_{m=1}^{M} \sum_{i=0}^{n_H-1} h_{i,l,m} z_m(k-i) + n_l(k), \end{aligned} \tag{7}$$

where $n_l(k)$ is the complex additive white Gaussian noise (AWGN) with power $\mathsf{E}\left[\left|n_l(k)\right|^2\right] = 2\sigma_n^2$. Collect the received signals $x_l(k)$ for $1 \le l \le L$ as $\boldsymbol{x}^{(h)}(k) = [x_1(k)\ x_2(k)\cdots x_L(k)]^{\rm T}$, which can be expressed as

$$\begin{aligned} \boldsymbol{x}^{(h)}(k) = &\begin{bmatrix} \boldsymbol{h}_{1,1}^{\rm T} & \boldsymbol{h}_{1,2}^{\rm T} & \cdots & \boldsymbol{h}_{1,M}^{\rm T} \\ \boldsymbol{h}_{2,1}^{\rm T} & \boldsymbol{h}_{2,2}^{\rm T} & \cdots & \boldsymbol{h}_{2,M}^{\rm T} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{h}_{L,1}^{\rm T} & \boldsymbol{h}_{L,2}^{\rm T} & \cdots & \boldsymbol{h}_{L,M}^{\rm T} \end{bmatrix} \begin{bmatrix} \boldsymbol{z}_1^{(h)}(k) \\ \boldsymbol{z}_2^{(h)}(k) \\ \vdots \\ \boldsymbol{z}_M^{(h)}(k) \end{bmatrix} \\ &+ \boldsymbol{n}^{(h)}(k) = \boldsymbol{H}\boldsymbol{z}^{(h)}(k) + \boldsymbol{n}^{(h)}(k), \end{aligned} \tag{8}$$

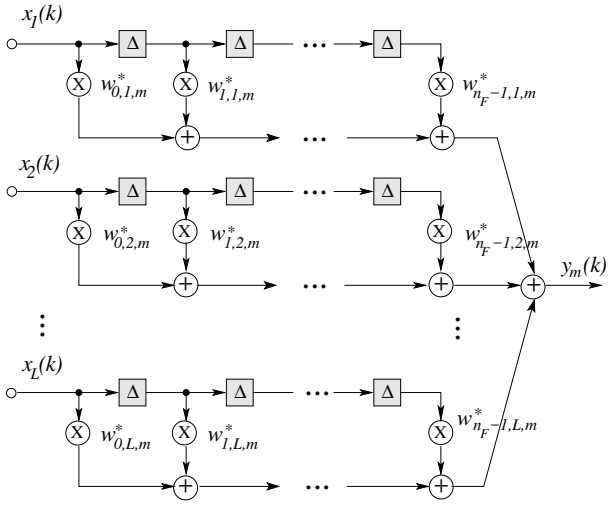

Fig. 2. Space-time equalizer for detecting the mth mobile user's transmitted signal.

in which the AWGN vector $\boldsymbol{n}^{(h)}(k) = [n_1(k)\cdots n_L(k)]^{\rm T}$.

2.2 Receiver multiuser detection

Assume that the BS knows the multiuser MIMO CIR matrix $\boldsymbol{H}$. If all the transmitters' HPAs are operating in the linear regions, the MUD for the MUs' transmitted signals consists of M space-time equalizers [3], [4], one for each mobile, as illustrated in Fig. 2. Specifically, each space-time equalizer has length n_F. Further define the mth space-time equalizer's weight vector associated with the BS's lth receive antenna as $\boldsymbol{w}_{l,m} = \left[w_{0,l,m}\ w_{1,l,m}\cdots w_{n_F-1,l,m}\right]^{\rm T}$, and denote the corresponding space-time equalizer's input signal vector by $\boldsymbol{x}_l(k) = \left[x_l(k)\ x_l(k-1)\cdots x_l(k-n_F+1)\right]^{\rm T}$. Then the output of the mth space-time equalizer is given by

$$\begin{aligned} y_m(k) &= \sum_{l=1}^{L} \boldsymbol{w}_{l,m}^{\rm H} \boldsymbol{x}_l(k) \\ &= \sum_{l=1}^{L} \sum_{i=0}^{n_F-1} w_{i,l,m}^* x_l(k-i),\ 1 \le m \le M. \end{aligned} \tag{9}$$

Since the HPAs are nonlinear, $y_m(k)$ is only a sufficient statistic for detecting $z_m(k-d) = \Psi(s_m(k-d))$, and it is not a sufficient statistic for detecting the transmitted data symbol $s_m(k-d)$, where d is the decision delay.

Based on linear convolution, $\boldsymbol{x}_l(k)$ can be expressed as $\boldsymbol{x}_l(k) = \sum_{m=1}^{M} \boldsymbol{c}_{l,m}\boldsymbol{z}_m$, where the $n_F \times (n_F + n_H - 1)$ CIR matrix $\boldsymbol{c}_{l,m}$ associated with the mth MU transmitter and the lth BS receiver antenna has the structure shown in (10) at the bottom of the previous page, and $\boldsymbol{z}_m(k) = \left[z_m(k)\ z_m(k-1)\cdots z_m(k-n_F-n_H+2)\right]^{\rm T}$, for $1 \le m \le M$.

$$\boldsymbol{c}_{l,m} = \begin{bmatrix} h_{0,l,m} & h_{1,l,m} & \cdots & h_{n_H-1,l,m} & 0 & \cdots & 0 \\ 0 & h_{0,l,m} & h_{1,l,m} & \cdots & h_{n_H-1,l,m} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & h_{0,l,m} & h_{1,l,m} & \cdots & h_{n_H-1,l,m} \end{bmatrix} \tag{10}$$

By defining the overall system CIR convolution matrix as

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{c}_{1,1} & \boldsymbol{c}_{1,2} & \cdots & \boldsymbol{c}_{1,M} \\ \boldsymbol{c}_{2,1} & \boldsymbol{c}_{2,2} & \cdots & \boldsymbol{c}_{2,M} \\ \vdots & \vdots & \cdots & \vdots \\ \boldsymbol{c}_{L,1} & \boldsymbol{c}_{L,2} & \cdots & \boldsymbol{c}_{L,M} \end{bmatrix}, \tag{11}$$

and the MIMO channel input vector as $\boldsymbol{z}(k) = \left[\boldsymbol{z}_1^{\mathrm{T}}(k) \cdots \boldsymbol{z}_M^{\mathrm{T}}(k)\right]^{\mathrm{T}}$, the space-time equalizer's input vector $\boldsymbol{x}(k) = \left[\boldsymbol{x}_1^{\mathrm{T}}(k) \cdots \boldsymbol{x}_L^{\mathrm{T}}(k)\right]^{\mathrm{T}}$ can be expressed as

$$\boldsymbol{x}(k) = \boldsymbol{C}\boldsymbol{z}(k) + \boldsymbol{n}(k), \tag{12}$$

in which the overall noise vector $\boldsymbol{n}(k) = \left[\boldsymbol{n}_1^{\mathrm{T}}(k) \ \boldsymbol{n}_2^{\mathrm{T}}(k) \cdots \boldsymbol{n}_L^{\mathrm{T}}(k)\right]^{\mathrm{T}}$ with $\boldsymbol{n}_l(k) = \left[n_l(k) \ n_l(k-1) \cdots n_l(k-n_F+1)\right]^{\mathrm{T}}$ for $1 \le l \le L$. Further define the overall weight vector of the mth space-time equalizer by $\boldsymbol{w}_m = \left[\boldsymbol{w}_{1,m}^{\mathrm{T}} \ \boldsymbol{w}_{2,m}^{\mathrm{T}} \cdots \boldsymbol{w}_{L,m}^{\mathrm{T}}\right]^{\mathrm{T}}$. The mth space-time equalizer (9) can be expressed concisely as

$$y_m(k) = \boldsymbol{w}_m^{\mathrm{H}} \boldsymbol{x}(k). \tag{13}$$

From [3], we have the following closed-form MMSE solution for $\boldsymbol{w}_m$:

$$\boldsymbol{w}_{(\mathrm{MMSE})m} = \left(\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}} + 2\sigma_n^2 \boldsymbol{I}\right)^{-1} \boldsymbol{C}_{[:(m-1)(n_F+n_H-1)+(d+1)]}, \tag{14}$$

for $1 \le m \le M$, where $\boldsymbol{I}$ is the $(Ln_F) \times (Ln_F)$ identity matrix and $\boldsymbol{C}_{[:i]}$ is the ith column of $\boldsymbol{C}$.

The space-time equalizer (13) provides the estimate $\widehat{z}_m(k-d)$ for $z_m(k-d)$. If the nonlinear inversion $\Psi^{-1}(\cdot)$ of the complex NHPA's nonlinear mapping $\Psi(\cdot)$ is known, the estimate of $s_m(k-d)$ is then given by

$$\widehat{s}_m(k-d) = \Psi^{-1}\left(\widehat{z}_m(k-d)\right). \tag{15}$$

2.3 Unique parametrization of MIMO uplink

It can be seen that in order to detect the MUs' data $s_m(k-d)$, $1 \le m \le M$, the BS needs to acquire the MIMO channel matrix $\boldsymbol{H}$ and to invert the unknown complex nonlinear mapping $\Psi(\cdot)$. This is a very challenging nonlinear estimation and inversion problem. First, the MIMO channel input $\boldsymbol{z}^{(h)}(k)$ is unknown to the receiver, and the BS cannot apply the standard least squares (LS) estimator to identify $\boldsymbol{H}$. Second, the model of the MUs' NHPAs, denoted as $\boldsymbol{z}^{(h)}$, is multiplicative with the MIMO CIR matrix $\boldsymbol{H}$ as the product $\boldsymbol{H} \cdot \boldsymbol{z}^{(h)}$. This implies that there are infinitely many equivalent pairs of the parametrization for the MIMO CIR matrix and the NHPAs' model, which causes ambiguity problem and imposes a significant challenge to the task of identifying both he MIMO CIR matrix and the NHPAs' model. In order to develop a meaningful identification procedure for both the linear MIMO CIR matrix and the NHPAs' model, it is necessary to derive a unique parametrization of the linear MIMO channel matrix and the M nonlinear transmitters.

First, we note that there are infinitely many pairs of the equivalent parametrization, which can be expressed as $\boldsymbol{H}\boldsymbol{U} \cdot \boldsymbol{U}^* \boldsymbol{z}^{(h)}$, where $\boldsymbol{U} \in \mathbb{C}^{(Mn_H) \times (Mn_H)}$ is any unitary matrix. Second, for any particular model for the MIMO channel matrix $\boldsymbol{H}\boldsymbol{U}$, there are also infinitely many pairs of the equivalent parametrization for the model of the NHPAs $\boldsymbol{U}^* \boldsymbol{z}^{(h)}$. In order to derive a unique parametrization of the linear MIMO channel matrix and the M nonlinear transmitters, therefore, we need: 1) first to determine a particular parametrization of the MIMO CIR matrix $\boldsymbol{H}\boldsymbol{U}$, and 2) next to derive a particular parametrization of the NHPAs' model $\boldsymbol{U}^* \boldsymbol{z}^{(h)}$. To achieve these two tasks, we re-express (8) equivalently as

$$\boldsymbol{x}^{(h)}(k) = \begin{bmatrix} \frac{1}{h_{0,1,1}} \boldsymbol{h}_{1,1}^{\mathrm{T}} & \frac{1}{h_{0,1,2}} \boldsymbol{h}_{1,2}^{\mathrm{T}} & \cdots & \frac{1}{h_{0,1,M}} \boldsymbol{h}_{1,M}^{\mathrm{T}} \\ \frac{1}{h_{0,1,1}} \boldsymbol{h}_{2,1}^{\mathrm{T}} & \frac{1}{h_{0,1,2}} \boldsymbol{h}_{2,2}^{\mathrm{T}} & \cdots & \frac{1}{h_{0,1,M}} \boldsymbol{h}_{2,M}^{\mathrm{T}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{h_{0,1,1}} \boldsymbol{h}_{L,1}^{\mathrm{T}} & \frac{1}{h_{0,1,2}} \boldsymbol{h}_{L,2}^{\mathrm{T}} & \cdots & \frac{1}{h_{0,1,M}} \boldsymbol{h}_{L,M}^{\mathrm{T}} \end{bmatrix} \times \begin{bmatrix} h_{0,1,1} \boldsymbol{z}_1^{(h)}(k) \\ \frac{h_{0,1,2}}{h_{0,1,1}} h_{0,1,1} \boldsymbol{z}_2^{(h)}(k) \\ \vdots \\ \frac{h_{0,1,M}}{h_{0,1,1}} h_{0,1,1} \boldsymbol{z}_M^{(h)}(k) \end{bmatrix} + \boldsymbol{n}^{(H)}(k). \tag{16}$$

From (16), we have a unique parametrized MIMO channel matrix $\boldsymbol{H}\boldsymbol{U}$ as

$$\boldsymbol{H} = \begin{bmatrix} \frac{1}{h_{0,1,1}} \boldsymbol{h}_{1,1}^{\mathrm{T}} & \frac{1}{h_{0,1,2}} \boldsymbol{h}_{1,2}^{\mathrm{T}} & \cdots & \frac{1}{h_{0,1,M}} \boldsymbol{h}_{1,M}^{\mathrm{T}} \\ \frac{1}{h_{0,1,1}} \boldsymbol{h}_{2,1}^{\mathrm{T}} & \frac{1}{h_{0,1,2}} \boldsymbol{h}_{2,2}^{\mathrm{T}} & \cdots & \frac{1}{h_{0,1,M}} \boldsymbol{h}_{2,M}^{\mathrm{T}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{h_{0,1,1}} \boldsymbol{h}_{L,1}^{\mathrm{T}} & \frac{1}{h_{0,1,2}} \boldsymbol{h}_{L,2}^{\mathrm{T}} & \cdots & \frac{1}{h_{0,1,M}} \boldsymbol{h}_{L,M}^{\mathrm{T}} \end{bmatrix}. \tag{17}$$

In (17), we still denote this equivalent linear MIMO channel matrix $\boldsymbol{H}\boldsymbol{U}$ by $\boldsymbol{H}$ for notational simplicity. From (16), we also note that the M nonlinear transmitters can be expressed as $z_m(k) = \frac{h_{0,1,m}}{h_{0,1,1}} \cdot \left(h_{0,1,1} \boldsymbol{z}_m^{(h)}(k)\right)$ for $1 \le m \le M$. Therefore, we have a unique parametrized NHPAs' model $\boldsymbol{U}^* \boldsymbol{z}^{(h)}$ as

$$z_m(k) = \zeta_m \Psi(s_m(k)), \ 1 \le m \le M, \tag{18}$$

with $\zeta_1 = 1$ and $\zeta_m = \frac{h_{0,1,m}}{h_{0,1,1}} \in \mathbb{C}$ for $2 \le m \le M$. Note that (18) corresponds to absorbing $h_{0,1,1}$ into the NHPA's response $\Psi(\cdot)$, and again for notational simplicity, we still denote this equivalent NHPA's response $h_{0,1,1}\Psi(\cdot)$ by $\Psi(\cdot)$.

