Single-carrier frequency domain equalisation for hammerstein communication systems using complex-valued neural networks

Article

Accepted Version

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2333555, IEEE Transactions on Signal Processing

Single-Carrier Frequency Domain Equalisation for Hammerstein Communication Systems Using Complex-Valued Neural Networks

Xia Hong, *Senior Member, IEEE,* Sheng Chen, *Fellow, IEEE,* Chris J. Harris and Emad Khalaf

***Abstract*—Single-carrier (SC) block transmission with frequency-domain equalisation (FDE) offers a viable transmission technology for combating the adverse effects of long dispersive channels encountered in high-rate broadband wireless communication systems. However, for high bandwidth-efficiency and high power-efficiency systems, the channel can generally be modelled by the Hammerstein system that includes the nonlinear distortion effects of the high power amplifier (HPA) at transmitter. For such nonlinear Hammerstein channels, the standard SC-FDE scheme no longer works. This paper advocates a complex-valued (CV) B-spline neural network based nonlinear SC-FDE scheme for Hammerstein channels. Specifically, We model the nonlinear HPA, which represents the CV static nonlinearity of the Hammerstein channel, by a CV B-spline neural network, and we develop two efficient alternating least squares schemes for estimating the parameters of the Hammerstein channel, including both the channel impulse response coefficients and the parameters of the CV B-spline model. We also use another CV B-spline neural network to model the inversion of the nonlinear HPA, 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 natural byproduct of the Hammerstein channel identification. Equalisation of the SC Hammerstein channel can then be accomplished by the usual one-tap linear equalisation in frequency domain as well as the inverse B-spline neural network model obtained in time domain. Extensive simulation results are included to demonstrate the effectiveness of our nonlinear SC-FDE scheme for Hammerstein channels.**

***Index Terms*—Single-carrier frequency domain equalisation, high power amplifier, Hammerstein channel, complex-valued B-spline neural network**

I. Introduction

It is well-known that for high-speed broadband communication applications with data rates in tens of Mbps or higher over wireless channels of typical delay spread in microseconds, the intersymbol interference (ISI) of wireless channels will span over tens or even hundreds of symbols. This causes th nightmare senario for time-domain (TD) equalisation, as an impractically long equaliser is required which suffers from excessively slow convergence and has poor performance. Orthogonal frequency-division multiplexing (OFDM) [1], [2] provides a low-complexity high-performance solution for mitigating long ISI. Owing to its virtues of resilience to frequency selective fading channels, OFDM has found its way into many recent wireless network standards. However, OFDM signals are notoriously known to have high peak-to-average power ratio (PAPR), which requires the high power amplifier (HPA) at the transmitter to have an extremely long linear dynamic range. This requirement may not be met by practical HPAs which exhibits nonlinear saturation characteristics [3]–[7]. Single-carrier (SC) block transmission with frequency-domain equalisation (FDE) [8], [9] offers a viable alternative solution for long ISI mitigation. Although the total complexity of a SC-FDE based transceiver is the same as that of an OFDM based transceiver, the SC-FDE transmitter does not require the fast Fourier transform (FFT) operation, and therefore it is better suited for uplink implementation. Therefore, the long term evolution advanced (LTE-A) has specified the standard for the uplink of the fourth generation (4G) systems based on the SC-FDE solution [10].

SC based high-rate broadband systems typically employ high-order quadrature amplitude modulation (QAM) signalling [11] for the sake of further enhancing the achievable bandwidth efficiency. The higher the order of QAM signalling, the better the bandwidth efficiency but also the higher the PAPR of the resulting transmit signal. This may drive the HPA at the transmitter into the nonlinear saturation region, which will significantly degrade the system's achievable bit error rate (BER) performance. Moreover, green communication [12] by emphasizing energy-efficiency aspect of communication favours high power-efficiency nonlinear HPAs, which however could not accommodate high bandwidth-efficiency transmission technologies. Furthermore, recently, millimeter-wave (mmW) communications have been attracting extensive attentions, owing to the huge amount of unlicensed bandwidth offered by mmW systems [13]–[15]. SC transmission provides a viable technology for mmW based beyond 4G (B4G) systems [15]. However, for mmW communications, the design of HPA encounters severe nonlinearity [16], [17]. Therefore, it is important to be able to effectively compensate the nonlinear distortions of the HPA in the design of a SC-FDE based B4G wireless system in order to achieve both high bandwidth efficiency and high power efficiency.

An effective approach to compensate for the nonlinear

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

S. Chen and C.J. Harris are with Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK (E-mails: sqc@ecs.soton.ac.uk, cjh@ecs.soton.ac.uk). S. Chen is also with Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia.

