Nonlinear Equalization of Hammerstein OFDM Systems

Xia Hong, *Senior Member, IEEE*, Sheng Chen, *Fellow, IEEE*, Yu Gong, *Member, IEEE*, and Chris J. Harris

***Abstract*—A practical orthogonal frequency-division multiplexing (OFDM) system can generally be modelled by the Hammerstein system that includes the nonlinear distortion effects of the high power amplifier (HPA) at transmitter. In this contribution, we advocate a novel nonlinear equalization scheme for OFDM Hammerstein systems. We model the nonlinear HPA, which represents the static nonlinearity of the OFDM Hammerstein channel, by a B-spline neural network, and we develop a highly effective alternating least squares algorithm for estimating the parameters of the OFDM Hammerstein channel, including channel impulse response coefficients and the parameters of the B-spline model. Moreover, we also use another B-spline neural network to model the inversion of the HPA's nonlinearity, and the parameters of this inverting B-spline model can easily be estimated using the standard least squares algorithm based on the pseudo training data obtained as a byproduct of the Hammerstein channel identification. Equalization of the OFDM Hammerstein channel can then be accomplished by the usual one-tap linear equalization as well as the inverse B-spline neural network model obtained. The effectiveness of our nonlinear equalization scheme for OFDM Hammerstein channels is demonstrated by simulation results.**

***Index Terms*—B-spline neural networks, De Boor algorithm, equalization, Hammerstein channel, nonlinear high power amplifier, orthogonal frequency-division multiplexing.**

I. Introduction

ORTHOGONAL FREQUENCY-DIVISION MULTIPLEXING (OFDM) [1], [2] has found its way into numerous recent wireless network standards, owing to its virtues of resilience to frequency selective fading channels. Both the modulation and demodulation operations of an OFDM system facilitate convenient low-complexity hardware implementations with the aid of the inverse fast Fourier transform (IFFT) and fast Fourier transform (FFT) operations. However, OFDM signals are notoriously known to have high peak to average power ratios, and a transmitted OFDM signal can be seriously distorted by the high power amplifier (HPA) at the transmitter, which exhibits nonlinear saturation characteristics [3]–[7]. Thus, the nonlinearities of the HPA at transmitter will significantly degrade the OFDM system's achievable bit error rate (BER) performance, and it is particularly critical to be able to effectively compensate for the nonlinear distortions of the HPA in the design of an OFDM wireless system.

An effective approach to compensate for the nonlinear distortions of HPA is to implement a digital predistorter at the transmitter, which is capable of achieving excellent performance, and various predistorter techniques have been developed [8]–[14]. Implementing the predistorter is attractive for the downlink, where the base station (BS) transmitter has the sufficient hardware and software capacities to accommodate the hardware and computational requirements for implementing digital predistorter. In the uplink, however, implementing predistorter at transmitter is difficult, because it is much more challenging for a pocket-size handset to absorb the additional hardware and computational complexity. Alternatively, the nonlinear distortions of the transmitter HPA can be dealt with at the BS receiver, which has sufficient hardware and software resources. With the nonlinear HPA at transmitter, the channel is a complex-valued (CV) nonlinear Hammerstein system and, moreover, the received signal is further impaired by the channel additive white Gaussian noise (AWGN). Therefore, inversion or equalization of the OFDM Hammerstein channel is not a trivial task.

Against this background, in this paper, we develop a highly effective nonlinear equalization scheme for OFDM Hammerstein channels based on the B-spline neural network. The reason that we adopt the B-spline neural network is because it has been demonstrated to be very effective in identification and inversion of CV Wiener systems [14], [15]. Specifically, we propose an efficient alternating least squares (ALS) identification algorithm for estimating the channel impulse response (CIR) coefficients together with the parameters of the B-spline neural network that models the HPA static nonlinearity of the OFDM Hammerstein channel. As linear equalization is naturally accomplished in OFDM systems by a simple yet effective one-tap equalization in frequency domain (FD), nonlinear equalization of the OFDM Hammerstein channel only additionally involves the inversion of the estimated B-spline neural network that models the HPA's nonlinearity. The previous works [14], [15] consider the inversion of a B-spline model as the root finding problem, and develop an iterative root finding procedure based on the Gauss-

Manuscript received October 17, 2013; revised February 14, 2014 and June 12, 2014; accepted September 02, 2014. Date of publication September 08, 2014; date of current version October 06, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Zhengdao Wang.

X. Hong is with the School of Systems Engineering, University of Reading, Reading RG6 6AY, U.K. (e-mail: x.hong@reading.ac.uk).

S. Chen is with the Department of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK. He is also with the Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia (e-mail: sqc@ecs.soton.ac.uk).

Y. Gong is with the School of Electronic, Electrical, and Systems Engineering, Loughborough University, Loughborough LE11 3TU, U.K. (e-mail: Y.Gong@lboro.ac.uk).

C. J. Harris is with the Department of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: cjh@ecs.soton.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2014.2355773

Newton algorithm. In this paper, we propose a much faster and more efficient alternative for inverting the HPA's nonlinearity. In particular, we also use another B-spline neural network to model the inversion of the HPA's nonlinearity. Although the HPA's output at the transmitter is unobservable at the receiver for identifying this inverse model, the pseudo training data obtained as a natural byproduct of the Hammerstein channel identification can be used to estimate the parameters of the inverting B-spline model using the standard least squares (LS) algorithm. Simulation results are presented to demonstrate the effectiveness of our proposed B-spline neural network based nonlinear equalization scheme for OFDM Hammerstein channels.

Throughout our discussion, a CV number $x \in \mathbb{C}$ is represented either by the rectangular form $x = x_R + \mathrm{j} \cdot x_I$, where $\mathrm{j} = \sqrt{-1}$, while $x_R = \Re[x]$ and $x_I = \Im[x]$ denote the real and imaginary parts of x, or alternatively by the polar form $x = |x| \cdot e^{\mathrm{j}\angle^x}$ with $|x|$ denoting the amplitude of x and $\angle^x$ its phase. The vector or matrix transpose and conjugate transpose operators are denoted by $()^{\mathrm{T}}$ and $()^{\mathrm{H}}$, respectively, while $()^{-1}$ stands for the inverse operation and the expectation operator is denoted by $\mathrm{E}\{\}$. Furthermore, $\boldsymbol{I}$ denotes the identity matrix with an appropriate dimension, and $\mathrm{diag}\{x_0, x_1, \cdots, x_{n-1}\}$ is the diagonal matrix with $x_0, x_1, \cdots, x_{n-1}$ as its diagonal elements.

II. OFDM Hammerstein Channel Model

For mathematical analysis tractability, we restrict to a non-turbo detection-decoding based OFDM system, where the detector and the channel decoder operate separately. The OFDM system considered has N subcarriers and employs the M-quadrature amplitude modulation (QAM). The sth FD OFDM symbol vector is expressed as

$$\boldsymbol{X}[s] = [X_0[s] X_1[s] \cdots X_{N-1}[s]]^{\mathrm{T}}, \tag{1}$$

where $X_n[s]$ is the CV data symbol at the nth subcarrier, which takes the value from the M-QAM symbol set

$$\mathbb{X} = \left\{ d(2l - \sqrt{M} - 1) + \mathrm{j} \cdot d(2q - \sqrt{M} - 1), 1 \le l, q \le \sqrt{M} \right\}, \tag{2}$$

with $2d$ being the minimum distance between symbol points. For notational simplification, we will drop the OFDM symbol index $[s]$ in the sequel. Feeding $\boldsymbol{X}$ through the N-point IFFT based modulator yields the time-domain (TD) OFDM signal

$$x_k = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} X_n e^{\frac{\mathrm{j}2\pi nk}{N}}, \quad 0 \le k \le N-1. \tag{3}$$

Define the FFT matrix $\boldsymbol{F} \in \mathbb{C}^{N \times N}$ given by

$$\boldsymbol{F} = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & e^{-\mathrm{j}2\pi/N} & \cdots & e^{-\mathrm{j}2\pi(N-1)/N} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & e^{-\mathrm{j}2\pi(N-1)/N} & \cdots & e^{-\mathrm{j}2\pi(N-1)(N-1)/N} \end{bmatrix}, \tag{4}$$

which has the orthogonal property of $\boldsymbol{F}^{\mathrm{H}}\boldsymbol{F} = \boldsymbol{F}\boldsymbol{F}^{\mathrm{H}} = \boldsymbol{I}$, and let

$$\boldsymbol{x} = [x_0\ x_1\ \cdots\ x_{N-1}]^{\mathrm{T}}. \tag{5}$$

Then, the IFFT based modulation operation (3) can be expressed concisely by

$$\boldsymbol{x} = \boldsymbol{F}^{\mathrm{H}} \boldsymbol{X}. \tag{6}$$

After adding the cyclic prefix (CP) of length N_{cp} to $\boldsymbol{x}$, the resultant TD OFDM signal

$$\bar{\boldsymbol{x}} = \left[x_{-N_{\mathrm{cp}}}\ x_{-N_{\mathrm{cp}}+1}\ \cdots\ x_{-1} \mid \boldsymbol{x}^{\mathrm{T}} \right]^{\mathrm{T}}, \tag{7}$$

in which $x_{-k} = x_{N-k}$ for $1 \le k \le N_{\mathrm{cp}}$, is amplified by the HPA and the actually transmitted TD signal vector