Compared to the nonlinear frequency-nonselective MIMO channel of [28], the nonlinear frequency-selective MIMO channel, namely, the nonlinear MIMO Hammerstein system (17) and (18), is much more complicated. In particular, MIMO channel matrix $\boldsymbol{H} \in \mathbb{C}^{n_H LM}$ is n_H times larger than the MIMO channel matrix of [28]. Identification of such a large-size MIMO nonlinear system, consisting of the frequency-selective MIMO CIR multiplicative with the nonlinear model of the M nonlinear transmitters, is much more difficult than the corresponding identification task in [28].

3. Proposed Nonlinear Space-time Equalization

As shown in the previous section, implementing the nonlinear space-time equalizer for multiuser nonlinear frequency-selective MIMO uplink requires the knowledge of the dispersive linear MIMO CIR matrix $\boldsymbol{H}$ as well as the inverse mappings of all the MUs' nonlinear HPAs. Since the dispersive linear MIMO CIR matrix $\boldsymbol{H}$ is cascaded with the M nonlinear transmitters, in order to acquire $\boldsymbol{H}$, it is necessary to jointly

estimate both the MIMO CIR matrix $\boldsymbol{H}$ and the model of the M NHPAs. Based on the unique parametrization of Subsection II-C, we develop a BSNN based approach for estimating both the MIMO CIR matrix $\boldsymbol{H}$ and the MUs' nonlinear transmitters $\zeta_m\Psi(\cdot)$ for $1 \le m \le M$. Furthermore, the results of this nonlinear identification enable us to acquire the inverse mappings $\big(\zeta_m\Psi\big)^{-1}(\cdot)$ of $\zeta_m\Psi(\cdot)$ for $1 \le m \le M$.

3.1 BSNN based parametrization

To jointly estimate the MIMO CIR matrix $\boldsymbol{H}$ (17) and the M NHPAs (18), we still need to parametrize the complex-valued NHPA $\Psi(\cdot)$ of (3). We propose to use a complex-valued BSNN for this parametrization. The reason why we choose the BSNN rather than other nonlinear models is because among all the universe approximators for the class of nonlinear continuous functions in the univariate dimension, the BSNN has the maximum robustness to estimation error [37]–[39]. In other words, it is an optimal choice for this task.

First, we establish some physical properties of the NHPA $\Psi(\cdot)$ and its input $s = s_R + \mathrm{j}s_I$, which are essential for our BSNN parametrization. Clearly, the NHPA $\Psi(\cdot)$ is a one-to-one continuous mapping, and therefore it is invertible. This establishes the physical base for identifying $\Psi(\cdot)$ as well as inverting it. The input to the NHPA s takes value from the QAM constellation $\mathbb{S}$ of (2). Observe from the QAM signaling (2) that the constellation points are symmetrically distributed, and they are both upper and lower bounded. In order words, the distributions of s_R and s_I are identical and symmetric. In addition, since $\big(-\sqrt{U}+1\big)d_{\mathrm{S}} < s_R, s_I < \big(\sqrt{U}-1\big)d_{\mathrm{S}}$, we can always specify some known finite real values, $U_{\min}$ and $U_{\max}$, such that $U_{\min} < s_R, s_I < U_{\max}$.

1) Univariate BSNN: Consider a generic continuous nonlinear real-valued function $y = f(u)$ defined in the univariate dimension of $u \in \mathbb{R}$, and its input u is both upper and lower bounded, i.e., $U_{\min} < u < U_{\max}$, with the known $U_{\min}$ and $U_{\max}$. We use a univariate BSNN with piecewise polynomial degree of P_o and N_{r} basis functions to model this nonlinear function. According to [29], the univariate BSNN is built upon the so-called knot sequence specified by $(N_{\mathrm{r}} + P_o + 1)$ knot values, denoted as $\{U_0, U_1, \cdots, U_{N_{\mathrm{r}}+P_o}\}$, with the following relationship

$$\begin{aligned} U_0 < U_1 < \cdots < U_{P_o-2} < U_{P_o-1} = U_{\min} < U_{P_o} < \cdots \\ < U_{N_{\mathrm{r}}} < U_{N_{\mathrm{r}}+1} = U_{\max} < U_{N_{\mathrm{r}}+2} < \cdots < U_{N_{\mathrm{r}}+P_o}. \end{aligned} \tag{19}$$

Since the input region is $\big(U_{\min},\ U_{\max}\big)$, we have $N_{\mathrm{r}}+1-P_o$ internal knots inside $\big(U_{\min},\ U_{\max}\big)$, two boundary knots ($U_{\min}$ and $U_{\max}$), and $2(P_o-1)$ external knots outside the input region. Given the set of predetermined knots (19), we can compute the set of N_{r} B-spline basis functions. Specifically, using the well-known De Boor recursion [29], we start from the zero-order basis functions

$$B_l^{(\mathrm{r},0)}(u) = \begin{cases} 1, & \text{if } U_{l-1} \le u < U_l, \\ 0, & \text{otherwise}, \end{cases} \quad 1 \le l \le N_{\mathrm{r}} + P_o, \tag{20}$$

and recursively compute the pth order basis functions with $p = 1, \cdots, P_o$

$$\begin{aligned} B_l^{(\mathrm{r},p)}(u) =& \frac{u - U_{l-1}}{U_{p+l-1} - U_{l-1}} B_l^{(\mathrm{r},p-1)}(u) \\ &+ \frac{U_{p+l} - u}{U_{p+l} - U_l} B_{l+1}^{(\mathrm{r},p-1)}(u), \end{aligned} \tag{21}$$

for $l = 1, \cdots, N_{\mathrm{r}} + P_o - p$. The BSNN model for $y = f(u)$ is then produced as

$$y = \sum_{i=1}^{N_{\mathrm{r}}} b_i B_i^{(\mathrm{r},P_o)}(u), \tag{22}$$

where b_i for $1 \le i \le N_{\mathrm{r}}$ are the BSNN model coefficients.

2) Structure determination: We now discuss how to determine the structure parameters, P_o and N_{r}, for the univariate BSNN model (22). For modeling the nonlinear functions commonly encountered in the real world, the polynomial degree $P_o = 3$ or 4 is sufficient [31]–[36]. Since the input region $\big(U_{\min},\ U_{\max}\big)$ is a bounded interval, using $N_{\mathrm{r}} = 6$ to 10 B-spline basis functions is also sufficient for accurately modeling over the interval $\big(U_{\min},\ U_{\max}\big)$. As regarding how to determine the knot sequence relationship (19), the two boundary knots can obviously be set to the known values $U_{\min}$ and $U_{\max}$, respectively, and the $N_{\mathrm{r}} + 1 - P_o$ internal knots can be uniformly spaced in the interval $\big(U_{\min},\ U_{\max}\big)$. The $2(P_o-1)$ external knots are used to equip the BSNN model with extrapolating capability outside the input region $\big(U_{\min},\ U_{\max}\big)$, and they can be set empirically. Since no data appears outside $\big(U_{\min},\ U_{\max}\big)$, the choice of these external knots does not really matter, in terms of modeling accuracy. Also the physical properties of the system to be modeled can be taken into account to improve the modeling performance. For our application with the symmetric QAM signals, the distribution of $u = s_R$ or $u = s_I$ is naturally symmetric, and therefore the knot sequence should be symmetrically distributed too.

3) Computational complexity: The computational complexity of the univariate BSNN model (22) depends on P_o, not N_{r}. This is because given any input $u \in \big(U_{\min},\ U_{\max}\big)$, it can be shown that no more than (P_o+1) basis functions are nonzero [40]. In [40], it further demonstrates that the complexity of the BSNN model (22) is no more than twice of the following polynomial basis model

$$y = \sum_{i=0}^{P_o} a_i u^i, \tag{23}$$

with the model coefficients a_i for $0 \le i \le P_o$, and the basis function set

$$1,\ u,\ u^2, \cdots, u^{P_o}. \tag{24}$$

4) Maximum robustness to estimation error: The works [37]–[39] have established the fact that among all the universe approximators for the class of nonlinear continuous functions in the univariate dimension, the BSNN model (22) has the optimal maximum robustness property. This optimal maximum robustness property of the BSNN is due to the convexity of its model bases, specifically, all the B-spline bases are positive

and they sum up to unity. This maximum robustness property provides the BSNN model with the maximum robustness to estimation error and enables the BSNN model to attain highly accurate estimation in noisy-data environments [28], [34], [36], [40]–[42], outperforming other universe approximators of similar complexity with non-convex model bases.

For example, the model bases (24) of the polynomial model (23) do not possess the convexity property, and our previous applications [28], [34], [36], [40]–[42] have all confirmed that the BSNN model significantly outperforms the polynomial model, particularly under highly noisy environments. In fact, the previous analysis given in [40] has explained exactly why. More specifically, recall the real-valued true nonlinear system $y = f(u)$ with $y, u \in \mathbb{R}$. Assume that this nonlinear function can be exactly modeled by the polynomial model (24) of degree P_o or by the BSNN model (22) of polynomial degree P_o with $N_{\rm r}$ basis functions. Because the training data are noisy, due to the estimation error, the estimated model coefficients are perturbed from their true values a_i to $\widehat{a}_i = a_i + \varepsilon_i$ for the polynomial model, and from their true values b_i to $\widehat{b}_i = b_i + \varepsilon_i$ for the BSNN model. Let us assume that all the estimation errors ε_i are bounded, namely, $\left|\varepsilon_i\right| < \varepsilon_{\max}$. The modeling error for the polynomial model satisfies the following condition

$$|y - \widehat{y}| = \left| \sum_{i=0}^{P_o} a_i u^i - \sum_{i=0}^{P_o} \widehat{a}_i u^i \right| < \varepsilon_{\max} \left| \sum_{i=0}^{P_o} u^i \right|. \qquad (25)$$

Observe that the upper bound of the modeling error for the polynomial model depends not only on the upper bound of the estimation error but also on the input value u and the polynomial degree P_o. For example, the higher the polynomial degree P_o is, the higher the modeling error of the polynomial model will be. By contrast, the modeling error $\left|y - \widehat{y}\right|$ for the BSNN model meets the following condition owe to the convexity of its model bases

$$\begin{aligned} \left|y - \widehat{y}\right| =& \left| \sum_{i=1}^{N_{\rm r}} b_i B_i^{({\rm r},P_o)}(u) - \sum_{i=1}^{N_{\rm r}} \widehat{b}_i B_i^{({\rm r},P_o)}(u) \right| \\ <& \varepsilon_{\max} \left| \sum_{i=1}^{N_{\rm r}} B_i^{({\rm r},P_o)}(u) \right| = \varepsilon_{\max}. \end{aligned} \qquad (26)$$

Clearly, the upper bound of the modeling error for the BSNN model only depends on the upper bound of the estimation error, and it does not depend on the input value x, the number of basis functions $N_{\rm r}$ or the polynomial degree P_o. Unlike the polynomial model, given the estimation error, the modeling error of the BSNN model is not amplified. This confirms that the BSNN model has the maximum robustness to estimation error or noise.

5) Complex-valued BSNN model for NHPA: The input $s = s_R + {\rm j} s_I$ to the NHPA (3) is complex-valued or bivariate and the output of the NHPA $\Psi(s)$ is also complex-valued. We now discuss how to construct the complex-valued BSNN model for the NHPA. First, based on the univariate BSNN modeling discussed in Subsection III-A1, for the inputs s_R and s_I, we can construct the two sets of the univariate B-spline basis functions, $B_r^{(R,P_o)}(s_R)$ for $1 \le r \le N_R$ and $B_i^{(I,P_o)}(s_I)$ for $1 \le i \le N_I$, respectively. Then by applying the tensor product between these two sets of univariate B-spline basis functions [30], we obtain the new set of bivariate B-spline basis functions $B_{r,i}^{(P_o)}(s) = B_r^{(R,P_o)}(s_R) B_i^{(I,P_o)}(s_I)$ for $1 \le r \le N_R$ and $1 \le i \le N_I$. This yields the following complex-valued BSNN model for the NHPA $\Psi(\cdot)$

$$\begin{aligned} \widehat{z} =& \widehat{\Psi}(s) = \sum_{r=1}^{N_R} \sum_{i=1}^{N_I} B_{r,i}^{(P_o)}(s) \theta_{r,i} \\ =& \sum_{r=1}^{N_R} \sum_{i=1}^{N_I} B_r^{(R,P_o)}(s_R) B_i^{(I,P_o)}(s_I) \theta_{r,i}, \end{aligned} \qquad (27)$$

where $\theta_{r,i} \in \mathbb{C}$, $1 \le r \le N_R$ and $1 \le i \le N_I$, are the complex-valued BSNN model coefficients. By collecting all the coefficients into a vector form

$$\boldsymbol{\theta} = \left[\theta_{1,1}\ \theta_{1,2} \cdots \theta_{r,i} \cdots \theta_{N_R,N_I}\right]^{\rm T} \in \mathbb{C}^{N_B}, \qquad (28)$$

where $N_B = N_R N_I$, the task of identifying the complex-valued NHPA $\Psi(\cdot)$ becomes one of estimating the complex-valued parameter vector $\boldsymbol{\theta}$.