$$\begin{aligned} \bar{\boldsymbol{w}} &= \left[w_{-N_{\mathrm{cp}}}\ w_{-N_{\mathrm{cp}}+1}\ \cdots\ w_{-1} \mid w_0\ w_1\ \cdots\ w_{N-1} \right]^{\mathrm{T}} \\ &= \left[w_{-N_{\mathrm{cp}}}\ w_{-N_{\mathrm{cp}}+1}\ \cdots\ w_{-1} \mid \boldsymbol{w}^{\mathrm{T}} \right]^{\mathrm{T}} \end{aligned} \tag{8}$$

is defined by

$$w_k = \Psi(x_k), \quad -N_{\mathrm{cp}} \le k \le N-1, \tag{9}$$

where $\Psi()$ represents the CV static nonlinearity of the transmitter HPA, and $w_{-k} = w_{N-k}$ for $1 \le k \le N_{\mathrm{cp}}$. We consider the solid state power amplifier [6], [7], whose nonlinearity $\Psi()$ is constituted by the HPA's amplitude response $A(r)$ and phase response $\Upsilon(r)$ given by

$$A(r) = \frac{g_a r}{\left(1 + \left(\frac{g_a r}{A_{\mathrm{sat}}}\right)^{2\beta_a}\right)^{\frac{1}{2\beta_a}}}, \tag{10}$$

$$\Upsilon(r) = \frac{\alpha_\phi r^{q_1}}{1 + \left(\frac{r}{\beta_\phi}\right)^{q_2}}, \tag{11}$$

where r denotes the amplitude of the input to the HPA, g_a is the small gain signal, β_a is the smoothness factor and A_{sat} is the saturation level, while the parameters of the phase response, α_ϕ. β_ϕ, q_1 and q_2, are adjusted to match the specific amplifier's characteristics. The NEC GaAs power amplifier used in the standardization [6], [7] has the the parameter set

$$\begin{aligned} &g_a = 19,\ \beta_a = 0.81,\ A_{\mathrm{sat}} = 1.4; \\ &\alpha_\phi = -48000,\ \beta_\phi = 0.123,\ q_1 = 3.8,\ q_2 = 3.7. \end{aligned} \tag{12}$$

Hence, given the input $x_k = |x_k| \cdot e^{\mathrm{j} \cdot \angle^{x_k}}$, the output of the HPA can be expressed as

$$w_k = A(|x_k|) \cdot e^{\mathrm{j} \cdot \left(\angle^{x_k} + \Upsilon(|x_k|)\right)}. \tag{13}$$

The operating status of the HPA may be specified by the output back-off (OBO), which is defined as the ratio of the maximum output power $P_{\max}$ of the HPA to the average output power P_{o} of the signal at the HPA output, given by

$$\mathrm{OBO} = 10 \cdot \log_{10} \frac{P_{\max}}{P_{\mathrm{o}}}. \tag{14}$$

The smaller OBO is, the more the HPA is operating into the nonlinear saturation region.

Let us denote the vector of the CIR coefficients by

$$\boldsymbol{h} = [h_0\ h_1\ \cdots\ h_{L_{\mathrm{cir}}}]^{\mathrm{T}}, \tag{15}$$

whose length satisfies $L_{\rm cir} \leq N_{\rm cp}$. It is assumed that $h_0 = 1$ because if this is not the case, h_0 can always be absorbed into the CV static nonlinearity $\Psi()$, and the CIR coefficients are re-scaled as h_i/h_0 for $0 \leq i \leq L_{\rm cir}$. At the receiver, after the CP removal, the channel-impaired received signals y_k are given by

$$y_k = \sum_{i=0}^{L_{\rm cir}} h_i w_{k-i} + e_k, \quad 0 \leq k \leq N-1, \tag{16}$$

in which $w_{k-i} = w_{N+k-i}$ for $k < i$, where $e_k = e_{k_R} + \mathrm{j} \cdot e_{k_I}$ is the TD AWGN with $E\{e_{k_R}^2\} = E\{e_{k_I}^2\} = \sigma_e^2$. Because $N_{\rm cp} \geq L_{\rm cir}$, the CP removal at the receiver automatically cancels the inter block interference and transfers the linear convolution channel into the circular one. Passing $\boldsymbol{y} = [y_0\ y_1\ \cdots\ y_{N-1}]^{\rm T}$ through the N-point FFT processor yields the FD received OFDM vector $\boldsymbol{Y} = [Y_0\ Y_1\ \cdots\ Y_{N-1}]^{\rm T} = \boldsymbol{F}\boldsymbol{y}$ whose elements are expressed by

$$Y_n = H_n W_n + \Xi_n, \quad 0 \leq n \leq N-1, \tag{17}$$

where $\Xi_n = \Xi_{n_R} + \mathrm{j} \cdot \Xi_{n_I}$ is the FD nth subcarrier AWGN with $E\{\Xi_{n_R}^2\} = E\{\Xi_{n_I}^2\} = \sigma_e^2$, and the frequency domain channel transfer function coefficients (FDCTFCs) H_n for $0 \leq n \leq N-1$ are given by the N-point FFT of $\boldsymbol{h}$

$$[H_0\ H_1\ \cdots\ H_{N-1}]^{\rm T} = \boldsymbol{F}\boldsymbol{h}, \tag{18}$$

while

$$\boldsymbol{W} = [W_0\ W_1\ \cdots\ W_{N-1}]^{\rm T} = \boldsymbol{F}\boldsymbol{w} \tag{19}$$

is the N-point FFT of $\boldsymbol{w}$. Note that $\boldsymbol{w}$ is unobservable and, therefore, neither $\boldsymbol{w}$ nor $\boldsymbol{W}$ is available at the receiver.

If we denote $\boldsymbol{\Xi} = [\Xi_0\ \Xi_1\ \cdots\ \Xi_{N-1}]^{\rm T}$, the received FD OFDM signal (17) can be expressed concisely as

$$\begin{aligned} \boldsymbol{Y} &= \mathrm{diag}\{H_0, H_1, \cdots, H_{N-1}\}\boldsymbol{W} + \boldsymbol{\Xi} \\ &= \mathrm{diag}\{H_0, H_1, \cdots, H_{N-1}\}\boldsymbol{F}\boldsymbol{w} + \boldsymbol{\Xi}. \end{aligned} \tag{20}$$

Given the FDCTFCs H_n for $0 \leq n \leq N-1$, the FD one-tap equalization can be carried out. The zero-forcing equalization, for example, is given by

$$\widetilde{W}_n = \frac{Y_n}{H_n}, \quad 0 \leq n \leq N-1. \tag{21}$$

If the HPA $\Psi()$ at the transmitter were linear, $\widetilde{W}_n$ would be an estimate of the transmitted data symbol X_n. But $\Psi()$ is nonlinear, and the linear equalization (21) alone is no longer sufficient for estimating $\boldsymbol{X}$. If the nonlinearity $\Psi()$ is known and it is invertible, then the effects of $\Psi()$ can be compensated by inverting it. Specifically, let us define $\widetilde{\boldsymbol{W}} = [\widetilde{W}_0\ \widetilde{W}_1\ \cdots\ \widetilde{W}_{N-1}]^{\rm T}$ and

$$\boldsymbol{\Psi}(\boldsymbol{x}) = [\Psi(x_0)\ \Psi(x_1)\ \cdots\ \Psi(x_{N-1})]^{\rm T}. \tag{22}$$

Performing the IFFT on $\widetilde{\boldsymbol{W}}$ yields

$$\widetilde{\boldsymbol{w}} = [\widetilde{w}_0\ \widetilde{w}_1\ \cdots\ \widetilde{w}_{N-1}]^{\rm T} = \boldsymbol{F}^{\rm H}\widetilde{\boldsymbol{W}} = \boldsymbol{\Psi}(\boldsymbol{x}) + \boldsymbol{F}^{\rm H}\widetilde{\boldsymbol{\Xi}}, \tag{23}$$

where $\widetilde{\boldsymbol{\Xi}} = \mathrm{diag}\{H_0^{-1}, H_1^{-1}, \cdots, H_{N-1}^{-1}\}\boldsymbol{\Xi}$. Thus, an estimate of the TD OFDM signal $\boldsymbol{x}$ is given by

$$\widehat{\boldsymbol{x}} = \boldsymbol{\Psi}^{-1}(\widetilde{\boldsymbol{w}}) = \left[\Psi^{-1}(\widetilde{w}_0)\ \Psi^{-1}(\widetilde{w}_1)\ \cdots\ \Psi^{-1}(\widetilde{w}_{N-1})\right]^{\rm T}, \tag{24}$$

which further yields the estimate of the FD OFDM symbol vector $\boldsymbol{X}$ by

$$\widehat{\boldsymbol{X}} = \boldsymbol{F}\widehat{\boldsymbol{x}}. \tag{25}$$

III. Nonlinear Equalization of OFDM Hammerstein System

The reliable detection of the transmitted OFDM data symbols depends on the ability of estimating the FDCTFCs H_n or the CIR coefficients h_i and the CV static nonlinearity $\Psi()$ of the transmitter HPA as well as the ability of inverting $\Psi()$. Note that the CV HPA's nonlinearity, (10) and (11), is completely unknown to the receiver and $\boldsymbol{w}$ is unobserved. We adopt the CV B-spline neural network [15], [16] to represent the mapping $\widehat{w} = \widehat{\Psi}(x_R + \mathrm{j} \cdot x_I) : \mathbb{C} \to \mathbb{C}$ that is the estimate of the underlying CV nonlinear function $\Psi()$. We then propose an efficient algorithm for jointly estimating H_n and $\Psi()$ based on the CV B-spline modeling of $\Psi()$. As a byproduct of this Hammerstein channel identification, we construct the pseudo training data $\widehat{\boldsymbol{w}}$, and this allows us to estimate another B-spline neural network that models $\Psi^{-1}()$. Before introducing the B-spline modelling of $\Psi()$, we point out that the HPA $\Psi()$ of (10) and (11) satisfies the following conditions.

1) $\Psi()$ is a one to one mapping, i.e. it is an invertible and continuous function.
2) x_R and x_I are upper and lower bounded by some finite and known real values, where $x = x_R + \mathrm{j} \cdot x_I$ denotes the input to the HPA $\Psi()$. Furthermore, the distributions of x_R and x_I are identical.

According to the property 2), we assume that $U_{\min} < x_{\rm od} < U_{\max}$, where $U_{\min}$ and $U_{\max}$ are known finite real values, while $x_{\rm od}$ symbolically represents either x_R or x_I, namely, the subscript $_{\rm od}$ is either $_R$ or $_I$.

A. Complex-Valued B-Spline Neural Network

A set of univariate B-spline basis functions based on $x_{\rm od}$ is parametrised by the order $(P_o - 1)$ of a piecewise polynomial and a knot sequence which is a set of values defined on the one-dimensional real line that break it up into a number of intervals. To have $N_{\rm od}$ basis functions, the knot sequence is specified by $(N_{\rm od} + P_o + 1)$ knot values, $\{U_0, U_1, \cdots, U_{N_{\rm od}+P_o}\}$, with

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

At each end, there are $P_o - 1$ "external" knots that are outside the input region and one boundary knot. As a result, the number of "internal" knots is $N_{\rm od} + 1 - P_o$. Given the set of predetermined knots (26), the set of $N_{\rm od}$ B-spline basis functions can be formed by using the De Boor recursion [17], yielding for $1 \leq l \leq N_{\rm od} + P_o$,

$$B_l^{({\rm od},0)}(x_{\rm od}) = \begin{cases} 1, & \text{if } U_{l-1} \leq x_{\rm od} < U_l, \\ 0, & \text{otherwise,} \end{cases} \tag{27}$$

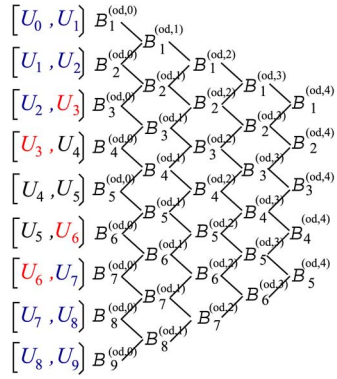


Fig. 1. Visualisation of the De Boor recursion for $P_o = 4$ and $N_{\rm od} = 5$, where $U_{\rm min} = U_3$ and $U_{\rm max} = U_6$.

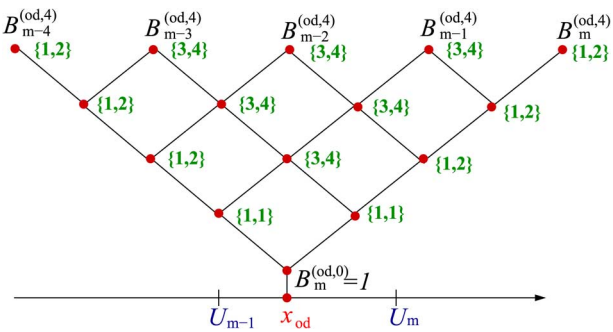


Fig. 2. Complexity of the B-spline model with $P_o = 4$ using the De Boor recursion, where $\{a, b\}$ beside a node indicates that it requires a additions and b multiplications to compute the basis function value at this node.

as well as for $l = 1, \cdots, N_{\rm od} + P_o - p$ and $p = 1, \cdots, P_o$,

$$B_l^{({\rm od},p)}(x_{\rm od}) = \frac{x_{\rm od} - U_{l-1}}{U_{p+l-1} - U_{l-1}} B_l^{({\rm od},p-1)}(x_{\rm od}) + \frac{U_{p+l} - x_{\rm od}}{U_{p+l} - U_l} B_{l+1}^{({\rm od},p-1)}(x_{\rm od}). \quad (28)$$

Here we have the superscript/subscript ${\rm od} = R$ or I.

The De Boor recursion is illustrated in Fig. 1. $P_o = 3$ to 4 is sufficient for most practical applications. The number of B-spline basis functions should be chosen to be sufficiently large to provide accurate approximation capability but not too large as to cause overfitting and to impose unnecessary computational complexity. The internal knots may be uniformly spaced in the interval $[U_{\rm min}, U_{\rm max}]$. The extrapolation capability of the B-spline model is influenced by the choice of the external knots. Note that there exist no data for $x_{\rm od} < U_{\rm min}$ and $x_{\rm od} > U_{\rm max}$ in identification but it is desired that the B-spline model has certain extrapolating capability outside the interval $[U_{\rm min}, U_{\rm max}]$. The external knots can be set empirically to meet the required extrapolation capability.

Because of the piecewise nature of B-spline functions, given a value $x_{\rm od} \in \mathbb{R}$, there are only $P_o + 1$ basis functions with nonzero values at most. This is advantageous as P_o can be set to a quite low value, e.g. $P_o = 4$ is often sufficient. The complexity of the De Boor recursion is, therefore, on the order of P_o^2. Fig. 2 shows the complexity of generating the B-spline basis function set for $P_o = 4$ using the De Boor recursion. Note that the complexity does not depend on the number of basis functions $N_{\rm od}$ employed. For the B-spline model with the polynomial degree $P_o = 4$, the total computational requirements are 26 additions and 38 multiplications at most.

Using the tensor product between the two sets of univariate B-spline basis functions [18], $B_l^{(R,P_o)}(x_R)$ for $1 \le l \le N_R$ and $B_m^{(I,P_o)}(x_I)$ for $1 \le m \le N_I$, a set of new B-spline basis functions $B_{l,m}^{(P_o)}(x)$ can be formed and used in the CV B-spline neural network, giving rise to

$$\widehat{w} = \widehat{\Psi}(x) = \sum_{l=1}^{N_R} \sum_{m=1}^{N_I} B_{l,m}^{(P_o)}(x)\theta_{l,m} = \sum_{l=1}^{N_R} \sum_{m=1}^{N_I} B_l^{(R,P_o)}(x_R) B_m^{(I,P_o)}(x_I)\theta_{l,m}, \quad (29)$$

where $\theta_{l,m} = \theta_{l,m_R} + {\rm j} \cdot \theta_{l,m_I} \in \mathbb{C}$, $1 \le l \le N_R$ and $1 \le m \le N_I$, are the CV weights.

Consider now using the CV B-spline neural network (29) to approximate the HPA nonlinearity $\Psi()$ over one OFDM symbol $\boldsymbol{x}$. Firstly, define the overall parameter vector $\boldsymbol{\theta} \in \mathbb{C}^{N_B}$, where $N_B = N_R \cdot N_I$, of the B-spline model (29) as

$$\boldsymbol{\theta} = [\theta_{1,1}\ \theta_{1,2}\ \cdots\ \theta_{l,m}\ \cdots\ \theta_{N_R,N_I}]^{\rm T}, \quad (30)$$

and the B-spline basis function matrix $\mathbf{B} \in \mathbb{R}^{N \times N_B}$ as

$$\mathbf{B} = \begin{bmatrix} B_{1,1}^{(P_o)}(x_0) & B_{1,2}^{(P_o)}(x_0) & \cdots & B_{N_R,N_I}^{(P_o)}(x_0) \\ B_{1,1}^{(P_o)}(x_1) & B_{1,2}^{(P_o)}(x_1) & \cdots & B_{N_R,N_I}^{(P_o)}(x_1) \\ \vdots & \vdots & \vdots & \vdots \\ B_{1,1}^{(P_o)}(x_{N-1}) & B_{1,2}^{(P_o)}(x_{N-1}) & \cdots & B_{N_R,N_I}^{(P_o)}(x_{N-1}) \end{bmatrix}. \quad (31)$$

Then the B-spline model (29) over $\boldsymbol{x}$ can be represented concisely by

$$\widehat{\boldsymbol{w}} = \mathbf{B}\boldsymbol{\theta} \quad (32)$$

where $\widehat{\boldsymbol{w}} = [\widehat{w}_0\ \widehat{w}_1\ \cdots\ \widehat{w}_{N-1}]^{\rm T}$ with $\widehat{w}_k = \widehat{\Psi}(x_k)$.