3.2 Nonlinear frequency-selective MIMO uplink identification

From Subsection II-C, the multiuser nonlinear frequency-selective MIMO uplink is the multiplicative cascade of the M nonlinear transmitters (18) with the MIMO CIR matrix (17). Further adopting the BSNN model for the NHPA given in Subsection III-A, we have the unique parametrization of this multiuser nonlinear frequency-selective MIMO uplink, which involves the parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{\zeta} = \left[\zeta_1\ \zeta_2 \cdots \zeta_M\right]^{\rm T}$, where $\zeta_1 = 1$, of the M complex-valued BSNNs as well as the multiuser MIMO CIR matrix $\boldsymbol{H}$, where $h_{0,1,m} = 1$ for $1 \le m \le M$.

1) Estimation signal representation: We first collect a block of K training data, $\{\boldsymbol{s}^{(h)}(k), \boldsymbol{x}^{(h)}(k)\}_{k=1}^{K}$, in which the training input $\boldsymbol{s}^{(h)}(k) = \left[\left(\boldsymbol{s}_1^{(h)}(k)\right)^{\rm T} \cdots \left(\boldsymbol{s}_M^{(h)}\right)^{\rm T}\right]^{\rm T}$ with $\boldsymbol{s}_m^{(h)}(k) = \left[s_m(k)\ s_m(k-1) \cdots s_m(k-n_H+1)\right]^{\rm T}$, and the desired output $\boldsymbol{x}^{(h)}(k) = [x_1(k) \cdots x_L(k)]^{\rm T}$. The outputs $\widehat{x}_l(k)$ of our nonlinear model for modeling the desired outputs $x_l(k)$ for $1 \le l \le L$ can be expressed by

$$\begin{aligned} \widehat{x}_l(k) =& \sum_{m=1}^{M} \sum_{q=0}^{n_H-1} h_{q,l,m} \widehat{z}_m(k-q) \\ =& \sum_{m=1}^{M} \sum_{q=0}^{n_H-1} \sum_{r=1}^{N_R} \sum_{i=1}^{N_I} B_{r,i}^{(P_o)}(s_m(k-q)) h_{q,l,m} \zeta_m \theta_{r,i}. \end{aligned} \qquad (29)$$

From (29), it can be seen that this nonlinear frequency-selective MIMO uplink identification is a very challenging nonlinear estimation problem because the parameters to be estimated enter the model in the nonlinear triple product form of $h_{q,l,m}\zeta_m\theta_{r,i}$. To devise an effective iterative estimation procedure, we need the regression representations that are 'linear' in $h_{q,l,m}$, ζ_m and $\theta_{r,i}$, respectively.

1) Linear in $\boldsymbol{H}$ regression model: Clearly, we can express the desired output matrix $\boldsymbol{X} \in \mathbb{C}^{L\times K}$ as

$$\boldsymbol{X} = \begin{bmatrix} x_1(1) & x_1(2) & \cdots & x_1(K) \\ x_2(1) & x_2(2) & \cdots & x_2(K) \\ \vdots & \vdots & \cdots & \vdots \\ x_L(1) & x_L(2) & \cdots & x_L(K) \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1^{\mathrm{T}} \\ \boldsymbol{x}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{x}_L^{\mathrm{T}} \end{bmatrix}. \quad (30)$$

Recalling the MIMO channel model (8), $\boldsymbol{X}$ can be further expressed as

$$\boldsymbol{X} = \boldsymbol{H}\boldsymbol{Q} + \boldsymbol{N}, \quad (31)$$

in which $\boldsymbol{N} \in \mathbb{C}^{L\times K}$ denotes the channel AWGN matrix, and the 'regression' matrix $\boldsymbol{Q} \in \mathbb{C}^{(Mn_H)\times K}$ is given by

$$\boldsymbol{Q} = \begin{bmatrix} \widehat{\boldsymbol{z}}_1(1) & \widehat{\boldsymbol{z}}_1(2) & \cdots & \widehat{\boldsymbol{z}}_1(K) \\ \widehat{\boldsymbol{z}}_2(1) & \widehat{\boldsymbol{z}}_2(2) & \cdots & \widehat{\boldsymbol{z}}_2(K) \\ \vdots & \vdots & \cdots & \vdots \\ \widehat{\boldsymbol{z}}_M(1) & \widehat{\boldsymbol{z}}_M(2) & \cdots & \widehat{\boldsymbol{z}}_M(K) \end{bmatrix}, \quad (32)$$

in which $\widehat{\boldsymbol{z}}_m(k) = \left[\widehat{z}_m(k)\ \widehat{z}_m(k-1)\cdots\widehat{z}_m(k-n_H+1)\right]^{\mathrm{T}}$ with

$$\widehat{z}_m(k-q) = \sum_{r=1}^{N_R}\sum_{i=1}^{N_I} B_{r,i}^{(P_o)}(s_m(k-q))\zeta_m\theta_{r,i}, \quad 0 \le q \le n_H-1, \quad (33)$$

for $1 \le m \le M$. The regression model (31) is indeed linear in $\boldsymbol{H}$ but its regression matrix $\boldsymbol{Q}$ is nonlinear in the parameter products $\zeta_m\theta_{r,i}$.

2) Linear in $\boldsymbol{\theta}$ regression model: Next, express the desired output vectors $\boldsymbol{x}_l \in \mathbb{C}^K$, $1 \le l \le L$, where $\boldsymbol{x}_l^{\mathrm{T}}$ is the lth row of $\boldsymbol{X}$, as

$$\boldsymbol{x}_l = \boldsymbol{P}_l\boldsymbol{\theta} + \boldsymbol{n}_l,\ 1 \le l \le L, \quad (34)$$

in which $\boldsymbol{n}_l \in \mathbb{C}^K$ is the corresponding channel AWGN vector, and the 'regression' matrix $\boldsymbol{P}_l \in \mathbb{C}^{K\times N_B}$ is given by

$$\boldsymbol{P}_l = \begin{bmatrix} \phi_{1,1}^{(l)}(1) & \phi_{1,2}^{(l)}(1) & \cdots & \phi_{N_R,N_I}^{(l)}(1) \\ \phi_{1,1}^{(l)}(2) & \phi_{1,2}^{(l)}(2) & \cdots & \phi_{N_R,N_I}^{(l)}(2) \\ \vdots & \vdots & \cdots & \vdots \\ \phi_{1,1}^{(l)}(K) & \phi_{1,2}^{(l)}(K) & \cdots & \phi_{N_R,N_I}^{(l)}(K) \end{bmatrix}, \quad (35)$$

with

$$\phi_{r,i}^{(l)}(k) = \sum_{m=1}^{M}\sum_{q=0}^{n_H-1} h_{q,l,m}\zeta_m B_{r,i}^{(P_o)}\big(s_m(k-q)\big), \quad (36)$$

for $1 \le r \le N_R$ and $1 \le i \le N_I$. Aggregating (34) for $1 \le l \le L$, we have

$$\sum_{l=1}^{L}\boldsymbol{x}_l = \sum_{l=1}^{L}\boldsymbol{P}_l\boldsymbol{\theta} + \sum_{l=1}^{L}\boldsymbol{n}_l \ \Rightarrow\ \overline{\boldsymbol{x}} = \overline{\boldsymbol{P}}\boldsymbol{\theta} + \overline{\boldsymbol{n}}. \quad (37)$$

This model is linear in $\boldsymbol{\theta}$ but its regression matrix $\overline{\boldsymbol{P}}$ is nonlinear in $h_{q,l,m}\zeta_m$.

3) Linear in $\boldsymbol{\zeta}$ regression model: Similarly, express $\boldsymbol{x}_l \in \mathbb{C}^K$, $1 \le l \le L$, as

$$\boldsymbol{x}_l = \boldsymbol{S}_l\boldsymbol{\zeta} + \boldsymbol{n}_l,\ 1 \le l \le L, \quad (38)$$

where the 'regression' matrix $\boldsymbol{S}_l \in \mathbb{C}^{K\times M}$ is given by

$$\boldsymbol{S}_l = \begin{bmatrix} \boldsymbol{h}_{l,1}^{\mathrm{T}}\boldsymbol{\psi}_1(1) & \boldsymbol{h}_{l,2}^{\mathrm{T}}\boldsymbol{\psi}_2(1) & \cdots & \boldsymbol{h}_{l,M}^{\mathrm{T}}\boldsymbol{\psi}_M(1) \\ \boldsymbol{h}_{l,1}^{\mathrm{T}}\boldsymbol{\psi}_1(2) & \boldsymbol{h}_{l,2}^{\mathrm{T}}\boldsymbol{\psi}_2(2) & \cdots & \boldsymbol{h}_{l,M}^{\mathrm{T}}\boldsymbol{\psi}_M(2) \\ \vdots & \vdots & \cdots & \vdots \\ \boldsymbol{h}_{l,1}^{\mathrm{T}}\boldsymbol{\psi}_1(K) & \boldsymbol{h}_{l,2}^{\mathrm{T}}\boldsymbol{\psi}_2(K) & \cdots & \boldsymbol{h}_{l,M}^{\mathrm{T}}\boldsymbol{\psi}_M(K) \end{bmatrix}, \quad (39)$$

in which $\boldsymbol{\psi}_m(k) = \left[\psi_m(k)\ \psi_m(k-1)\cdots\psi_m(k-n_H+1)\right]^{\mathrm{T}}$ and

$$\psi_m(k-q) = \sum_{r=1}^{N_R}\sum_{i=1}^{N_I} B_{r,i}^{(P_o)}\big(s_m(k-q)\big)\theta_{r,i}, \quad (40)$$

for $0 \le q \le n_H-1$ and $1 \le m \le M$. Hence we have

$$\sum_{l=1}^{L}\boldsymbol{x}_l = \sum_{l=1}^{L}\boldsymbol{S}_l\boldsymbol{\zeta} + \sum_{l=1}^{L}\boldsymbol{n}_l \ \Rightarrow\ \overline{\boldsymbol{x}} = \overline{\boldsymbol{S}}\boldsymbol{\zeta} + \overline{\boldsymbol{n}}. \quad (41)$$

This model is linear in $\boldsymbol{\zeta}$ but its regression matrix $\overline{\boldsymbol{S}}$ is nonlinear in $h_{q,l,m}\theta_{r,i}$.

2) Iterative ALS procedure: We extend the estimation algorithm of [28] originally developed for efficiently identifying the multiuser nonlinear narrowband MIMO uplink model to our current application of the multiuser nonlinear frequency-selective MIMO uplink, and derive a new iterative ALS procedure for estimating $\boldsymbol{H}$, $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$. The basic idea is that if we alternatively fix the two parameters among the triple parameters $\boldsymbol{H}$, $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$, the third parameter can be obtained by the least squares (LS) estimator. The estimation procedure involves two iterative loops with three-stage ALS estimations of $\boldsymbol{H}$, $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$, respectively, as detailed below.

Algorithm: Two-loop three-stage ALS estimation.
Step 1. Estimation procedure initialization.
1.1. Set the maximum number of the outer loop iterations $\varsigma_{\max}$ and the maximum number of the inner loop iterations $\upsilon_{\max}$.
1.2. Initialize $\boldsymbol{H}$ and $\boldsymbol{\zeta}$ to $\boldsymbol{H}^{[0]}$ and $\boldsymbol{\zeta}^{[0]}$. Specifically, by assuming that all the transmitters' HPAs are linear, we have the 'approximate' regression model $\boldsymbol{X} \approx \boldsymbol{S}\boldsymbol{H} + \boldsymbol{N}$, where the regression matrix $\boldsymbol{S} \in \mathbb{C}^{(n_HM)\times K}$ is given by

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{s}_1(1) & \boldsymbol{s}_1(2) & \cdots & \boldsymbol{s}_1(K) \\ \boldsymbol{s}_2(1) & \boldsymbol{s}_2(2) & \cdots & \boldsymbol{s}_2(K) \\ \vdots & \vdots & \cdots & \vdots \\ \boldsymbol{s}_M(1) & \boldsymbol{s}_M(2) & \cdots & \boldsymbol{s}_M(K) \end{bmatrix}, \quad (42)$$

with $\boldsymbol{s}_m(k) = \left[s_m(k)\ s_m(k-1)\cdots s_m(k-n_H+1)\right]^{\mathrm{T}}$ for $1 \le m \le M$. Then we can set $\boldsymbol{H}^{[0]}$ to the following standard LS estimate

$$\widehat{\boldsymbol{H}}^{[0]} = \boldsymbol{X}\boldsymbol{S}^{\mathrm{H}}\left(\boldsymbol{S}\boldsymbol{S}^{\mathrm{H}}\right)^{-1}. \quad (43)$$

To meet the unique parametrization of the MIMO CIR matrix as discussed in Subsection II-C, we 'normalize' $\widehat{\boldsymbol{H}}^{[0]}$ according to

$$\widehat{\boldsymbol{h}}_{l,m}^{[0]} = \frac{1}{\widehat{h}_{0,1,m}^{[0]}}\widehat{\boldsymbol{h}}_{l,m}^{[0]},\ 1 \le l \le L,\ 1 \le m \le M. \quad (44)$$

All the elements of $\widehat{\boldsymbol{\zeta}}^{[0]}$ can be initialized to 1, i.e., $\widehat{\zeta}_m^{[0]} = 1$ for $1 \le m \le M$.