Remarks: B-splines have been widely studied in the subjects of approximation theory and numerical analysis, owing to their many good properties, including numerical stability. Specifically, the B-spline basis functions as model basis have the best approximation capability according to the Stone Weierstrass approximation theorem, i.e. the basis function is complete. Although any polynomial function can also be used to approximate a continuous function, the B-spline functions are proven to be optimally stable bases [19]–[21]. One critical and practical aspect to consider in the evaluation of a model representation is the stability with respect to perturbation of the model parameters, because in any identification, the data are inevitably noisy, which will perturb the model parameters away from their true values. A significant advantage of using the B-spline model with De Boor algorithm for functional approximation over many other polynomial forms is its superior numerical stability [19]–[21]. We now elaborate this aspect. Assume that the true system can be represented by the polynomial model of degree P_o as

$$y_{\rm od} = \sum_{i=0}^{P_o} a_i \cdot x_{\rm od}^i,$$

as well as by the following B-spline model exactly

$$y_{\rm od} = \sum_{i=1}^{N_{\rm od}} b_i \cdot B_i^{({\rm od},P_o)}(x_{\rm od}).$$

Because the identification data are noisy, the estimated model coefficients are perturbed from their true values to $\widehat{a}_i = a_i + \varepsilon_i$ for the polynomial model, and to $\widehat{b}_i = b_i + \varepsilon_i$ for the B-spline model. Assume that all the estimation noises ε_i are bounded, namely, $|\varepsilon_i| < \varepsilon_{\max}$. The upper bound of $|y_{\rm od} - \widehat{y}_{\rm od}|$ for the B-spline model can be worked out as follows

$$\begin{aligned}|y_{\rm od} - \widehat{y}_{\rm od}| &= \left| \sum_{i=1}^{N_{\rm od}} b_i \cdot B_i^{({\rm od},P_o)}(x_{\rm od}) - \sum_{i=1}^{N_{\rm od}} \widehat{b}_i \cdot B_i^{({\rm od},P_o)}(x_{\rm od}) \right| \\ &< \varepsilon_{\max} \cdot \left| \sum_{i=1}^{N_{\rm od}} B_i^{({\rm od},P_o)}(x_{\rm od}) \right| = \varepsilon_{\max}.\end{aligned}$$

Observe that the upper bound of the B-spline model output perturbation only depends on the upper bound of the perturbation noise, and it does not depend on the input value $x_{\rm od}$, the number of basis functions $N_{\rm od}$ or the polynomial degree P_o. This confirms that the B-spline model has the maximum numerical robustness, which is well known. Optimality of the B-spline model in terms of numerical stability is due to the convexity of its model bases, i.e. they are all positive and sum to one. By contrast, the upper bound of $|y_{\rm od} - \widehat{y}_{\rm od}|$ for the polynomial model can be worked out as follows

$$|y_{\rm od} - \widehat{y}_{\rm od}| = \left| \sum_{i=0}^{P_o} a_i \cdot x_{\rm od}^i - \sum_{i=0}^{P_o} \widehat{a}_i \cdot x_{\rm od}^i \right| < \varepsilon_{\max} \cdot \left| \sum_{i=0}^{P_o} x_{\rm od}^i \right|$$

Observe that the upper bound of the polynomial model output perturbation depends not only on the upper bound of the perturbation noise but also on the input value $x_{\rm od}$ and the polynomial degree P_o. The higher the polynomial degree P_o, the more serious the polynomial model may be perturbed, a well-known drawback of using polynomial modeling.

Fig. 3(a) plots a quadratic polynomial function $y_R = 0.001x_R^2 - 0.02x_R + 0.1$ defined over $x_R \in [0, 20]$ in solid line. Based on the knot sequence of $\{-5, -4, 0, 20, 24, 25\}$, this function is modeled as a quadratic B-spline model of $y_R = 0.14B_1^{(R,2)}(x_R) - 0.10B_2^{(R,2)}(x_R) + 0.14B_3^{(R,2)}(x_R)$, which is depicted in Fig. 3(b) in solid line. We now draw three noises ε_i, $1 \le i \le 3$, from a uniformly distributed random number in $[-0.0001, 0.0001]$, and add them to the three parameters in the two models, respectively, to simulate the effects of the noise in identification. Fig. 3(a) and Fig. 3(b) depict the ten sets of the perturbed functions in dashed line generated by perturbing the two models, respectively. It can be clearly seen from Fig. 3(a) that the polynomial model is seriously perturbed, but there is no noticeable change in Fig. 3(b) for the quadratic B-spline model. We next draw three perturbation noises from a uniformly distributed random number in $[-0.001, 0.001]$, and add them to the three parameters of the B-spline model. Again, the B-spline model is hardly affected, as can be seen from Fig. 3(c). We then draw three perturbation noises from a uniformly distributed random number in $[-0.01, 0.01]$ to add to the three B-spline parameters, and the results obtained are shown in Fig. 3(d). Observe from Fig. 3(a) and Fig. 3(d) that, despite of the fact that the strength of the perturbation noise added to the B-spline model coefficients is 100 times larger than that added to the polynomial model coefficients, the B-spline model is much less seriously perturbed than the polynomial model.

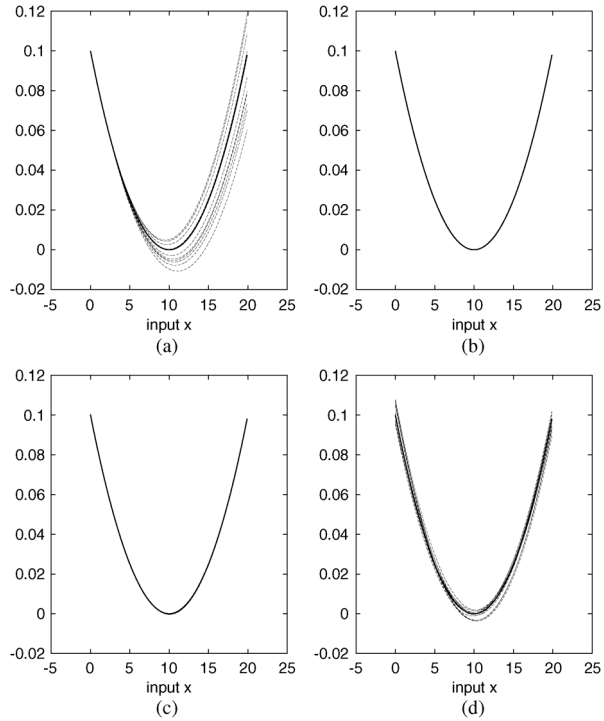


Fig. 3. (a) The polynomial model with three perturbation noises drawn from a uniformly distributed random number in $[-0.0001, 0.0001]$, (b) the B-spline model with three perturbation noises drawn from a uniformly distributed random number in $[-0.0001, 0.0001]$, (c) the B-spline model with three perturbation noises drawn from a uniformly distributed random number in $[-0.001, 0.001]$, and (d) the B-spline model with three perturbation noises drawn from a uniformly distributed random number in $[-0.01, 0.01]$.

B. Identification of the OFDM Hammerstein Channel

The identification of the OFDM Hammerstein channel involves estimating the parameter vector $\boldsymbol{\theta}$ of the B-spline neural network (29) that represents the nonlinearity $\Psi()$ as well as the CIR coefficient vector $\boldsymbol{h}$. Given a block of K training samples $\{x_k, y_k\}_{k=0}^{K-1}$, where $K \le N$, the task can be formulated as the one that minimises the cost function

$$J(\boldsymbol{h}, \boldsymbol{\theta}) = \frac{1}{K} \sum_{k=0}^{K-1} |\widehat{e}_k|^2 = \frac{1}{K} \sum_{k=0}^{K-1} |y_k - \widehat{y}_k|^2, \qquad (33)$$

subject to the constraint of $h_0 = 1$, in which the model prediction $\widehat{y}_k$ is given by

$$\widehat{y}_k = \sum_{i=0}^{L_{\rm cir}} h_i \widehat{w}_{k-i} = \sum_{i=0}^{L_{\rm cir}} h_i \sum_{l=1}^{N_R} \sum_{m=1}^{N_I} B_{l,m}^{(P_o)}(x_{k-i}) \theta_{l,m}, \qquad (34)$$

where $x_{k-i} = x_{N+k-i}$ if $k < i$. Note that (34) can be viewed as two different linear regression models, namely, one is with respect to $\boldsymbol{h}$ when fixing $\boldsymbol{\theta}$ and the other is with respect to $\boldsymbol{\theta}$

given a fixed $\boldsymbol{h}$, each problem having a closed-form solution. According to [22], [23], the estimates of $\boldsymbol{\theta}$ and $\boldsymbol{h}$ are unbiased, irrespective the optimization algorithm used.

Specifically, denote $\boldsymbol{y} = [y_0 \ y_1 \ \cdots \ y_{K-1}]^{\mathrm{T}}$ and $\widehat{\boldsymbol{e}} = [\widehat{e}_0 \ \widehat{e}_1 \ \cdots \ \widehat{e}_{K-1}]^{\mathrm{T}}$. Then over the training data set, the system can be represented as

$$\boldsymbol{y} = \boldsymbol{P}\boldsymbol{h} + \widehat{\boldsymbol{e}} = \boldsymbol{Q}\boldsymbol{\theta} + \widehat{\boldsymbol{e}}, \tag{35}$$

where the regression matrices $\boldsymbol{P} \in \mathbb{C}^{K\times(L_{\mathrm{cir}}+1)}$ and $\boldsymbol{Q} \in \mathbb{C}^{K\times N_B}$ are given respectively

$$\boldsymbol{P} = \begin{bmatrix} \widehat{w}_0 & \widehat{w}_{-1} & \cdots & \widehat{w}_{-L_{\mathrm{cir}}} \\ \vdots & \vdots & \vdots & \vdots \\ \widehat{w}_k & \widehat{w}_{k-1} & \cdots & \widehat{w}_{k-L_{\mathrm{cir}}} \\ \vdots & \vdots & \vdots & \vdots \\ \widehat{w}_{K-1} & \widehat{w}_{K-2} & \cdots & \widehat{w}_{K-1-L_{\mathrm{cir}}} \end{bmatrix}, \tag{36}$$

$$\boldsymbol{Q} = \begin{bmatrix} \varphi_{1,1}(0) & \cdots & \varphi_{l,m}(0) & \cdots & \varphi_{N_R,N_I}(0) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \varphi_{1,1}(k) & \cdots & \varphi_{l,m}(k) & \cdots & \varphi_{N_R,N_I}(k) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \varphi_{1,1}(K-1) & \cdots & \varphi_{l,m}(K-1) & \cdots & \varphi_{N_R,N_I}(K-1) \end{bmatrix}, \tag{37}$$

in which

$$\widehat{w}_k = \widehat{\Psi}(x_k) = \sum_{l=1}^{N_R}\sum_{m=1}^{N_I} B_{l,m}^{(P_o)}(x_k)\theta_{l,m}, \tag{38}$$

$$\varphi_{l,m}(k) = 1 + \sum_{i=1}^{L_{\mathrm{cir}}} h_i B_{l,m}^{(P_o)}(x_{k-i}), \tag{39}$$

with $x_k = x_{N+k}$ if $k < 0$.

We adopt the following ALS procedure to estimate $\boldsymbol{h}$ and $\boldsymbol{\theta}$, which is a coordinate gradient descent algorithm [25], [24]. The global convergence of the generic coordinate gradient descent algorithm for quasi convex cost functions was given in [24]. The cost function (33) is quasi convex and moreover, unlike a generic coordinate gradient descent algorithm, in our case we have the closed-form solutions for both $\boldsymbol{h}$ and $\boldsymbol{\theta}$. Therefore, our ALS procedure guarantees to converge fast to an unbiased estimate of $\boldsymbol{h}$ and $\boldsymbol{\theta}$ jointly.

Initialisation: Initialise $\widehat{w}_k = x_k$ in $\boldsymbol{P}$ of (36). Calculate $\boldsymbol{h}$ as the LS estimate given by

$$\widehat{\boldsymbol{h}}^{(0)} = \left(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P}\right)^{-1}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{y}. \tag{40}$$

Then obtain $\widehat{\boldsymbol{h}}^{(0)}$ by normalising $h_i \leftarrow h_i/h_0$ for $0 \le i \le L_{\mathrm{cir}}$.

ALS Estimation: For $1 \le \tau \le \tau_{\max}$, where $\tau_{\max}$ is the maximum number of iterations, perform:

a) Fix $\boldsymbol{h}$ to $\widehat{\boldsymbol{h}}^{(\tau-1)}$ in $\boldsymbol{Q}$ of (37). The LS estimate of $\widehat{\boldsymbol{\theta}}^{(\tau)}$ is readily given by

$$\widehat{\boldsymbol{\theta}}^{(\tau)} = \left(\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{Q}\right)^{-1}\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{y}. \tag{41}$$

b) For $\boldsymbol{P}$ of (36), fix $\widehat{w}_i$ according to (38) based on $\widehat{\boldsymbol{\theta}}^{(\tau)}$. Calculate

$$\widehat{\boldsymbol{h}}^{(\tau)} = \left(\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P}\right)^{-1}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{y}. \tag{42}$$

Then obtain $\widehat{\boldsymbol{h}}^{(\tau)}$ by normalising $h_i \leftarrow h_i/h_0$ for $0 \le i \le L_{\mathrm{cir}}$.

A few iterations, i.e. a very small $\tau_{\max}$, are sufficient for the above ALS estimation procedure to converge to a joint unbiased estimate of $\boldsymbol{h}$ and $\boldsymbol{\theta}$ that is a minimum solution for minimising the cost function (33).

C. Inversion of the OFDM Hammerstein Channel's Static Nonlinear Function

Given the CV Hammerstein channel's static nonlinearity $\Psi()$, we wish to compute its inversion defined by $x_k = \Psi^{-1}(w_k)$. This task is identical to find the CV root of $w_k = \Psi(x_k)$, given w_k, which can be solved iteratively [14], [15]. Given the estimated nonlinearity $\widehat{\Psi}()$ and during the data detection phase, the strategy of [14], [15] requires to iteratively calculate the root of $\widetilde{w}_k = \widehat{\Psi}(\widehat{x}_k)$ for each linearly equalised TD received signal sample $\widetilde{w}_k$ in order to obtain the TD transmitted signal estimate $\widehat{x}_k$. In order to avoid the iterative root finding procedure for every sample $\widetilde{w}_k$, we adopt an alternative strategy by constructing a mapping $x_k = \Phi(w_k, \boldsymbol{\alpha}) = \Psi^{-1}(w_k)$ also based on the CV B-spline neural network of Section III-A, where $\boldsymbol{\alpha}$ denotes the associated parameter vector of this inverting B-spline model. In order to learn the mapping $x_k = \Phi(w_k, \boldsymbol{\alpha})$, however, a training data set $\{w_k, x_k\}$ would be needed but w_k is unobservable and, therefore, is not available. Fortunately, as a byproduct of the OFDM Hammerstein channel identification presented in Section III-B, we already obtain an estimate for $\boldsymbol{w}$ as $\widehat{\boldsymbol{w}} = \boldsymbol{B}\boldsymbol{\theta}^{(\tau_{\max})}$. Therefore, we may construct the pseudo training data set $\{\widehat{w}_k, x_k\}_{k=0}^{N-1}$ to estimate $\boldsymbol{\alpha}$.

Specifically, define two knots sequences similar to (26) for w_R and w_I. Similar to (29), we have[1]

$$\begin{aligned}\widehat{x} = \widehat{\Phi}(w, \boldsymbol{\alpha}) &= \sum_{l=1}^{N_R}\sum_{m=1}^{N_I} B_{l,m}^{(P_o)}(w)\alpha_{l,m} \\ &= \sum_{l=1}^{N_R}\sum_{m=1}^{N_I} B_{l}^{(R,P_o)}(w_R)B_{m}^{(I,P_o)}(w_I)\alpha_{l,m},\end{aligned} \tag{43}$$

where $B_l^{(R,P_o)}(w_R)$ and $B_m^{(I,P_o)}(w_I)$ are respectively calculated based on (27) and (28), while

$$\boldsymbol{\alpha} = [\alpha_{1,1} \ \alpha_{1,2} \ \cdots \ \alpha_{l,m} \ \cdots \ \alpha_{N_R,N_I}]^{\mathrm{T}}, \tag{44}$$

Here again for notational simplicity, we assume that the same number of basis functions and polynomial order are used for the two B-spline neural networks $\Psi(x_n)$ and $\Phi(w_n)$. Over the pseudo training data set $\{\widehat{w}_k, x_k\}_{k=0}^{N-1}$, the regression matrix $\widetilde{\mathbf{B}} \in \mathbb{R}^{N\times N_B}$ can be formed as

$$\widetilde{\mathbf{B}} = \begin{bmatrix} B_{1,1}^{(P_o)}(\widehat{w}_0) & B_{1,2}^{(P_o)}(\widehat{w}_0) & \cdots & B_{N_R,N_I}^{(P_o)}(\widehat{w}_0) \\ B_{1,1}^{(P_o)}(\widehat{w}_1) & B_{1,2}^{(P_o)}(\widehat{w}_1) & \cdots & B_{N_R,N_I}^{(P_o)}(\widehat{w}_1) \\ \vdots & \vdots & \vdots & \vdots \\ B_{1,1}^{(P_o)}(\widehat{w}_{N-1}) & B_{1,2}^{(P_o)}(\widehat{w}_{N-1}) & \cdots & B_{N_R,N_I}^{(P_o)}(\widehat{w}_{N-1}) \end{bmatrix}. \tag{45}$$

and the LS solution for $\boldsymbol{\alpha}$ is readily given by $\widetilde{\boldsymbol{\alpha}} = (\widetilde{\mathbf{B}}^{\mathrm{T}}\widetilde{\mathbf{B}})^{-1}\widetilde{\mathbf{B}}^{\mathrm{T}}\boldsymbol{x}$.