Step 2. Start of the outer iterative loop. **For** $1 \leq \varsigma \leq \varsigma_{\max}$, **do**:
Step 3. Fix the unknown MIMO CIR $\boldsymbol{H}$ in the regression matrices $\overline{\boldsymbol{P}}$ and $\overline{\boldsymbol{S}}$ to $\widehat{\boldsymbol{H}}^{[\varsigma-1]}$, and initialize $\widehat{\boldsymbol{\zeta}}^{[0}=\widehat{\boldsymbol{\zeta}}^{[\varsigma-1]}$.
3.1. Start of the inner iterative loop. **For** $1 \leq \upsilon \leq \upsilon_{\max}$, **do**:
3.2. Fix the unknown $\boldsymbol{\zeta}$ in the regression matrix $\overline{\boldsymbol{P}}$ to $\widehat{\boldsymbol{\zeta}}^{[\upsilon-1}$ and denote the resulting $\overline{\boldsymbol{P}}$ as $\overline{\boldsymbol{P}}^{[\upsilon}$, which is now free from the unknown $\boldsymbol{H}$ and $\boldsymbol{\zeta}$. The closed-form regularized LS estimate of $\boldsymbol{\theta}$ can readily be obtained as[1]

$$\widehat{\boldsymbol{\theta}}^{[\upsilon}=\left(\left(\overline{\boldsymbol{P}}^{[\upsilon}\right)^{\mathrm{H}}\overline{\boldsymbol{P}}^{[\upsilon}+\lambda\boldsymbol{I}_{N_B}\right)^{-1}\left(\overline{\boldsymbol{P}}^{[\upsilon}\right)^{\mathrm{H}}\overline{\boldsymbol{x}}, \tag{45}$$

where λ is a very small positive regularization parameter, and $\boldsymbol{I}_{N_B}$ denotes the $N_B \times N_B$ identity matrix.
3.3. Fix the unknown $\boldsymbol{\theta}$ in the regression matrix $\overline{\boldsymbol{S}}$ to $\widehat{\boldsymbol{\theta}}^{[\upsilon}$ and denote the resulting $\overline{\boldsymbol{S}}$ as $\overline{\boldsymbol{S}}^{[\upsilon}$, which is then free from the unknown $\boldsymbol{H}$ and $\boldsymbol{\theta}$. The closed-form LS estimate of $\boldsymbol{\zeta}$ is readily given by

$$\widehat{\boldsymbol{\zeta}}^{[\upsilon}=\left(\left(\overline{\boldsymbol{S}}^{[\upsilon}\right)^{\mathrm{H}}\overline{\boldsymbol{S}}^{[\upsilon}\right)^{-1}\left(\overline{\boldsymbol{S}}^{[\upsilon}\right)^{\mathrm{H}}\overline{\boldsymbol{x}}. \tag{46}$$

To meet the unique parametrization of the M NHPAs, we normalize $\widehat{\boldsymbol{\zeta}}^{[\upsilon}$ with

$$\widehat{\zeta}_m^{[\upsilon}=\widehat{\zeta}_m^{[\upsilon}/\widehat{\zeta}_1^{[\upsilon},\ 1 \leq m \leq M. \tag{47}$$

3.4. End of the inner iterative loop.

For the fixed MIMO CIR matrix $\widehat{\boldsymbol{H}}^{[\varsigma-1]}$, we obtain the estimated parameter vectors of the M NHPAs as $\widehat{\boldsymbol{\theta}}^{[\varsigma]}=\widehat{\boldsymbol{\theta}}^{[\upsilon_{\max}}$ and $\widehat{\boldsymbol{\zeta}}^{[\varsigma]}=\widehat{\boldsymbol{\zeta}}^{[\upsilon_{\max}}$.
Step 4. In the regression matrix $\boldsymbol{Q}$, fix the unknown $\boldsymbol{\theta}$ to $\widehat{\boldsymbol{\theta}}^{[\varsigma]}$ and the unknown $\boldsymbol{\zeta}$ to $\widehat{\boldsymbol{\zeta}}^{[\varsigma]}$. The resultant $\boldsymbol{Q}$ is denoted as $\boldsymbol{Q}^{[\varsigma]}$, which becomes independent of the unknown $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$. The closed-form LS estimate of $\boldsymbol{H}$ is then given by

$$\widehat{\boldsymbol{H}}^{[\varsigma]}=\boldsymbol{X}\left(\boldsymbol{Q}^{[\varsigma]}\right)^{\mathrm{H}}\left(\boldsymbol{Q}^{[\varsigma]}\left(\boldsymbol{Q}^{[\varsigma]}\right)^{\mathrm{H}}\right)^{-1}, \tag{48}$$

which is followed by the normalization operation

$$\widehat{\boldsymbol{h}}_{l,m}^{[\varsigma]}=\frac{1}{\widehat{h}_{0,1,m}^{[\varsigma]}}\widehat{\boldsymbol{h}}_{l,m}^{[\varsigma]},\ 1 \leq l \leq L,\ 1 \leq m \leq M, \tag{49}$$

to meet the unique parametrization of the MIMO CIR matrix.
Step 5. End of the outer iterative loop.

At the end of **Algorithm**, we obtain the estimates $\widehat{\boldsymbol{H}}=\widehat{\boldsymbol{H}}^{[\varsigma_{\max}]}$, $\widehat{\boldsymbol{\theta}}=\widehat{\boldsymbol{\theta}}^{[\varsigma_{\max}]}$ and $\widehat{\boldsymbol{\zeta}}=\widehat{\boldsymbol{\zeta}}^{[\varsigma_{\max}]}$ for the multiuser nonlinear frequency-selective MIMO uplink.

3) Unbiasedness and efficiency analysis: Observe that our proposed identification procedure for the multiuser nonlinear frequency-selective MIMO uplink model contains the two iterative loops, namely, the outer iterative loop of **Step 1** to **Step 5** and the inner iterative loop of **Step 3.1** to **Step 3.4**, together with the three stages of ALS estimation, namely, the closed-form LS estimates (45), (46) (with the normalization (47)) and (48) (with the normalization (49)). The outer loop iterates between the two stages of estimating the model of the M NHPAs and estimating the multiuser MIMO CIR matrix. Within the first stage of the outer iterative loop, the inner loop iterates between the two stages of estimating $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$, which together forms the model of the M nonlinear transmitters.

We now analyze why the proposed iterative ALS procedure converges extremely fast. The initial estimate for the unknown MIMO CIR matrix is given by $\widehat{\boldsymbol{H}}^{[0]}$ of (43), which is an estimate of $\boldsymbol{H}$ scaled by the NHPAs' complex-valued gains. It can readily be seen that with the normalization operation (44), $\widehat{\boldsymbol{H}}^{[0]}$ is an unbiased unique estimate of $\boldsymbol{H}$ in (17). Therefore, give this unbiased estimate $\widehat{\boldsymbol{H}}^{[0]}$ of $\boldsymbol{H}$, the inner iterative loop converges to the unique estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$ very fast, owing to the unique parametrization of the NHPAs and the closed-form LS estimates of (45) and (46) (with the normalization (47)). Given this accurate NHPAs' model, **Step 4** of the outer iterative loop can further improve the accuracy of the estimate for the MIMO CIR matrix. Thus, the outer iterative loop with $\varsigma_{\max}=1$ iteration is in fact sufficient. To further enhance the overall estimation accuracy of the multiuser nonlinear frequency-selective MIMO uplink model, we may set $\varsigma_{\max}=2$.

It is worth emphasizing that the uniqueness of the solutions $\boldsymbol{H}$, $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$ is guaranteed by our unique parametrization of the multiuser nonlinear MIMO system, (16) to (18). In the simulation study, we will further investigate empirically the unbiasedness and efficiency property of our proposed iterative ALS estimation procedure.

3.3 Inverting the NHPAs

To implement the nonlinear STE based MUD for the multiuser nonlinear frequency-selective MIMO uplink, we also require the inverse mappings of the M NHPAs. Mathematically, the complex-valued inverse mappings of the M NHPAs are defined by

$$\begin{aligned} s_m(k)=&\left(\zeta_m\Psi\right)^{-1}(z_m(k))=\Phi_m(z_m(k)) \\ =&\Phi_m\left(\zeta_m\Psi\left(s_m\right)\right), 1 \leq m \leq M. \end{aligned} \tag{50}$$

It can be seen that the inverse mapping $\Phi_m(\cdot)$ of the mth NHPA maps the output z_m of the NHPA back to the NHPA's input s_m. This is a challenging complex-valued nonlinear inversion problem.

Since the BSNN is a universe nonlinear approximator with the maximum robustness to estimation error, as discussed in Subsection III-A, it is ideal for this nonlinear inversion problem. Thus, we employ another complex-valued BSNN to model $\Phi_m(\cdot)$. By defining the two knots sequences similar to (19) for the real and imaginary parts of $z_m=z_{m_R}+\mathrm{j}z_{m_I}$, respectively, the BSNN model for $\Phi_m(\cdot)$ can be constructed as

$$\begin{aligned} \widehat{s}_m=&\widehat{\Phi}_m(z_m)=\sum_{r=1}^{N_R}\sum_{i=1}^{N_I}B_{r,i}^{(P_o)}(z_m)\alpha_{r,i}^{(m)} \\ =&\sum_{r=1}^{N_R}\sum_{i=1}^{N_I}B_r^{(R,P_o)}(z_{m_R})B_i^{(I,P_o)}(z_{m_I})\alpha_{r,i}^{(m)}, \end{aligned} \tag{51}$$

for $1 \leq m \leq M$, where the two sets of the basis functions, $B_r^{(R,P_o)}(z_{m_R})$ for $1 \leq r \leq N_R$ and $B_i^{(I,P_o)}(z_{m_I})$ for $1 \leq i \leq N_I$, are similarly calculated according to the De Boor recursion (20) and (21). It can be seen that inverting the NHPA

[1]Since the size of $\boldsymbol{\theta}$ is relatively large, the regularization is applied to avoid ill-conditioning and enhance estimation accuracy.

$\zeta_m \Psi(\cdot)$ becomes the task of estimating the BSNN's parameter vector $\boldsymbol{\alpha}^{(m)} \in \mathbb{C}^{N_B}$ given by

$$\boldsymbol{\alpha}^{(m)} = \left[\alpha_{1,1}^{(m)} \ \alpha_{1,2}^{(m)} \cdots \alpha_{r,i}^{(m)} \cdots \alpha_{N_R,N_I}^{(m)}\right]^{\mathrm{T}}. \quad (52)$$

If we can acquire the training input-output data $\left\{z_m(k), s_m(k)\right\}_{k=1}^{K}$, then this estimation problem is easily solved. However, the input $z_m(k)$ for this identification task is unobserved and therefore unavailable. To overcome this problem, we utilize the identification results for the multiuser nonlinear frequency-selective MIMO uplink. Specifically, in this identification, we have obtained the M BSNN models $\widehat{\zeta}_m \widehat{\Psi}\left(\cdot; \widehat{\boldsymbol{\theta}}\right)$, $1 \le m \le M$, for the M NHPAs. Therefore, we can calculate the estimate of $z_m(k)$ as $\widehat{\widehat{z}}_m(k) = \widehat{\zeta}_m \widehat{\Psi}\left(s_m(k); \widehat{\boldsymbol{\theta}}\right)$, and use this 'pseudo' input $\widehat{\widehat{z}}_m(k)$ to substitute for the unknown true input $z_m(k)$. This enables us to construct the training data $\left\{\widehat{\widehat{z}}_m(k), s_m(k)\right\}_{k=1}^{K}$ for this inverse modeling. The downside is that $\widehat{\widehat{z}}_m(k)$ is not the true training input and it is highly noisy, which may potentially introduce bias in the estimate. Since we employ the BSNN as the inverse model, we can rely on the maximum robustness property of BSNN to combat this problem. After constructing the training data $\left\{\widehat{\widehat{z}}_m(k), s_m(k)\right\}_{k=1}^{K}$, we can form the linear in $\boldsymbol{\alpha}^{(m)}$ regression model from which the LS estimate of $\boldsymbol{\alpha}^{(m)}$ is readily obtained. Specifically, by defining the desired output vector as

$$\boldsymbol{s}_m = \left[s_m(1) \ s_m(2) \cdots s_m(K)\right]^{\mathrm{T}}, \quad (53)$$

and the regression matrix $\widetilde{\mathbf{B}}_m \in \mathbb{R}^{K \times N_B}$ as

$$\widetilde{\boldsymbol{B}}_m = \begin{bmatrix} B_{1,1}^{(P_o)}\left(\widehat{\widehat{z}}_m(1)\right) & \cdots & B_{N_R,N_I}^{(P_o)}\left(\widehat{\widehat{z}}_m(1)\right) \\ B_{1,1}^{(P_o)}\left(\widehat{\widehat{z}}_m(2)\right) & \cdots & B_{N_R,N_I}^{(P_o)}\left(\widehat{\widehat{z}}_m(2)\right) \\ \vdots & \cdots & \vdots \\ B_{1,1}^{(P_o)}\left(\widehat{\widehat{z}}_m(K)\right) & \cdots & B_{N_R,N_I}^{(P_o)}\left(\widehat{\widehat{z}}_m(K)\right) \end{bmatrix}, \quad (54)$$

the closed-form LS estimate $\boldsymbol{\alpha}^{(m)}$ is readily given by

$$\widehat{\boldsymbol{\alpha}}^{(m)} = \left(\widetilde{\boldsymbol{B}}_m^{\mathrm{T}} \widetilde{\boldsymbol{B}}_m\right)^{-1} \widetilde{\boldsymbol{B}}_m^{\mathrm{T}} \boldsymbol{s}_m. \quad (55)$$

Although the training input $\widehat{\widehat{z}}_m(k)$ is noisy, the optimal maximum robustness property of the BSNN as discussed in Subsection III-A4 ensures that the LS estimate (55) is unbiased and highly accurate.

4. Simulation Evaluation

4.1 Simulation system setup

The simulated multiuser nonlinear frequency-selective MIMO uplink is specified in Table I. Since the system has L receiver antennas and M users, we define the multiuser MIMO system's average signal-to-noise ratio (SNR) as the ratio of the total transmitted signal power over the total noise power, given by

$$\text{Average SNR} = \frac{\sum_{m=1}^{M} \sigma_{z_m}^2}{L \cdot 2\sigma_n^2}, \quad (56)$$

where $\sigma_{z_m}^2 = E\{|z_m(k)|^2\}$ is the average power of the mth MU's transmitted signal. The BSNNs used for modeling and inverse modeling of NHPA are specified in Table II. As explained in Subsection III-A2, choosing $N_R = N_I = 8$ and $P_o = 4$ is adequate for our application. The knot sequences in Table II are chosen to cover the NHPA's operating range and match the 64-QAM signals. Observe that the knot sequences for s_R and s_I are identical, and they are symmetric, since the distributions of s_R and s_I are symmetric and identical. Similarly, the knot sequences for z_R and z_I are identical, and they are symmetric.