[1]In order to avoid repetitions, we keep the same B-spline notations of Section III-A

TABLE I
EMPIRICALLY DETERMINED KNOT SEQUENCES

Knot sequence for x_R and x_I	-10,	-9,	-0.3,	-0.1,	-0.05,	-0.02,	0,	0.02,	0.05,	0.1,	0.3,	9,	10
Knot sequence for w_R and w_I	-20,	-10,	-3.5,	-2	-0.5,	-0.2,	0,	0.2,	0.5,	2,	3.5,	10,	20

TABLE II
IDENTIFICATION RESULTS FOR THE CIR COEFFICIENT VECTOR $\boldsymbol{h}$ OF THE HAMMERSTEIN CHANNEL

	True Parameters	Parameter estimate under			
		$E_{\rm b}/N_{\rm o} = 0$ dB OBO $= 5$ dB	$E_{\rm b}/N_{\rm o} = 10$ dB OBO $= 5$ dB	$E_{\rm b}/N_{\rm o} = 0$ dB OBO $= 3$ dB	$E_{\rm b}/N_{\rm o} = 10$ dB OBO $= 3$ dB
h_0	1	1	1	1	1
h_1	$-0.2145 - {\rm j}0.1867$	$-0.2143 - {\rm j}0.1872$	$-0.2144 - {\rm j}0.1869$	$-0.2141 - {\rm j}0.1874$	$-0.2143 - {\rm j}0.1869$
h_2	$0.0399 + {\rm j}0.3675$	$0.0396 + {\rm j}0.3684$	$0.0399 + {\rm j}0.3678$	$0.0396 + {\rm j}0.3686$	$0.0399 + {\rm j}0.3678$
h_3	$-0.0900 + {\rm j}0.4053$	$-0.0903 + {\rm j}0.4051$	$-0.0901 + {\rm j}0.4052$	$-0.0902 + {\rm j}0.4051$	$-0.0895 + {\rm j}0.4052$
h_4	$-0.0893 + {\rm j}0.1287$	$-0.0895 + {\rm j}0.1265$	$-0.0894 + {\rm j}0.1280$	$-0.0896 + {\rm j}0.1261$	$-0.0895 + {\rm j}0.1279$
h_5	$-0.1117 + {\rm j}0.3035$	$-0.1131 + {\rm j}0.3035$	$-0.1122 + {\rm j}0.3035$	$-0.1134 + {\rm j}0.3035$	$-0.1122 + {\rm j}0.3035$
h_6	$-0.0766 - {\rm j}0.0264$	$-0.0765 - {\rm j}0.0269$	$-0.0766 - {\rm j}0.0265$	$-0.0765 - {\rm j}0.0269$	$-0.0766 - {\rm j}0.0265$
h_7	$0.0623 - {\rm j}0.0668$	$0.0621 - {\rm j}0.0668$	$0.0623 - {\rm j}0.0668$	$0.0621 - {\rm j}0.0670$	$0.0623 - {\rm j}0.0669$
h_8	$0.0282 + {\rm j}0.0324$	$0.0283 + {\rm j}0.0325$	$0.0282 + {\rm j}0.0325$	$0.0284 + {\rm j}0.0326$	$0.0282 + {\rm j}0.0325$
h_9	$-0.0395 - {\rm j}0.0291$	$-0.0393 - {\rm j}0.0304$	$-0.0394 - {\rm j}0.0295$	$-0.0391 - {\rm j}0.0308$	$-0.0393 - {\rm j}0.0297$

IV. SIMULATION STUDY

We considered a Hammerstein OFDM system in which the HPA employed was described by (10) and (11) with the parameter set given in (12). The number of subcarriers was $N = 2048$ and 64-QAM was used. We assumed a quasi-static Rayleigh multipath channel with an exponentially decreasing power delay profile, where the average power for the lth path was given by

$$E\left\{|h_l|^2\right\} = e^{-\frac{l^2}{\gamma^2}}, \quad 0 \le l \le L_{\rm cir}, \tag{46}$$

with γ being the channel degradation factor. In the simulation study, we set $\gamma = 3$ and $L_{\rm cir} = 9$. The CIR coefficients h_l for $0 \le l \le L_{\rm cir}$ remained constant during the communication session. We used a full OFDM pilot symbol with $K = N = 2048$ training samples in the joint estimation of the CIR coefficient vector $\boldsymbol{h}$ and the parameter vector $\boldsymbol{\theta}$ of the B-spline model for $\Psi()$ as well as the estimation of the parameter vector $\boldsymbol{\alpha}$ of the B-spline model for $\Psi^{-1}()$. The piecewise quartic polynomial of $P_o = 4$ was chosen as the B-spline basis function, and the number of B-spline basis functions was set to $N_R = N_I = 8$. Owing to the symmetric distribution of x_R and x_I, the knot sequence for x_R was set to be identical to that for x_I. Similarly, the knot sequences for w_R and w_I were chosen to be identical. The empirically determined knot sequences covering different HPA operating conditions are listed in Table I. The system's signal-to-noise ratio (SNR) was defined as ${\rm SNR} = E_{\rm b}/N_{\rm o}$, where $E_{\rm b}$ was the average power of the input signal x_k to the HPA and $N_{\rm o} = 2\sigma_e^2$ was the channel AWGN's power.

The identification experiments were conducted under the HPA operation conditions of ${\rm OBO} = 5$ dB and ${\rm OBO} = 3$ dB, respectively, as well as given the two SNR conditions of ${\rm SNR} = 0$ dB and ${\rm SNR} = 10$ dB, respectively. The identification results of the linear subsystem in the Hammerstein channel under these four experimental conditions are summarised in Table II, while the modelling results of the HPA static nonlinearity $\Psi()$ by the B-spline neural network $\widehat{\Psi}()$ for all the four simulation conditions are illustrated in Fig. 4. It can be seen from Table II that the CIR estimates achieve high accuracy for all the four conditions. The results of Fig. 4 clearly demonstrate the capability of the proposed CV B-spline neural network to accurately model the HPA's static nonlinearity, where it can be observed that the maximum deviation of the estimated phase response from the HPA's true phase response is less than 0.05 even under the adverse condition of ${\rm OBO} = 3$ dB and ${\rm SNR} = 0$ dB.

The combined response of the HPA's true nonlinearity $\Psi()$ and its inverse estimate $\widehat{\Phi}()$ obtained under the condition of ${\rm OBO} = 5$ dB and ${\rm SNR} = 10$ dB is depicted in Fig. 5. The result of Fig. 5 demonstrates the capability of the proposed CV B-spline neural network to accurately model the inversion of the HPA's nonlinearity based only on the pseudo training data. The effectiveness of the proposed nonlinear equalization scheme is illustrated in Fig. 6, where the nonlinear equalizer is constructed based on the estimated CIR $\widehat{\boldsymbol{h}}$ and inverse mapping $\widehat{\Phi}()$ obtained under ${\rm OBO} = 5$ dB and $E_{\rm b}/N_{\rm o} = 10$ dB.

In order to demonstrated the effectiveness of our proposed B-spline based nonlinear equalizer, we also investigated the polynomial based nonlinear equalizer. For this polynomial based nonlinear equalizer, we employed the polynomial model with the polynomial basis set

$$\begin{aligned}\{1, x_R, x_I, x_R^2, x_R x_I, x_I^2, x_R^3, x_R^2 x_I, x_R x_I^2, x_I^3,\\ x_R^4, x_R^3 x_I, x_R^2 x_I^2, x_R x_I^3, x_I^4\}\end{aligned} \tag{47}$$

to model the HPA's nonlinearity $\Psi()$, and the same ALS algorithm was used to identify the parameters of this polynomial model as well as the CIR coefficient vector $\boldsymbol{h}$. Furthermore, another polynomial model with the same polynomial basis set (47) was adopted to model $\Psi^{-1}()$, and the parameters of this inverting polynomial model were estimated using the LS algorithm based on the same pseudo training data. Without any channel coding, the raw achievable BER performance of the B-spline based nonlinear equalizer are plotted in Fig. 7 under three different operating conditions of the HPA, in comparison

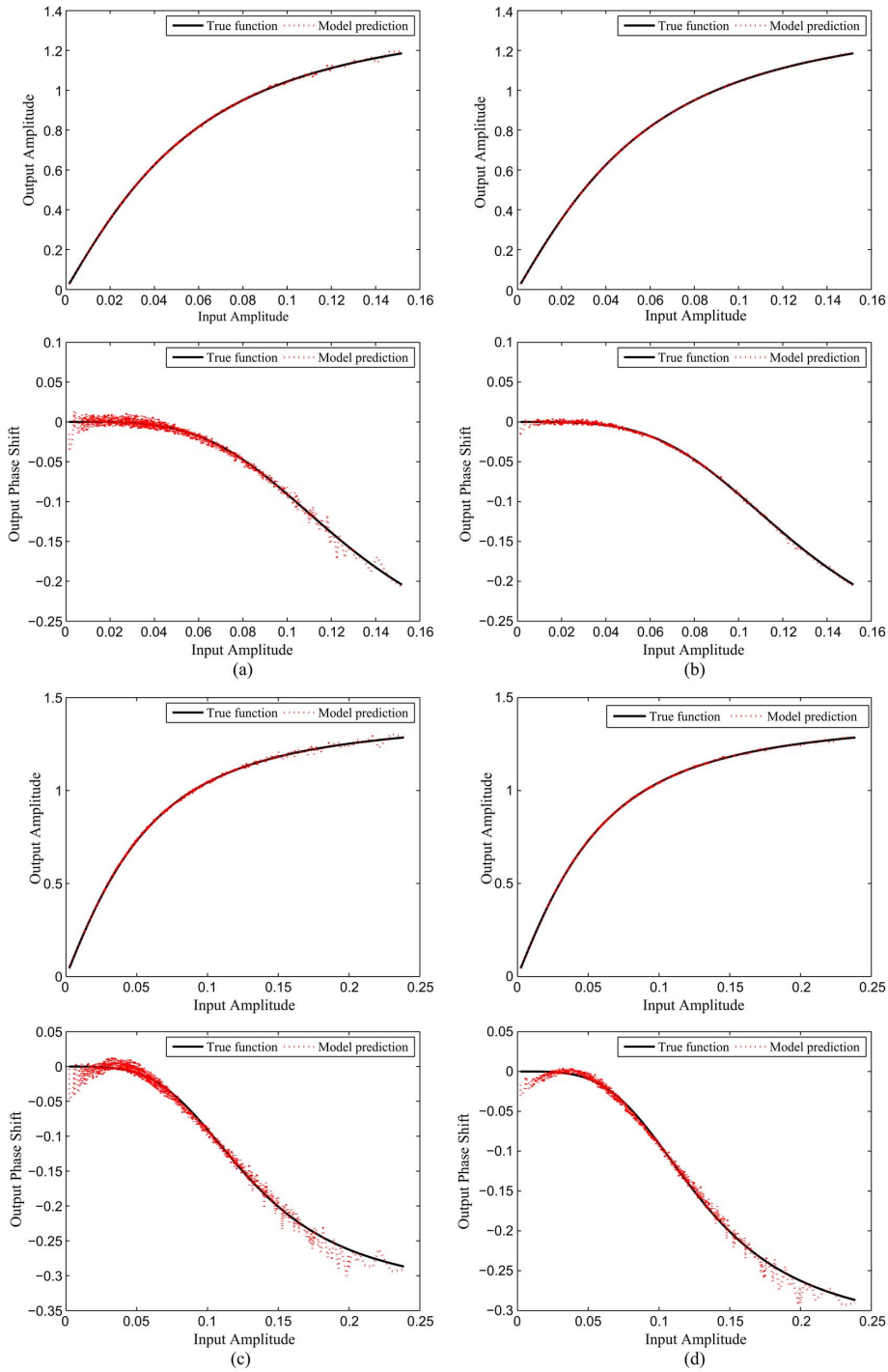


Fig. 4. Comparison of the HPA's static nonlinearity $\Psi()$ and the estimated static nonlinearity $\widehat{\Psi}()$ under: (a) OBO $= 5$ dB, $E_{\rm b}/N_{\rm o} = 0$ dB; (b) OBO $= 5$ dB, $E_{\rm b}/N_{\rm o} = 10$ dB; (c) OBO $= 3$ dB, $E_{\rm b}/N_{\rm o} = 0$ dB; and (d) OBO $= 3$ dB, $E_{\rm b}/N_{\rm o} = 10$ dB.

to the BER performance obtained by the standard linear equalizer and the polynomial based nonlinear equalizer. Clearly, the standard linear equalizer is incapable of compensating the nonlinear distortions of the Hammerstein channel and its attainable BER performance is poor even under the HPA operating condition of OBO $= 10$ dB, as can be seen from Fig. 7. By contrast, the proposed nonlinear equalizer constructed based on the estimated CIR and inverse mapping of the HPA is able to correct most of the nonlinear distortions and attains a much better BER performance. Also observe from Fig. 7 that the B-spline based nonlinear equalizer significantly outperforms the polynomial based nonlinear equalizer, particularly under the adverse operating conditions of OBO $= 3$ dB and 5 dB.

Next, we added a channel coder to the transmitter, and performed the channel decoding after equalization at the receiver. In this coded and 64-QAM modulated OFDM Hammerstein system, the source bits were encoded with the NASA standard 1/2 rate convolutional code [7, (133,171)], where the constraint length is 7, the generator polynomials are 133 and 171, while at the receiver, the demodulated bits were fed into the Viterbi

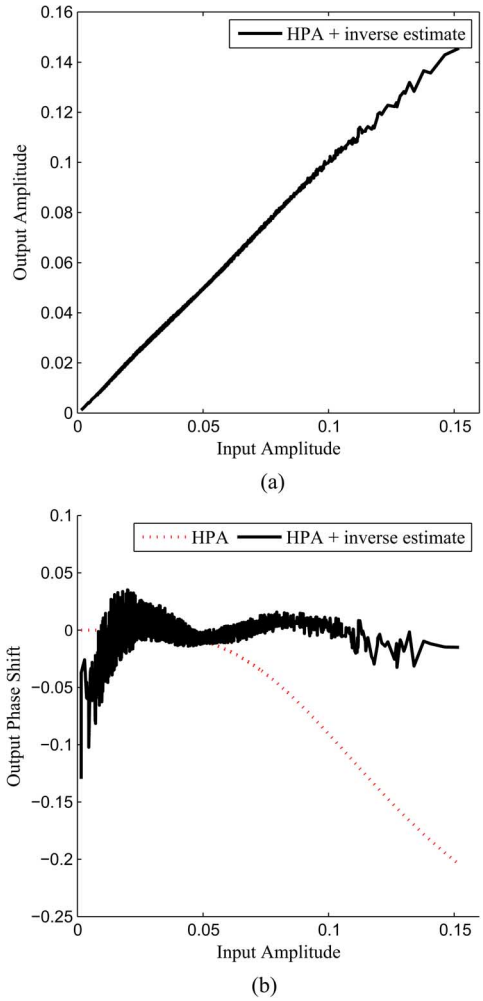


Fig. 5. Combined response of the HPA $\Psi()$ and its estimated inversion $\widehat{\Phi}()$ obtained under $\mathrm{OBO} = 5$ dB and $E_\mathrm{b}/N_\mathrm{o} = 10$ dB: (a) combined amplitude response, and (b) combined phase response.

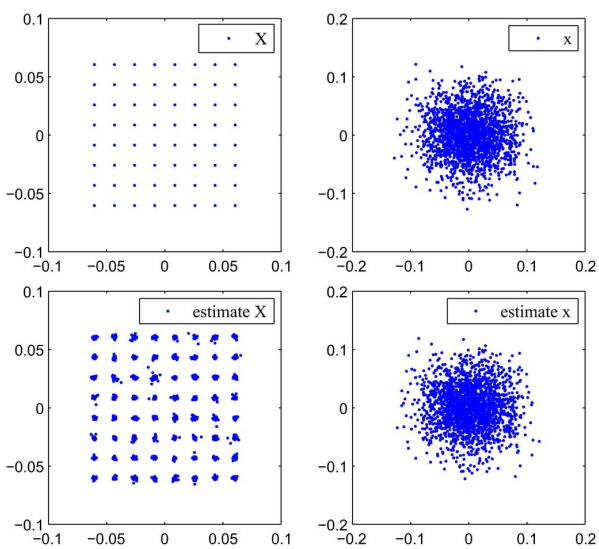


Fig. 6. Effectiveness of the proposed nonlinear equaliser based on the estimated CIR $\widehat{\boldsymbol{h}}$ and inverse mapping $\widehat{\Phi}()$ obtained under $\mathrm{OBO} = 5$ dB and $E_\mathrm{b}/N_\mathrm{o} = 10$ dB, where the top two plots depict one transmitted FD OFDM symbol $\boldsymbol{X}$ and its TD signal $\boldsymbol{x}$, while the bottom two plots show the corresponding estimated $\widehat{\boldsymbol{X}}$ and $\widehat{\boldsymbol{x}}$ obtained by the nonlinear equalizer.

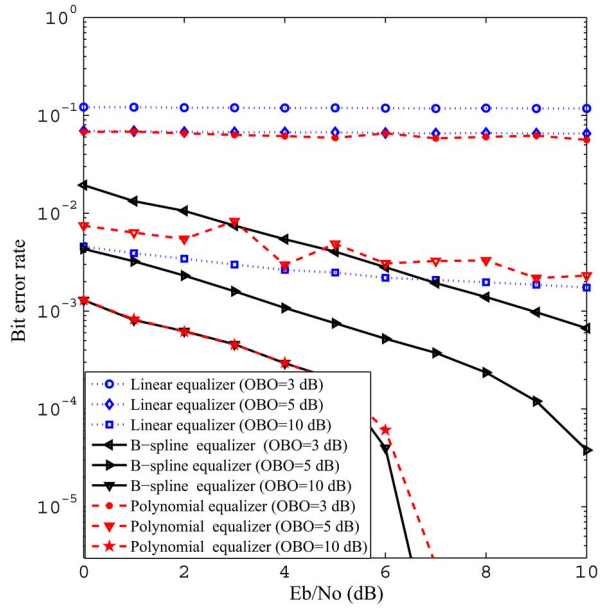


Fig. 7. The uncoded bit error rate performance comparison of the proposed B-spline based nonlinear equalizer with the standard linear equalizer and the polynomial based nonlinear equalizer.

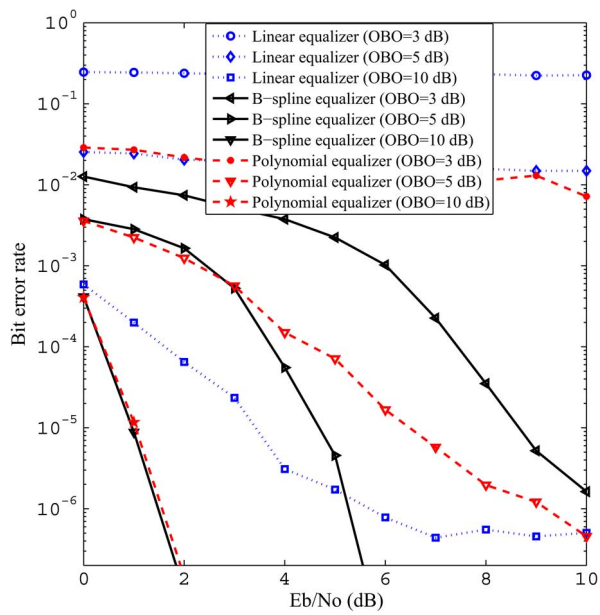


Fig. 8. The coded bit error rate performance comparison of the proposed B-spline based nonlinear equalizer with the standard linear equalizer and the polynomial based nonlinear equalizer.