Since the total number of model parameters for this multiuser nonlinear frequency-selective MIMO channel is $L \times M \times n_H + N_B + M = 112$, the number of training samples is set to $K = 1000$ for ensuring the estimation accuracy. For the iterative ALS procedure, we set the number of outer-loop iterations to $\varsigma_{\max} = 2$ and the number of inner-loop iterations to $\upsilon_{\max} = 2$. As explained in Subsection III-B3, this choice is sufficient for the iterative ALS procedure to converge to the unique and accurate estimates of $\boldsymbol{H}$, $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$.

TABLE I
PARAMETERS OF SIMULATED MULTIUSER NONLINEAR FREQUENCY-SELECTIVE MIMO UPLINK

BS antennas: $L = 5$; MUs: $M = 3$; modulation: 64-QAM; CIR length: $n_H = 3$
NHPA: (4) and (5) with $g_a = 19$, $\beta_a = 0.81$, $A_{\text{sat}} = 1.4$, $\alpha_\phi = -48000$, $\beta_\phi = 0.123$, $q_1 = 3.8$, $q_2 = 3.7$
Space-time equalizer length: $n_F = 10$, and decision delay: $d = 5$

TABLE II
STRUCTURE PARAMETERS OF B-SPLINE NEURAL NETWORK.

Polynomial degree: $P_o = 4$, number of basis functions: $N_R = N_I = 8$
Knot sequence for s_R and s_I (modeling of NHPA) -10.0, -9.0, -1.0, -**0.9**, -0.05, -0.02, 0.0, 0.02, 0.05, **0.9**, 1.0, 9.0, 10.0
Knot sequence for z_R and z_I (inverse modeling of NHPA) -20.0, -18.0, -3.0, -**1.4**, -0.8, -0.4, 0.0, 0.4, 0.8, **1.4**, 3.0, 18.0, 20.0

4.2 Estimation results by our BSNN approach

We first demonstrate that the proposed BSNN based identification algorithms presented in Subsections III-B and III-C are

TABLE III
UNIQUE PARAMETRIZED TRUE MULTIUSER NONLINEAR FREQUENCY-SELECTIVE MIMO CHANNEL. $L = 5$, $M = 3$ AND $n_H = 3$. THE NHPA IS SPECIFIED IN TABLE I.

NHPAs' true weightings $\boldsymbol{\zeta}^{\mathrm{T}}$		
1	1	1
True $\boldsymbol{H} = \left[\left[\boldsymbol{h}_{l,m}^{\mathrm{T}}, 1 \le m \le 3\right] \in \mathbb{C}^{1\times 9}\right], 1 \le l \le 5$		
1	$0.4740 + \mathrm{j}1.1054$	$0.3705 - \mathrm{j}0.7751$
1	$0.3755 + \mathrm{j}0.4018$	$1.6995 - \mathrm{j}0.2905$
1	$-0.1295 - \mathrm{j}1.4125$	$-0.5323 - \mathrm{j}0.4941$
$0.3291 - \mathrm{j}0.1268$	$1.0269 + \mathrm{j}0.4665$	$-0.5798 + \mathrm{j}0.8334$
$-0.5858 - \mathrm{j}0.2308$	$-0.3396 + \mathrm{j}0.1845$	$-0.2193 - \mathrm{j}0.3347$
$1.3517 - \mathrm{j}1.3128$	$-0.6780 + \mathrm{j}0.9676$	$0.8737 - \mathrm{j}0.3385$
$-0.1278 + \mathrm{j}0.6590$	$0.0567 - \mathrm{j}0.2107$	$-0.4374 - \mathrm{j}0.5615$
$-0.5436 - \mathrm{j}0.5148$	$0.7399 + \mathrm{j}0.2869$	$-0.5403 + \mathrm{j}0.7881$
$0.0122 + \mathrm{j}0.9869$	$0.3670 + \mathrm{j}0.4122$	$0.1809 + \mathrm{j}0.2305$
$-1.0084 - \mathrm{j}0.4358$	$-0.0909 - \mathrm{j}0.4223$	$-0.8884 - \mathrm{j}0.4641$
$-0.2137 - \mathrm{j}0.2550$	$-0.1393 - \mathrm{j}0.3626$	$-0.2465 + \mathrm{j}0.0176$
$-1.1740 + \mathrm{j}0.7498$	$-1.7164 + \mathrm{j}0.6888$	$-0.6179 + \mathrm{j}0.6992$
$-0.6067 - \mathrm{j}0.7319$	$0.1464 + \mathrm{j}0.5121$	$-0.4454 + \mathrm{j}0.4105$
$0.0466 - \mathrm{j}0.5741$	$0.8389 - \mathrm{j}0.9315$	$0.1460 - \mathrm{j}0.7706$
$1.0872 + \mathrm{j}1.0012$	$-0.8176 + \mathrm{j}1.3148$	$1.8309 + \mathrm{j}0.5452$

TABLE IV
BSNN BASED IDENTIFICATION RESULTS FOR THE MULTIUSER NONLINEAR FREQUENCY-SELECTIVE MIMO CHANNEL OF TABLE III. THE OBO IS 3 DB AND THE AVERAGE SNR IS 20 DB. THE RESULTS ARE OBTAINED OVER 100 INDEPENDENT RUNS, AND ARE PRESENTED AS: AVERAGE ESTIMATE (±STANDARD DEVIATION). THE BSNN ESTIMATED NHPAS ARE DEPICTED IN FIG. 3.

Estimated weightings of MUs' HPAs $\widehat{\boldsymbol{\zeta}}$ by BSNN approach		
1	$1.0000 + \mathrm{j}\,0.0000$ $(\pm 0.0023 \pm \mathrm{j}\,0.0024)$	$1.0000 - \mathrm{j}\,0.0002$ $(\pm 0.0023 \pm \mathrm{j}\,0.0027)$
Estimated MIMO channel matrix $\widehat{\boldsymbol{H}} = \left[\left[\widehat{\boldsymbol{h}}_{l,m}^{\mathrm{T}}, 1 \le m \le 3\right] \in \mathbb{C}^{1\times 9}\right]$, $1 \le l \le 5$, by BSNN approach		
1	$0.4735 + \mathrm{j}\,1.1052$ $(\pm 0.0032 \pm \mathrm{j}\,0.0032)$	$0.3704 - \mathrm{j}\,0.7748$ $(\pm 0.0023 \pm \mathrm{j}\,0.0026)$
1	$0.3754 + \mathrm{j}\,0.4020$ $(\pm 0.0020 \pm \mathrm{j}\,0.0019)$	$1.6995 - \mathrm{j}\,0.2908$ $(\pm 0.0036 \pm \mathrm{j}\,0.0033)$
1	$-0.1297 - \mathrm{j}\,1.4128$ $(\pm 0.0032 \pm \mathrm{j}\,0.0032)$	$-0.5323 - \mathrm{j}\,0.4944$ $(\pm 0.0024 \pm \mathrm{j}\,0.0023)$
$0.3291 - \mathrm{j}\,0.1270$ $(\pm 0.0017 \pm \mathrm{j}\,0.0022)$	$1.0270 + \mathrm{j}\,0.4665$ $(\pm 0.0025 \pm \mathrm{j}\,0.0029)$	$-0.5800 + \mathrm{j}\,0.8330$ $(\pm 0.0027 \pm \mathrm{j}\,0.0026)$
$-0.5859 - \mathrm{j}\,0.2308$ $(\pm 0.0021 \pm \mathrm{j}\,0.0020)$	$-0.3399 + \mathrm{j}\,0.1846$ $(\pm 0.0019 \pm \mathrm{j}\,0.0019)$	$-0.2191 - \mathrm{j}\,0.3343$ $(\pm 0.0019 \pm \mathrm{j}\,0.0017)$
$1.3521 - \mathrm{j}\,1.3126$ $(\pm 0.0044 \pm \mathrm{j}\,0.0038)$	$-0.6780 + \mathrm{j}\,0.9675$ $(\pm 0.0026 \pm \mathrm{j}\,0.0026)$	$0.8741 - \mathrm{j}\,0.3383$ $(\pm 0.0022 \pm \mathrm{j}\,0.0027)$
$-0.1277 + \mathrm{j}\,0.6592$ $(\pm 0.0020 \pm \mathrm{j}\,0.0021)$	$0.0565 - \mathrm{j}\,0.2107$ $(\pm 0.0022 \pm \mathrm{j}\,0.0018)$	$-0.4371 - \mathrm{j}\,0.5613$ $(\pm 0.0022 \pm \mathrm{j}\,0.0024)$
$-0.5436 - \mathrm{j}\,0.5150$ $(\pm 0.0023 \pm \mathrm{j}\,0.0022)$	$0.7398 + \mathrm{j}\,0.2869$ $(\pm 0.0022 \pm \mathrm{j}\,0.0023)$	$-0.5402 + \mathrm{j}\,0.7881$ $(\pm 0.0025 \pm \mathrm{j}\,0.0026)$
$0.0120 + \mathrm{j}\,0.9866$ $(\pm 0.0021 \pm \mathrm{j}\,0.0025)$	$0.3672 + \mathrm{j}\,0.4126$ $(\pm 0.0023 \pm \mathrm{j}\,0.0020)$	$0.1809 + \mathrm{j}\,0.2304$ $(\pm 0.0018 \pm \mathrm{j}\,0.0019)$
$-1.0081 - \mathrm{j}\,0.4359$ $(\pm 0.0025 \pm \mathrm{j}\,0.0029)$	$-0.0909 - \mathrm{j}\,0.4222$ $(\pm 0.0021 \pm \mathrm{j}\,0.0022)$	$-0.8883 - \mathrm{j}\,0.4642$ $(\pm 0.0027 \pm \mathrm{j}\,0.0025)$
$-0.2138 - \mathrm{j}\,0.2547$ $(\pm 0.0019 \pm \mathrm{j}\,0.0019)$	$-0.1395 - \mathrm{j}\,0.3628$ $(\pm 0.0020 \pm \mathrm{j}\,0.0017)$	$-0.2468 + \mathrm{j}\,0.0179$ $(\pm 0.0019 \pm \mathrm{j}\,0.0018)$
$-1.1740 + \mathrm{j}\,0.7497$ $(\pm 0.0032 \pm \mathrm{j}\,0.0028)$	$-1.7168 + \mathrm{j}\,0.6883$ $(\pm 0.0037 \pm \mathrm{j}\,0.0041)$	$-0.6180 + \mathrm{j}\,0.6993$ $(\pm 0.0027 \pm \mathrm{j}\,0.0026)$
$-0.6065 - \mathrm{j}\,0.7320$ $(\pm 0.0023 \pm \mathrm{j}\,0.0026)$	$0.1464 + \mathrm{j}\,0.5121$ $(\pm 0.0021 \pm \mathrm{j}\,0.0021)$	$-0.4451 + \mathrm{j}\,0.4104$ $(\pm 0.0024 \pm \mathrm{j}\,0.0023)$
$0.0466 - \mathrm{j}\,0.5740$ $(\pm 0.0019 \pm \mathrm{j}\,0.0021)$	$0.8388 - \mathrm{j}\,0.9316$ $(\pm 0.0029 \pm \mathrm{j}\,0.0029)$	$0.1457 - \mathrm{j}\,0.7707$ $(\pm 0.0020 \pm \mathrm{j}\,0.0022)$
$1.0872 + \mathrm{j}\,1.0016$ $(\pm 0.0035 \pm \mathrm{j}\,0.0032)$	$-0.8175 + \mathrm{j}\,1.3143$ $(\pm 0.0035 \pm \mathrm{j}\,0.0032)$	$1.8308 + \mathrm{j}\,0.5451$ $(\pm 0.0039 \pm \mathrm{j}\,0.0037)$

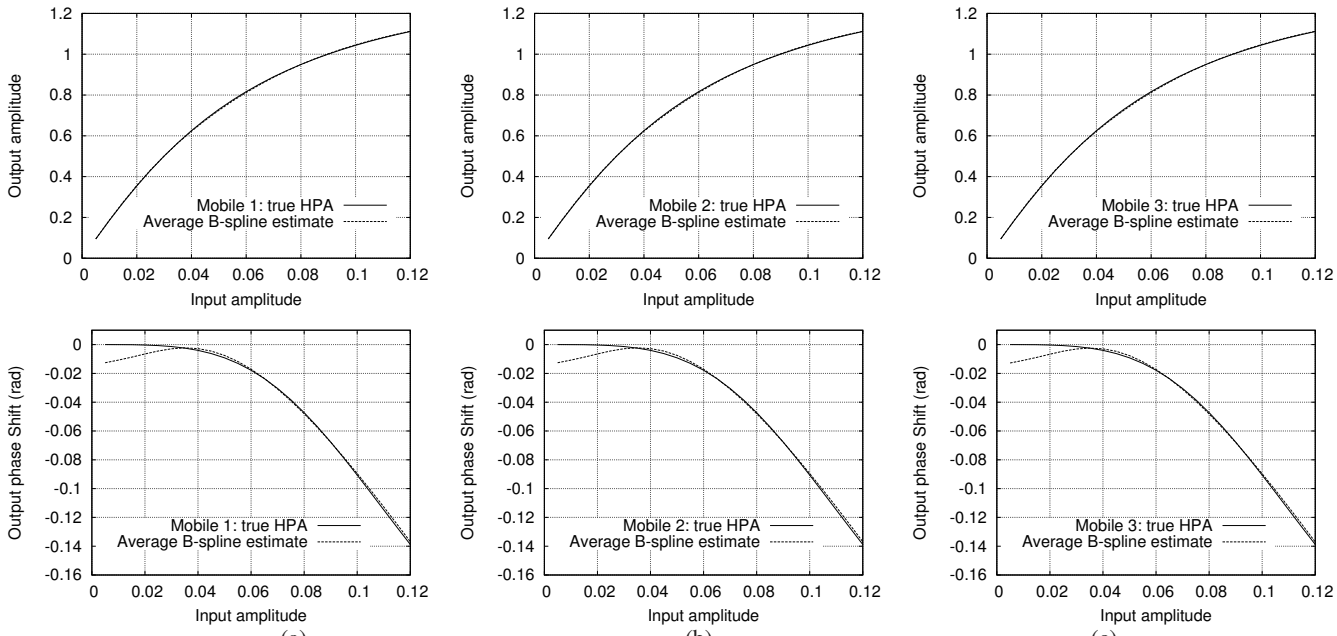

Fig. 3. The unique parametrized true NHPA's mapping $\zeta_m \Psi(\cdot)$ in comparison with the BSNN estimated NHPA mapping $\widehat{\zeta}_m \widehat{\Psi}(\cdot)$ averaged over 100 identification runs given OBO of 3 dB and average SNR of 20 dB: (a) MU 1, (b) MU 2, and (c) MU 3.