decoder with trace back length of 20 [26]. Fig. 8 compares the coded BER performance of the B-spline based nonlinear equalizer with those of the linear equalizer as well as the polynomial based nonlinear equalizer under three HPA operating conditions. By comparing Fig. 8 with Fig. 7, we can see that the coded BER of the linear equalizer under the HPA operating condition of $\mathrm{OBO} = 3$ dB is actually worse than the uncoded case. This is because for channel coding to be effective, the uncoded BER must be below certain threshold; otherwise channel decoding actually degrades the achievable BER. Given $\mathrm{OBO} = 3$ dB, the uncoded BER obtained by the linear equalizer is higher than

this threshold, owing to the very serious nonlinear distortion of the channel. Under the operating condition of $\mathrm{OBO} = 5$ dB, the channel coding does improve the BER of the linear equalizer but only marginally, and the coded BER of the linear equalizer has a high BER floor, again owing to the serious nonlinear distortion of the channel. This is because channel coding is very effective in cleaning the errors caused by random noise but is unable to correct the errors caused by the nonlinear distortion at all. For $\mathrm{OBO} = 10$ dB, the channel is almost linear, and the channel coding significantly enhances the achievable BER of the linear equalizer, as can be observed by the coded BER of the linear equalizer for $\mathrm{OBO} = 10$ dB in Fig. 8. Clearly, in this case most of the errors are caused by the AWGN. Observe that this coded BER curve exhibits a BER floor around 10^{-6}. This is obviously caused by the very mild channel nonlinearity under $\mathrm{OBO} = 10$ dB, for which the channel coding is completely powerless to correct. By contrast, as most of the nonlinear distortions are corrected by the proposed nonlinear equalizer, it benefits more from the channel coding, as is self-evident by comparing Fig. 8 with Fig. 7. The results of Fig. 8 again confirm that the B-spline based nonlinear equalizer attains better BER performance than the polynomial based nonlinear equalizer, especially under the highly nonlinear operating conditions of $\mathrm{OBO} = 3$ dB and 5 dB.

V. Conclusions

A novel nonlinear equalization scheme has been developed for the Hammerstein OFDM system, where the nonlinear distortion is caused by the high power amplifier at transmitter. We have proposed to use a CV B-spline neural network for modelling the HPA's static nonlinearity as well as to use another B-spline neural network for modelling the inverse mapping of the HPA's nonlinearity. Our novel contribution includes deriving a highly efficient alternating least squares algorithm for estimating the CIR coefficients and the parameters of the B-spline neural network that models the static nonlinearity of the Hammerstein channel. Moreover, as a natural byproduct of this Hammerstein channel identification, the pseudo training data can be constructed to effectively estimate the inverse B-spline neural network that models the inverse mapping of the HPA nonlinearity. All the three estimates, the CIR coefficients, the parameters of the B-spline model and the parameters of the inverse B-spline model, have the closed-form LS solutions. Simulation results obtained have demonstrated that our proposed identification procedure is capable of accurately estimating the Hammerstein channel as well as the inverse mapping of the channel's static nonlinearity. The results obtained also confirm the effectiveness of the proposed nonlinear equalizer constructed based on the estimated CIR and inverse B-spline mapping. Our future work will investigate applying turbo detection-decoding technique to this challenging nonlinear equalization problem.

References

[1] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Mag.*, vol. 28, no. 5, pp. 5–14, May 1990.

[2] L. Hanzo, M. Münster, B. J. Choi, and T. Keller, *OFDM and MC-CDMA for Broadband Multi-User Communications, WLANs, and Broadcasting*. Chichester, U.K.: Wiley, 2003.

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

[4] M. Honkanen and S.-G. Häggman, "New aspects on nonlinear power amplifier modeling in radio communication system simulations," in *Proc. PIMRC'97*, Helsinki, Finland, Sep. 1–4, 1997, pp. 844–848.

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

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

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

[8] 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. Commun.*, vol. 52, no. 1, pp. 159–165, Jan. 2004.

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

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

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

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

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

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

[15] X. Hong, S. Chen, and C. J. Harris, "Complex-valued B-spline neural networks for modeling and inverse of Wiener systems," in *Complex-Valued Neural Networks: Advances and Applications*, A. Hirose, Ed. Hoboken, NJ, USA: Wiley, 2013, ch. 9, pp. 209–233.

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

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

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

[19] J. M. Pena, "B-spline and optimal stability," *Math. Comput.*, vol. 66, no. 220, pp. 1555–1560, Oct. 1997.

[20] T. Lyche and J. M. Pena, "Optimally stable multivariate bases," *Adv. Comput. Math.*, vol. 20, no. 1–3, pp. 149–159, Jan. 2004.

[21] E. Mainar and J. M. Pena, "Optimal stability of bivariate tensor product B-bases," *J. Numer. Anal., Ind. Appl.Math.*, vol. 6, no. 3–4, pp. 95–104, 2011.

[22] A. V. Ivanov, "An asymptotic expansion for the distribution of the least squares estimator of the non-linear regression parameter," *SIAM Theory Probabil. Its Appl.*, vol. 21, no. 3, pp. 557–570, 1977.

[23] C.-F. Wu, "Asymptotic theory of nonlinear least squares estimation," *Ann. Statist.*, vol. 9, no. 3, pp. 501–513, 1981.

[24] Z. Q. Luo and P. Tseng, "On the convergence of the coordinate descent method for convex differentiable minimization," *J. Optimiz. Theory Appl.*, vol. 72, no. 1, pp. 7–35, Jan. 1991.

[25] R. J. Hathaway and J. C. Bezdek, "Grouped coordinate minimization using Newton's method for inexact minimization in one vector coordinate," *J. Optimiz. Theory Appl.*, vol. 71, no. 3, pp. 503–516, Dec. 1991.

[26] C. R. Lahmeyer and K.-M. Cheung, "Long Decoding Runs for Galileo's Convolutional Codes," The Interplanetary Network Progress Report 42-95, Nov. 15, 1988 [Online]. Available: http://ipnpr.jpl.nasa.gov/progress_report/42-95/95M.PDF

Xia Hong (A'01–SM'02) received her university education at National University of Defense Technology, P. R. China (BSc, 1984, MSc, 1987), and University of Sheffield, UK (PhD, 1998), all in automatic control.

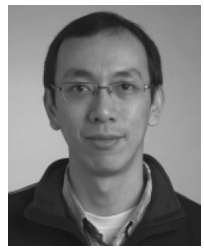
She worked as a research assistant in Beijing Institute of Systems Engineering, Beijing, China from 1987–1993. She worked as a research fellow in the Department of Electronics and Computer Science at University of Southampton from 1997–2001. She is currently a Professor at School of Systems Engineering, University of Reading. She is actively engaged in research into nonlinear systems identification, data modelling, estimation and intelligent control, neural networks, pattern recognition, learning theory and their applications. She has published over 140 research papers, and coauthored a research book.

Professor Hong was awarded a Donald Julius Groen Prize by IMechE in 1999.

Sheng Chen (M'90–SM'97–F'08) received his BEng degree from the East China Petroleum Institute, Dongying, China, in 1982, and his PhD degree from the City University, London, in 1986, both in control engineering. In 2005, he was awarded the higher doctoral degree, Doctor of Sciences (DSc), from the University of Southampton, Southampton, UK.

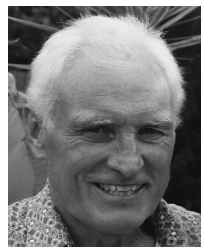
From 1986 to 1999, He held research and academic appointments at the Universities of Sheffield, Edinburgh and Portsmouth, all in UK. Since 1999, he has been with Electronics and Computer Science, the University of Southampton, UK, where he currently holds the post of Professor in Intelligent Systems and Signal Processing. Dr Chen's research interests include adaptive signal processing, wireless communications, modelling and identification of nonlinear systems, neural network and machine learning, intelligent control system design, evolutionary computation methods and optimisation. He has published over 500 research papers.

Dr. Chen is a Fellow of IET, a Distinguished Adjunct Professor at King Abdulaziz University, Jeddah, Saudi Arabia, and an ISI highly cited researcher in engineering (March 2004). In 2014, he was elected as a Fellow of the United Kingdom Royal Academy of Engineering.



Yu Gong (M'06) received his BEng and MEng degrees from the University of Electronics and Science Technology of China in 1992 and 1995, respectively, both in electronic engineering. He received his PhD degree in communications from the National University of Singapore in 2002.

After his PhD, Dr Gong took several research positions in Institute of Inforcomm Research in Singapore and Queen's University of Belfast in UK, respectively. From 2006 and 2012, Dr Gong was an academic staff in the School of Systems Engineering, the University of Reading, UK. Since July 2021, he has been with the School of Electronic, Electrical and Systems Engineering, Loughborough University, UK. Dr Gong's research interests are in the area of signal processing and communications including wireless communications, cooperative networks, non-linear and non-stationary system identification, and adaptive filters.



Chris J. Harris received his BSc and MA degrees from the University of Leicester and the University of Oxford in UK, respectively, and his PhD degree from the University of Southampton, UK, in 1972. He was awarded the higher doctoral degree, the Doctor of Sciences (DSc), by the University of Southampton in 2001. He is Emeritus Reseach Professor at the University of Southampton, having previously held senior academic appointments at Imperial College, Oxford and Manchester Universities, as well as Deputy Chief Scientist for the UK Government.

Professor Harris was awarded the IEE senior Achievement Medal for Data Fusion research and the IEE Faraday Medal for distinguished international research in Machine Learning. He was elected to the UK Royal Academy of Engineering in 1996. He is the co-author of over 450 scientific research papers during a 45 year research career.