capable of attaining the unbiased and accurate estimates of the MIMO CIR matrix and the BSNN models of the NHPAs at the MUs' transmitters as well as the BSNN inverse models of the transmitters' NHPAs. For this purpose, we consider the true MIMO CIR matrix $\boldsymbol{H}$ and the true NHPAs' weights $\boldsymbol{\zeta}$ for a unique parametrized multiuser nonlinear frequency-selective MIMO channel as given in Table III, where for the clear representation purpose, each row of $\boldsymbol{H}$ is re-arranged into three subrows:

$$\boldsymbol{h}_{l,1}^{\mathrm{T}}\ \boldsymbol{h}_{l,2}^{\mathrm{T}}\ \boldsymbol{h}_{l,3}^{\mathrm{T}} \Rightarrow \begin{matrix} \boldsymbol{h}_{l,1}^{\mathrm{T}} \\ \boldsymbol{h}_{l,2}^{\mathrm{T}} \\ \boldsymbol{h}_{l,3}^{\mathrm{T}} \end{matrix},\ 1 \le l \le L. \tag{57}$$

In this set of experiments, we set the NHPAs' OBO to 3 dB and the system's average SNR to 20 dB. The BSNN based identification scheme and the nonlinear STE based MUD presented in Section III are applied to this multiuser nonlinear frequency-selective MIMO uplink. The results are obtained over 100 independent identification experiments.

1) Accuracy of MIMO CIR matrix estimate: The MIMO CIR matrix estimate $\widehat{\boldsymbol{H}}$ obtained by the proposed BSNN based estimator is tabulated in Table IV, where the estimation results are presented as average estimate with standard deviation. Observe from Table IV that the BSNN based estimate $\widehat{\boldsymbol{H}}$ is a very accurate unbiased estimate of the true MIMO CIR matrix $\boldsymbol{H}$ given in Table III, with very small estimation error standard deviations.

2) Accuracy of BSNN estimates of NHPAs: The estimated NHPAs' weighting vector $\widehat{\boldsymbol{\zeta}}$ obtained by the BSNN based estimator closely matches the true NHPAs' weighting vector $\boldsymbol{\zeta}$, as can be clearly seen from Table IV. Fig. 3 compares the

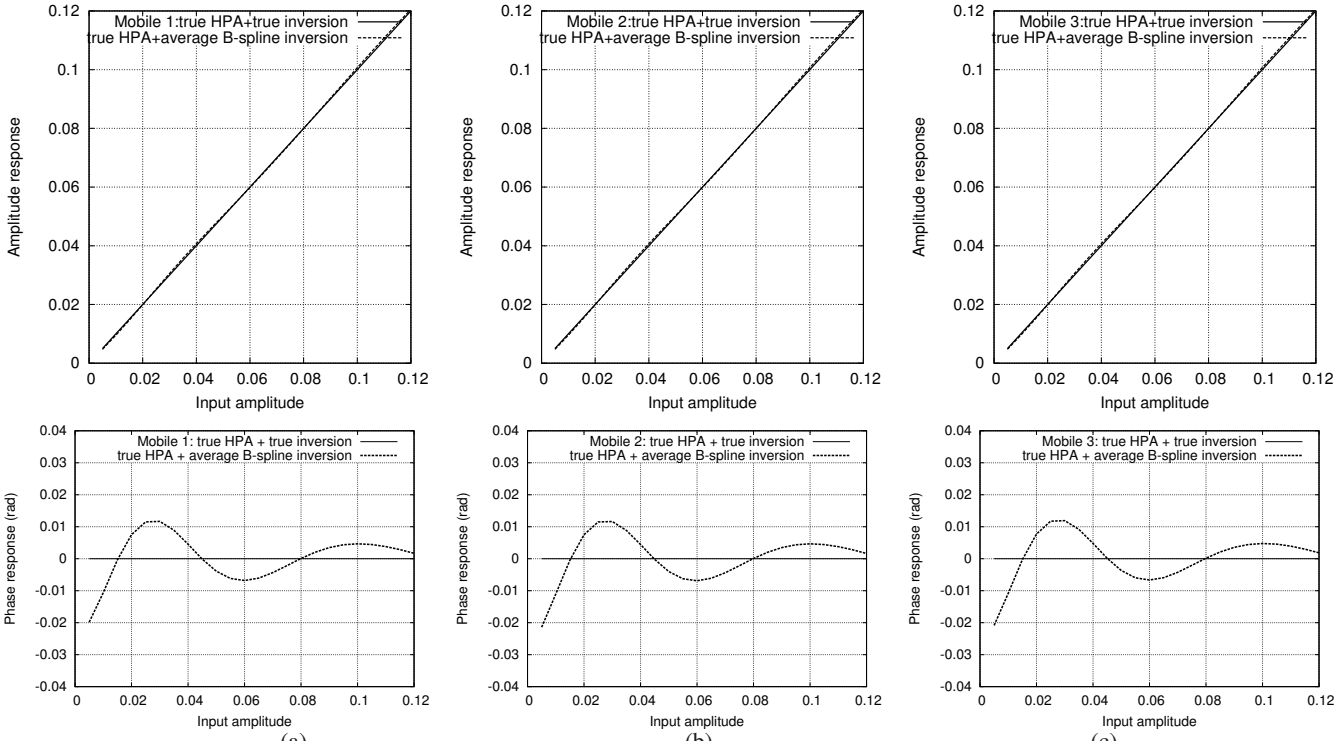

Fig. 4. The ideal combined response of the true NHPA $\zeta_m \Psi(\cdot)$ and its true inversion $\Phi_m(\cdot)$ in comparison with the combined response of the true HPA $\zeta_m \Psi(\cdot)$ and the estimated BSNN inversion $\widehat{\Phi}_m(\cdot)$ averaged over 100 identification runs given the OBO of 3 dB and the average SNR of 20 dB: (a) MU 1, (b) MU 2, and (c) MU 3.

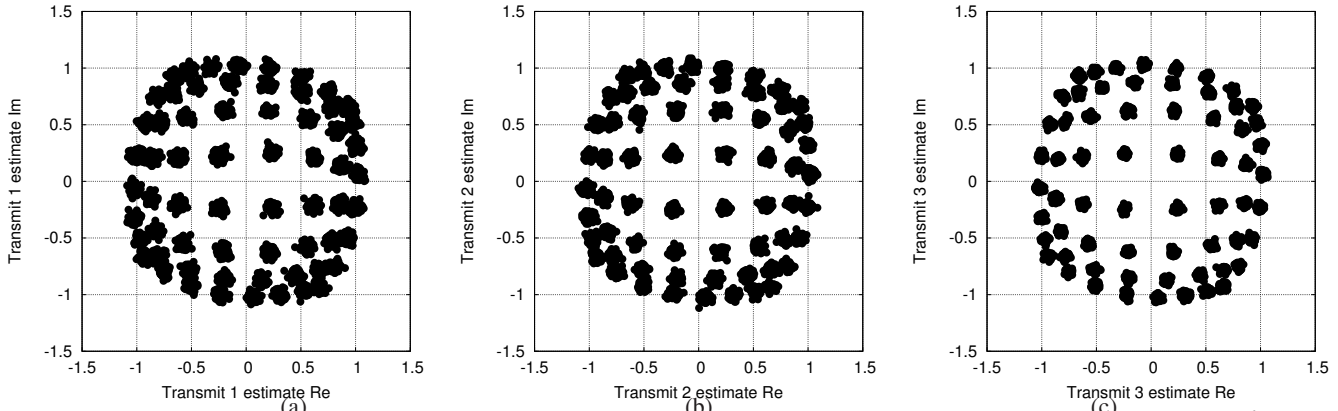

Fig. 5. Detected MUs' transmitted signals, $z_m(k)$, $1 \leq k \leq 3$, by the MMSE space-time equalizer using the estimated MIMO CIR matrix $\widehat{\boldsymbol{H}}$ obtained by the BSNN based estimation scheme at a typical identification run given OBO of 3 dB and average SNR of 20 dB: (a) $\widehat{z}_1(k-5)$, (b) $\widehat{z}_2(k-5)$, and (c) $\widehat{z}_3(k-5)$.

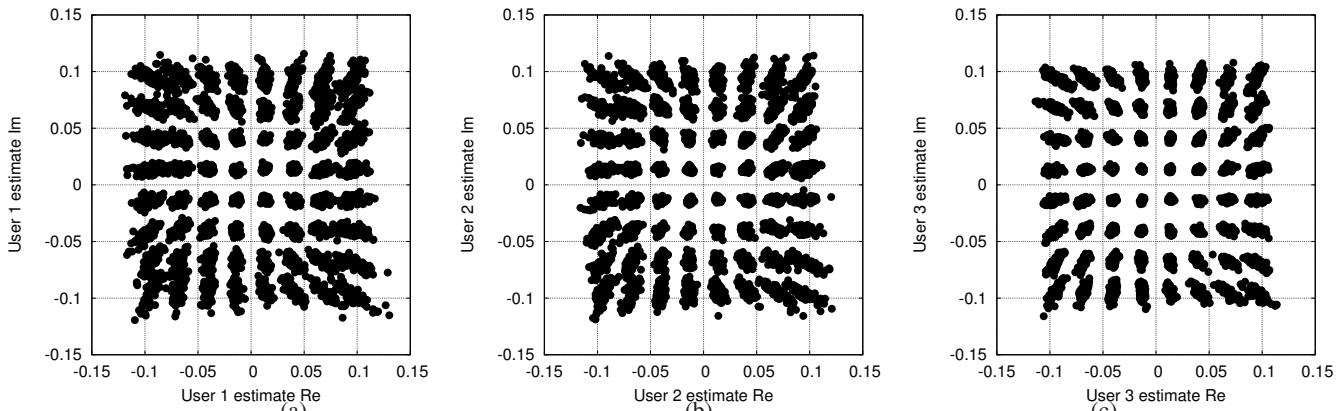

Fig. 6. Detected MUs' transmitted 64-QAM symbols, $s_m(k)$, $1 \leq m \leq 3$, by the nonlinear MMSE space-time equalizer based MUD using the estimated $\widehat{\boldsymbol{H}}$ and the BSNN inversions $\widehat{\Phi}_m(\cdot)$ obtained by the BSNN based estimation scheme at a typical identification run given OBO of 3 dB and average SNR of 20 dB: (a) $\widehat{s}_1(k-5)$, (b) $\widehat{s}_2(k-5)$, and (c) $\widehat{s}_3(k-5)$.

TABLE V
LINEAR LS ESTIMATE FOR THE TRUE MIMO CHANNEL MATRIX $\boldsymbol{H}$ OF MULTIUSER NONLINEAR FREQUENCY-SELECTIVE MIMO CHANNEL. THE OBO IS 3 DB AND THE AVERAGE SNR IS 20 DB. THE RESULTS ARE OBTAINED OVER 100 INDEPENDENT RUNS, AND ARE PRESENTED AS: AVERAGE ESTIMATE (±STANDARD DEVIATION).

Linear LS estimate (43) $\widehat{\boldsymbol{H}}^{[0]} = \left[[(\widehat{\boldsymbol{h}}_{l,m}^{[0]})^{\mathrm{T}}, 1 \le m \le 3] \in \mathbb{C}^{1\times 9} \right], 1 \le l \le 5$		
$10.5037 - \mathrm{j}0.9566$ ($\pm 0.1368 \pm \mathrm{j}0.1215$)	$6.0066 + \mathrm{j}11.1395$ ($\pm 0.1139 \pm \mathrm{j}0.1278$)	$3.1474 - \mathrm{j}8.4847$ ($\pm 0.1183 \pm \mathrm{j}0.1432$)
$10.4996 - \mathrm{j}0.9583$ ($\pm 0.1260 \pm \mathrm{j}0.1184$)	$4.3061 + \mathrm{j}3.8817$ ($\pm 0.1028 \pm \mathrm{j}0.1349$)	$17.5568 - \mathrm{j}4.6649$ ($\pm 0.1413 \pm \mathrm{j}0.1154$)
$10.4797 - \mathrm{j}0.9443$ ($\pm 0.1309 \pm \mathrm{j}0.1221$)	$-2.7116 - \mathrm{j}14.6836$ ($\pm 0.1099 \pm \mathrm{j}0.1346$)	$-6.0624 - \mathrm{j}4.6978$ ($\pm 0.1301 \pm \mathrm{j}0.1274$)
$3.3110 - \mathrm{j}1.6360$ ($\pm 0.1179 \pm \mathrm{j}0.1248$)	$11.2270 + \mathrm{j}3.9250$ ($\pm 0.1197 \pm \mathrm{j}0.1057$)	$-5.2922 + \mathrm{j}9.2857$ ($\pm 0.1101 \pm \mathrm{j}0.1009$)
$-6.3678 - \mathrm{j}1.8751$ ($\pm 0.1130 \pm \mathrm{j}0.1107$)	$-3.3683 + \mathrm{j}2.2609$ ($\pm 0.1089 \pm \mathrm{j}0.1069$)	$-2.6414 - \mathrm{j}3.2991$ ($\pm 0.1190 \pm \mathrm{j}0.1083$)
$12.9403 - \mathrm{j}15.0679$ ($\pm 0.1186 \pm \mathrm{j}0.1057$)	$-6.1817 + \mathrm{j}10.7898$ ($\pm 0.1185 \pm \mathrm{j}0.1090$)	$8.8338 - \mathrm{j}4.3716$ ($\pm 0.1151 \pm \mathrm{j}0.1004$)
$-0.7102 + \mathrm{j}7.0421$ ($\pm 0.0820 \pm \mathrm{j}0.0850$)	$0.3957 - \mathrm{j}2.2574$ ($\pm 0.0841 \pm \mathrm{j}0.0811$)	$-5.1194 - \mathrm{j}5.4768$ ($\pm 0.0831 \pm \mathrm{j}0.0859$)
$-6.1958 - \mathrm{j}4.8718$ ($\pm 0.0828 \pm \mathrm{j}0.0769$)	$8.0256 + \mathrm{j}2.3119$ ($\pm 0.0864 \pm \mathrm{j}0.0794$)	$-4.9283 + \mathrm{j}8.7706$ ($\pm 0.0720 \pm \mathrm{j}0.0744$)
$1.0621 + \mathrm{j}10.3361$ ($\pm 0.0720 \pm \mathrm{j}0.0977$)	$4.2427 + \mathrm{j}3.9866$ ($\pm 0.0853 \pm \mathrm{j}0.0825$)	$2.1151 + \mathrm{j}2.2470$ ($\pm 0.0852 \pm \mathrm{j}0.0763$)
$-11.0036 - \mathrm{j}3.6347$ ($\pm 0.1145 \pm \mathrm{j}0.1149$)	$-1.3577 - \mathrm{j}4.3382$ ($\pm 0.1102 \pm \mathrm{j}0.1123$)	$-9.7641 - \mathrm{j}4.0338$ ($\pm 0.1133 \pm \mathrm{j}0.1178$)
$-2.4888 - \mathrm{j}2.4631$ ($\pm 0.1145 \pm \mathrm{j}0.1159$)	$-1.7963 - \mathrm{j}3.6669$ ($\pm 0.1038 \pm \mathrm{j}0.1204$)	$-2.5767 + \mathrm{j}0.4046$ ($\pm 0.1148 \pm \mathrm{j}0.1288$)
$-11.6016 + \mathrm{j}8.9805$ ($\pm 0.1205 \pm \mathrm{j}0.1031$)	$-17.3620 + \mathrm{j}8.8487$ ($\pm 0.1294 \pm \mathrm{j}0.0982$)	$-5.8282 + \mathrm{j}7.9291$ ($\pm 0.1160 \pm \mathrm{j}0.1159$)
$-7.0849 - \mathrm{j}7.1076$ ($\pm 0.1373 \pm \mathrm{j}0.1471$)	$2.0245 + \mathrm{j}5.2562$ ($\pm 0.1213 \pm \mathrm{j}0.1403$)	$-4.2894 + \mathrm{j}4.7420$ ($\pm 0.1373 \pm \mathrm{j}0.1430$)
$-0.0367 - \mathrm{j}6.0743$ ($\pm 0.1442 \pm \mathrm{j}0.1321$)	$7.9309 - \mathrm{j}10.5453$ ($\pm 0.1296 \pm \mathrm{j}0.1181$)	$0.7785 - \mathrm{j}8.2380$ ($\pm 0.1362 \pm \mathrm{j}0.1305$)
$12.3697 + \mathrm{j}9.4738$ ($\pm 0.1281 \pm \mathrm{j}0.1386$)	$-7.3169 + \mathrm{j}14.5895$ ($\pm 0.1263 \pm \mathrm{j}0.1260$)	$19.7142 + \mathrm{j}3.9957$ ($\pm 0.1485 \pm \mathrm{j}0.1206$)
Normalized linear LS estimate (44) $\widehat{\boldsymbol{H}}^{[0]}$		
1	$0.4713 + \mathrm{j}1.1035$ ($\pm 0.0197 \pm \mathrm{j}0.0207$)	$0.3702 - \mathrm{j}0.7741$ ($\pm 0.0151 \pm \mathrm{j}0.0149$)
1	$0.3733 + \mathrm{j}0.4038$ ($\pm 0.0126 \pm \mathrm{j}0.0145$)	$1.6986 - \mathrm{j}0.2892$ ($\pm 0.0226 \pm \mathrm{j}0.0219$)
1	$-0.1314 - \mathrm{j}1.4130$ ($\pm 0.0200 \pm \mathrm{j}0.0196$)	$-0.5338 - \mathrm{j}0.4964$ ($\pm 0.0147 \pm \mathrm{j}0.0144$)
$0.3267 - \mathrm{j}0.1260$ ($\pm 0.0114 \pm \mathrm{j}0.0124$)	$1.0263 + \mathrm{j}0.4672$ ($\pm 0.0139 \pm \mathrm{j}0.0164$)	$-0.5796 + \mathrm{j}0.8312$ ($\pm 0.0180 \pm \mathrm{j}0.0150$)
$-0.5853 - \mathrm{j}0.2320$ ($\pm 0.0125 \pm \mathrm{j}0.0114$)	$-0.3376 + \mathrm{j}0.1845$ ($\pm 0.0105 \pm \mathrm{j}0.0101$)	$-0.2210 - \mathrm{j}0.3344$ ($\pm 0.0124 \pm \mathrm{j}0.0109$)
$1.3533 - \mathrm{j}1.3159$ ($\pm 0.0222 \pm \mathrm{j}0.0238$)	$-0.6771 + \mathrm{j}0.9686$ ($\pm 0.0170 \pm \mathrm{j}0.0162$)	$0.8734 - \mathrm{j}0.3384$ ($\pm 0.0158 \pm \mathrm{j}0.0157$)
$-0.1277 + \mathrm{j}0.6588$ ($\pm 0.0090 \pm \mathrm{j}0.0095$)	$0.0568 - \mathrm{j}0.2097$ ($\pm 0.0090 \pm \mathrm{j}0.0087$)	$-0.4362 - \mathrm{j}0.5612$ ($\pm 0.0123 \pm \mathrm{j}0.0129$)
$-0.5432 - \mathrm{j}0.5136$ ($\pm 0.0112 \pm \mathrm{j}0.0100$)	$0.7381 + \mathrm{j}0.2876$ ($\pm 0.0110 \pm \mathrm{j}0.0122$)	$-0.5411 + \mathrm{j}0.7859$ ($\pm 0.0130 \pm \mathrm{j}0.0135$)
$0.0124 + \mathrm{j}0.9874$ ($\pm 0.0117 \pm \mathrm{j}0.0131$)	$0.3676 + \mathrm{j}0.4135$ ($\pm 0.0098 \pm \mathrm{j}0.0095$)	$0.1810 + \mathrm{j}0.2307$ ($\pm 0.0089 \pm \mathrm{j}0.0079$)
$-1.0077 - \mathrm{j}0.4378$ ($\pm 0.0129 \pm \mathrm{j}0.0110$)	$-0.0909 - \mathrm{j}0.4213$ ($\pm 0.0111 \pm \mathrm{j}0.0114$)	$-0.8872 - \mathrm{j}0.4649$ ($\pm 0.0140 \pm \mathrm{j}0.0160$)
$-0.2139 - \mathrm{j}0.2541$ ($\pm 0.0095 \pm \mathrm{j}0.0092$)	$-0.1381 - \mathrm{j}0.3619$ ($\pm 0.0101 \pm \mathrm{j}0.0120$)	$-0.2469 + \mathrm{j}0.0160$ ($\pm 0.0112 \pm \mathrm{j}0.0123$)
$-1.1747 + \mathrm{j}0.7511$ ($\pm 0.0186 \pm \mathrm{j}0.0184$)	$-1.7188 + \mathrm{j}0.6895$ ($\pm 0.0207 \pm \mathrm{j}0.0224$)	$-0.6193 + \mathrm{j}0.7008$ ($\pm 0.0141 \pm \mathrm{j}0.0147$)
$-0.6078 - \mathrm{j}0.7320$ ($\pm 0.0139 \pm \mathrm{j}0.0154$)	$0.1459 + \mathrm{j}0.5137$ ($\pm 0.0115 \pm \mathrm{j}0.0122$)	$-0.4458 + \mathrm{j}0.4108$ ($\pm 0.0168 \pm \mathrm{j}0.0157$)
$0.0489 - \mathrm{j}0.5741$ ($\pm 0.0140 \pm \mathrm{j}0.0125$)	$0.8400 - \mathrm{j}0.9277$ ($\pm 0.0208 \pm \mathrm{j}0.0212$)	$0.1446 - \mathrm{j}0.7714$ ($\pm 0.0160 \pm \mathrm{j}0.0146$)
$1.0901 + \mathrm{j}1.0022$ ($\pm 0.0253 \pm \mathrm{j}0.0238$)	$-0.8170 + \mathrm{j}1.3186$ ($\pm 0.0184 \pm \mathrm{j}0.0210$)	$1.8319 + \mathrm{j}0.5464$ ($\pm 0.0247 \pm \mathrm{j}0.0260$)

BSNN estimated NHPA mapping $\widehat{\zeta}_m\widehat{\Psi}(\cdot)$ averaged over 100 independent runs with the true HHPA's mapping $\zeta_m\Psi(\cdot)$. It can be seen that the amplitude response of $\widehat{\zeta}_m\widehat{\Psi}(\cdot)$ is almost identical to the true NHPA's amplitude response, and the estimation error of the phase response of $\widehat{\zeta}_m\widehat{\Psi}(\cdot)$ is no more than 0.01 radian.

3) Accuracy of BSNN inversions of NHPAs: We now verify the accuracy of the BSNN inversion estimates $\widehat{\Phi}_m(\cdot)$, $1 \le m \le 3$, obtained by the proposed BSNN inverting scheme. From (50), it can be seen that the ideal combined response of the true NHPA $\zeta_m\Psi(\cdot)$ and its true inversion $\Phi_m(\cdot)$ satisfies $s_m = \Phi_m\big(\zeta_m\Psi(s_m)\big)$. Therefore, we generate the combined response of the true NHPA $\zeta_m\Psi(\cdot)$ and its BSNN estimated inversion $\widehat{\Phi}_m(\cdot)$, and compare this combined response with the ideal combined response in Fig. 4. Observe that the combined response $\widehat{\Phi}_m\big(\zeta_m\Psi(\cdot)\big)$ matches well the ideal combined response $\Phi_m\big(\zeta_m\Psi(\cdot)\big)$, and we have

$$\widehat{\Phi}_m\big(\zeta_m\Psi(s_m)\big) \approx s_m, \ 1 \le m \le 3. \tag{58}$$

More specifically, the combined magnitude response is almost identical to the ideal combined magnitude response, while the error between the combined phase response and the ideal combined phase response is no more than 0.02 radian. This clearly demonstrates the accuracy of our proposed BSNN inversion scheme based on the noisy pseudo training data.

4) Overall effectiveness of BSNN based estimation procedure: To further illustrate the overall effectiveness of our design, the estimated MIMO CIR matrix $\widehat{\boldsymbol{H}}$ and the BSNN inversions $\widehat{\Phi}_m(\cdot)$, $1 \le m \le 3$, of the transmitters' NHPAs, obtained by the BSNN based estimation procedure in a typical identification run, are used to constructed the nonlinear STE based MUD. Fig. 5 depicts the detected MUs' transmitted signals $z_m(k)$, $1 \le m \le 3$, by the MMSE space-time equalizer. The MUs' transmitted 64-QAM data are detected by passing the detected transmitted signals $\widehat{z}_m(k-5)$ through the estimated BSNN inversion $\widehat{\Phi}_m(\cdot)$ to compensate for the distortion of the transmitters' NHPAs, for $1 \le m \le 3$, which are shown in Fig. 6.

5) Empirical evidence of unbiasedness and efficiency: The estimation results of Table IV as well as Figs. 3 to 6 clearly demonstrate the accuracy and efficiency of our proposed BSNN based estimation procedure for the multiuser nonlinear frequency-selective MIMO uplink. Specifically, these empirical results show that the estimated MIMO CIR matrix $\widehat{\boldsymbol{H}}$ is an unbiased and accurate estimate for the true MIMO CIR matrix $\boldsymbol{H}$, while the identified BSNN inversions $\widehat{\Phi}_m(\cdot)$ are unbiased and accurate estimates for the true NHPAs' inversion mappings $\Phi_m(\cdot)$, for $1 \le m \le M$. Since we only set the numbers of iterations in both the outer loop and the inner loop to 2, the fast convergence of our proposed iterative ALS procedure is self-evident.

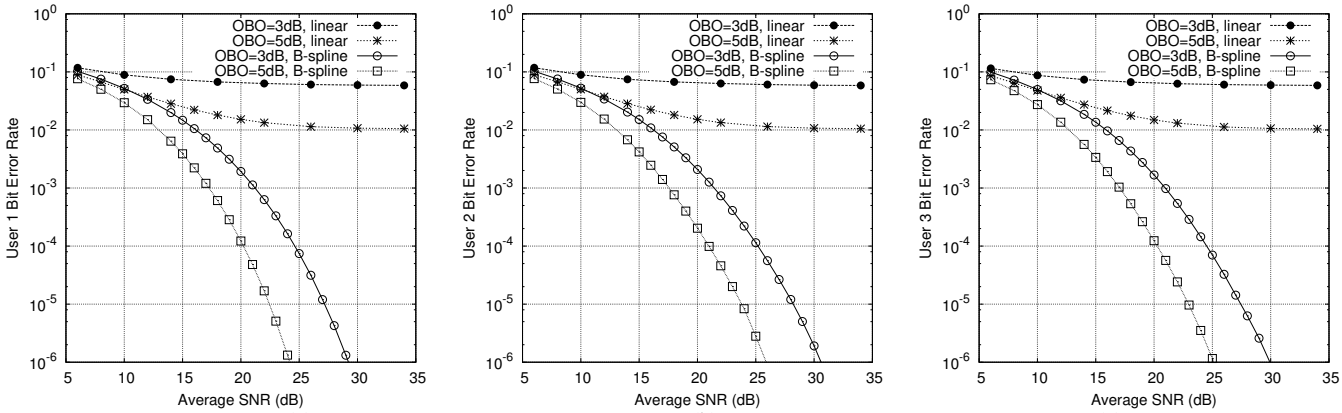

Fig. 7. Average bit error rate performance comparison for the proposed BSNN assisted nonlinear space-time equalization based MUD and the standard MMSE linear space-time equalization scheme over 100 MIMO channel realizations, given the two OBO values of 3 dB and 5 dB: (a) MU 1, (b) MU 2, and (c) MU3.

In Subsection III-B3, we point out that the linear LS estimate $\widehat{\boldsymbol{H}}^{[0]}$ (43) is an estimate of the MIMO CIR matrix scaled by the MUs' NHPAs' complex-valued gains, and its normalized version (44) is an unbiased estimate of the true MIMO CIR matrix $\boldsymbol{H}$, although its estimation accuracy may be poor. We now supply the empirical evidence to support this analysis. In Table V, we list the linear LS estimate $\widehat{\boldsymbol{H}}^{[0]}$ (43) and its normalized version (44). By comparing this normalized linear LS estimate with the true MIMO CIR matrix $\boldsymbol{H}$ given in Table III, it is clear that the normalized $\widehat{\boldsymbol{H}}^{[0]}$ is an unbiased estimate of the true MIMO CIR matrix $\boldsymbol{H}$. Since this normalized linear LS estimate is used as the initial estimate of the MIMO channel matrix in our iterative ALS estimator, it is not surprising that our iterative ALS estimator converges very fast. Moreover, by comparing this normalized linear LS estimate $\widehat{\boldsymbol{H}}^{[0]}$ with our BSNN approach based estimate $\widehat{\boldsymbol{H}}$ of Table IV, it can be seen that with only two iterations, the estimation accuracy of the latter is significantly better than that of the former, since the estimation error standard deviations of $\widehat{\boldsymbol{H}}$ are around six times smaller than those of the initial $\widehat{\boldsymbol{H}}^{[0]}$. Hence our estimation results also offer the empirical evidence to support the analysis of Subsection III-B3.

4.3 Bit error rate performance

We now evaluate the ultimate performance metric of our design, namely, its achievable bit error rate (BER). We consider the rich scattered wireless environment, where the entries of the MIMO CIR matrix follow the independent complex Gaussian distribution $\mathcal{CN}(0,1)$. In the simulation, we randomly generate the multiuser MIMO channel matrix $\boldsymbol{H}$ by drawing its coefficients $h_{i,l,m}$ for $0 \le i \le n_H - 1$, $1 \le l \le L$ and $1 \le m \le M$ from $\mathcal{CN}(0,1)$. A total of 100 channel realizations or MIMO CIR matrices are drawn. For each MIMO channel realization, joint estimates of $\boldsymbol{H}$, $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$ are obtained using the identification algorithms of Section III with $K = 1000$ training data. Based on the estimated $\widehat{\boldsymbol{H}}$, $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\zeta}}$, the BSNN assisted nonlinear STE based MUD is implemented and 10^8 64-QAM data symbols are transmitted by each MU for the BS to calculate the BER. The average BER performance over the generated 100 MIMO channel realizations achieved by our proposed BSNN assisted nonlinear STE based MUD are depicted in Fig. 7, given the two OBO values of 3 dB and 5 dB.

It is worth recapping that our proposed scheme is the first effective and practical multiuser nonlinear STE scheme for the single-carrier multiuser nonlinear frequency-selective MIMO uplink, and there exists no other effective and practical nonlinear MUD schemes in the literature to compare with. The existing STE based MUD schemes for single-carrier multiuser frequency-selective MIMO uplink typically assume a linear frequency-selective MIMO channel, which clearly no longer work for the single-carrier multiuser nonlinear frequency-selective MIMO uplink. To demonstrate this fact, we also implement the linear STE for this single-carrier multiuser nonlinear frequency-selective MIMO uplink. Specifically, we first estimate the equivalent linear MIMO channel matrix $\widehat{\boldsymbol{H}}^{[0]}$ using the linear LS estimate of (43), and then design the linear MMSE space-time equalizer based on $\widehat{\boldsymbol{H}}^{[0]}$. The BER performance achieved by this linear space-time equalizer are also shown in Fig. 7 for the comparison with our BSNN assisted nonlinear STE based MUD. Not surprisingly, this linear space-time equalizer exhibits a high BER floor at the BER level of 10^{-2} even under the OBO of 5 dB, because it cannot compensate for the nonlinear distortion of the transmitters' NHPAs.

5. Concluding Remarks

A BSNN assisted space-time equalization based MUD has been proposed for the single-carrier multiuser nonlinear frequency-selective MIMO uplink employing high-throughput QAM transmission and with NHPAs at MUs' transmitters. First, we have developed a unique parametrization of the multiuser frequency-selective MIMO CIR matrix and the MUs' nonlinear transmitters as well as a BSNN parametrization of the transmitter's NHPA. Second, we have proposed a highly efficient and accurate iterative ALS estimation procedure to jointly estimate the MIMO CIR matrix and the BSNN models of the MUS' NHPAs. Third, the BSNN inverse models for the MUs' NHPAs have also been estimated. Based on the estimated MIMO CIR matrix and the constructed BSNN inversion models of the NHPAs, a BSNN assisted space-time equalization based MUD has been implemented for the single-carrier multiuser nonlinear frequency-selective MIMO

uplink. Simulation results have demonstrated that our proposed iterative ALS procedure converges very fast to the unbiased and accurate estimates of both the dispersive MIMO CIR matrix and the MUs' NHPAs. Simulation results have also confirmed the effectiveness of the BSNN assisted space-time equalization based MUD scheme, in terms of achievable BER performance.

References

[1] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Communications*, vol. 49, no. 12, pp. 2198–2206, Dec. 2001.

[2] Y. Jiang, J. Li, and W. W. Hager, "Joint transceiver design for MIMO communications using geometric mean decomposition," *IEEE Trans. Signal Processing*, vol. 53, no. 10, pp. 3791–3803, Oct. 2005.

[3] S. Chen, A. Livingstone, and L. Hanzo, "Minimum bit-error rate design for space-time equalization-based multiuser detection," *IEEE Trans. Communications*, vol. 54, no. 5, pp. 824–832, May 2006.

[4] 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 Processing*, vol. 54, no. 8, pp. 3090–3098, Aug. 2006.

[5] S. Chen, A. Livingstone, H.-Q. Du, and L. Hanzo, "Adaptive minimum symbol error rate beamforming assisted detection for quadrature amplitude modulation," *IEEE Trans. Wireless Communications*, vol. 7, no. 4, pp. 1140–1145, Apr. 2008.

[6] S. Chen, W. Yao, and L. Hanzo, "Semi-blind adaptive spatial equalisation for MIMO systems with high-order QAM signalling," *IEEE Trans. Wireless Communications*, vol. 7, no. 11, pp. 4486–4491, Nov. 2008.

[7] W. Yao, S. Chen, and L. Hanzo, "A transceiver design based on uniform channel decomposition and MBER vector perturbation," *IEEE Trans. Vehicular Technology*, vol. 59, no. 6, pp. 3153–3159, Jul. 2010.

[8] W. Yao, S. Chen, and L. Hanzo, "Generalised MBER-based vector precoding design for multiuser transmission," *IEEE Trans. Vehicular Technology*, vol. 60, no. 2, pp. 739–745, Feb. 2011.

[9] S. Sugiura, S. Chen, and L. Hanzo, "Generalized space-time shift keying designed for flexible diversity-, multiplexing- and complexity-tradeoffs," *IEEE Trans. Wireless Communications*, vol. 10, no. 4, pp. 1144–1153, Apr. 2011.

[10] 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 Communications*, vol. 18, no. 9, pp. 4326–4343, Sep. 2019.

[11] A. A. M. Saleh, "Frequency-independent and frequency-dependent nonlinear models of TWT amplifiers," *IEEE Trans. Communications*, vol. COM-29, no. 11, pp. 17151720, Nov. 1981.

[12] M. Honkanen and S.-G. Häggman, "New aspects on nonlinear power amplifier modeling in radio communication system simulations," in *Proc. 8th International Symposium on Personal, Indoor and Mobile Radio Communications* (Helsinki, Finland), Sep. 1-4, 1997, pp. 844–848.

[13] 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. Microwave Theory and Techniques*, vol. 46, no. 12, pp. 2531–2540, Dec. 1998.

[14] C.-S. Choi, Y. Shoji, H. Harada, R. Funada, S. Kato, K. Maruhashi, I. Toyoda, and K. Takahashi, "RF impairment models 60 GHz band SYS/PHY simulation," Document IEEE 802.15-06-0477-01-003c, Nov. 2006. https://mentor.ieee.org/802.15/dcn/06/15-06-0477-01-003c-rf-impairment-models-60ghz-band-sysphy-simulation.pdf

[15] *60 GHz impairments modeling*, Document IEEE 802.11-09/1213r1, Nov. 2009.

[16] L. Hanzo, S. X. Ng, T. Keller, and W. Webb, *Quadrature Amplitude Modulation: From Basics to Adaptive Trellis-Coded, Turbo-Equalised and Space-Time Coded OFDM, CDMA and MC-CDMA Systems*. Chichester, UK: John Wiley, 2004.

[17] L. Ding, G. T. Zhou, D. R. Morgan, Z. Ma, J. S. Kenney, J. Kim, and C. R. Giardina, "A robust digital baseband predistorter constructed using memory polynomials," *IEEE Trans. Communications*, vol. 52, no. 1, pp. 159–165, Jan. 2004.

[18] D. Zhou and V. E. DeBrunner, "Novel adaptive nonlinear predistorters based on the direct learning algorithm," *IEEE Trans. Signal Processing*, vol. 55, no. 1, pp. 120–133, Jan. 2007.

[19] 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. Vehicular Technology*, vol. 57, no. 2, pp. 882–892, Mar. 2008.

[20] S. Choi, E.-R. Jeong, and Y. H. Lee, "Adaptive predistortion with direct learning based on piecewise linear approximation of amplifier nonlinearity," *IEEE J. Selected Topics in Signal Processing*, vol. 3, no. 3, pp. 397–404, Jun. 2009.

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

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

[23] 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 and Systems I*, vol. 60, no. 6, pp. 1584–1594, Jun. 2013.

[24] S. Chen, S. X. Ng, E. Khalaf, A. Morfeq, and N. Alotaib, "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, 2021.

[25] C. A. R. Fernandes, *Nonlinear MIMO Communication Systems: Channel Estimation and Information Recovery Using Volterra Models*. PhD These, Université de Nice Sophia Antipolis, France, 2009.

[26] C. A. R. Fernandes, J. C. M. Mota, and G. Favier, "MIMO Volterra modeling for nonlinear communication channels," *J. Brazilian Society Neural Networks*, vol. 8, no. 2, pp. 71–92, 2010.

[27] C. A. R. Fernandes, G. Favier, and J. C. M. Mota, "PARAFAC-based channel estimation and data recovery in nonlinear MIMO spread spectrum communication systems," *Signal Processing*, vol. 91, no. 2, pp. 311–322, Feb. 2011.

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

[29] C. De Boor, *A Practical Guide to Splines*. New York: Spring Verlag, 1978.

[30] C. J. Harris, X. Hong, and Q. Gan, *Adaptive Modelling, Estimation and Fusion from Data: A Neurofuzzy Approach*. Berlin: Springer-Verlag, 2002.

[31] X. Hong and S. Chen, "Modeling of complex-valued Wiener systems using B-spline neural network," *IEEE Trans. Neural Networks*, vol 22, no. 5, pp. 818–825, May 2011.

[32] X. Hong, R. J. Mitchell, and S. Chen, "Modeling and control of Hammerstein system using B-spline approximation and the inverse of De Boor algorithm," *Int. J. Systems Science*, vol. 43, no. 10, pp. 1976–1984, 2012.

[33] X. Hong and S. Chen, "The system identification and control of Hammerstein system using non-uniform rational B-spline neural network and particle swarm optimization," *Neurocomputing*, vol. 82, pp.216–223, 2012.

[34] S. Chen, X. Hong, J. B. Gao, and C. J. Harris, "Complex-valued B-spline neural networks for modeling and inverting Hammerstein systems," *IEEE Trans. Neural Networks and Learning Systems*, vol. 25, no. 9, pp. 1673–1685, Sep. 2014.

[35] X. Hong, S. Iplikci, S. Chen, and K. Warwick, "A model-based PID controller for Hammerstein systems using B-spline neural networks," *Int. J. Adaptive Control and Signal Processing*, vol. 28, nos.3-5, pp. 412–428, 2014.

[36] 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 Networks and Learning Systems*, vol. 28, no. 12, pp. 2872–2884, Dec. 2017.

[37] J. M. Pena, "B-spline and optimal stability," *Mathematics of Computation*, vol. 66, no. 220, pp. 1555–1560, Oct. 1997.

[38] T. Lyche and J. M. Pena, "Optimally stable multivariate bases," *Advances in Computational Mathematics*, vol. 20, nos. 1-3, pp. 149–159, Jan. 2004.

[39] E. Mainar and J. M. Pena, "Optimal stability of bivariate tensor product B-bases," *J. Numerical Analysis, Industrial and Applied Mathematics*, vol. 6 nos. 3-4, pp. 95–104, 2011.

[40] 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 Communications*, vol. 16, no. 5, pp. 3341–3354, May 2017.

[41] X. Hong, S. Chen, C. J. Harris, and E. Khalaf, "Single-carrier frequency domain equalization for Hammerstein communication systems using complex-valued neural networks," *IEEE Trans. Signal Processing*, vol. 62, no. 17, pp. 4467–4478, Sep. 2014.

[42] S. Chen, X. Hong, E. Khalaf, A. Morfeq, and N. D. Alotaibi, "Adaptive B-spline neural network based nonlinear equalization for high-order QAM systems with nonlinear transmit high power amplifier," *Digital Signal Processing*, vol. 40, pp. 238–249, May 2015.