E. Khalaf is with Electrical and Computer Engineering Department, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia (E-mail: ekhalaf@kau.edu.sa)

This project was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, under grant No. (5-135-35-HiCi). The authors, therefore, acknowledge technical and financial support of DSR.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2333555, IEEE Transactions on Signal Processing

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 [18]–[24]. 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 much more difficult, because it is extremely challenging for a pocket-size handset to absorb the additional hardware and computational complexity. Therefore, the predistorter option is not viable for the SC-FDE based uplink system. 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, nonlinear inversion or equalisation of the SC-FDE based CV Hammerstein channel is a challenging task.

In this contribution, we propose an efficient nonlinear SC-FDE scheme for Hammerstein channels based on the CV B-spline neural network. Motivated by our previous works [24]–[26], which demonstrate the effectiveness of the CV B-spline neural network approach for identification and inversion of CV Wiener systems, we adopt a CV B-spline neural network to model the CV static nonlinearity of the Hammerstein channel, and we develop two highly efficient alternating least squares (ALS) identification algorithms for estimating the channel impulse response (CIR) coefficients as well as the parameters of the CV B-spline neural network that models the HPA's CV static nonlinearity. As linear equalisation is naturally accomplished in SC-FDE based systems by a one-tap equalisation in frequency domain (FD), nonlinear SC-FDE of the Hammerstein channel additionally involves the inversion of the estimated CV B-spline neural network that models the HPA's nonlinearity in TD. The previous works [24], [26] considers the inversion of a B-spline model as the root finding problem, and develop an iterative root finding procedure based on the Gauss-Newton algorithm for inverting the estimated B-spline neural network mode. This approach requires to carry out the iterative root finding procedure for detecting every data symbol. We propose a much faster and more efficient alternative for inverting the HPA's nonlinearity. Specifically, we use another CV B-spline neural network to model the inversion of the HPA's CV nonlinearity. Although the HPA's output at the transmitter is unobservable at the receiver for identifying this CV 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. The effectiveness of our proposed CV B-spline neural network based SC-FDE scheme for Hammerstein channels is demonstrated in an extensive simulation study.

To the best of our knowledge, this is the first practical and effective scheme proposed for compensating the transmitter HPA's nonlinearity at the receiver for SC-FDE based systems. It should be emphasized that the scheme developed in a recent paper [27] is not applicable to the SC-FDE based system, and it can only be applied to a TD equalisation based system. More specifically, for pure TD transmission systems, the work [27] develops a highly-complicated, high-training-overhead and high-complexity nonlinear TD based equalisation scheme for the Hammerstein channel. Firstly, a specially designed unity-PAPR training sequence has to be adopted to identify the HPA biased CIR. Linear TD equalisation is carried out with the equaliser order set to the data frame length, based on the estimated biased CIR. Then a second training sequence with the same data modulation scheme and same data frame length has to be employed to identify the so-called the distortion constellation set (DCS). Finally, the estimated DCS is used for data detection based on the TD linearly equalised received signal sequence, which has a much higher complexity then a standard data detection. It is clearly that the scheme of [27] not only suffers from the drawbacks of high-training-overhead and high-complexity but also cannot be applied to SC-FDE based systems. By contrast, the B-spline based approach adopted in this manuscript for SC-FDE based systems can be applied to the same pure TD transmission system considered in [27]. In fact, we have applied our B-spline based nonlinear equalisation scheme to a similar TD transmission system in [28], which offers lower training overhead and lower complexity than the scheme developed in [27].

The rest of this paper is organized as follows. Section II presents the Hammerstein channel model and summarises the requirements of nonlinear SC-FDE given the Hammerstein channel. Section III details our proposed CV B-spline neural network based nonlinear SC-FDE scheme, while the simulation study is presented in Section IV to demonstrate the excellent performance of our proposed nonlinear SC-FDE scheme. Our conclusions are offered in Section V.

Throughout this contribution, a CV number $x \in \mathbb{C}$ is represented either by the rectangular form $x = x_R + \mathsf{j} \cdot x_I$, where $\mathsf{j} = \sqrt{-1}$, while $x_R = \Re[x]$ and $x_I = \Im[x]$ denote the real and imaginary parts of x, or alternatively by the polar form $x = |x| \cdot e^{\mathsf{j}\angle^x}$ 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. Hammerstein Channel Model for SC-FDE

In our SC block based transmission system, each transmit block consists of N data symbols expressed as

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

where $[s]$ denotes the block index. We assume that $x_k[s]$, $0 \le k \le N-1$, take the values from the M-QAM symbol set

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

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2333555, IEEE Transactions on Signal Processing

where $2d$ is the minimum distance between symbol points. For notational simplification, the block index $[s]$ is dropped in the sequel. Adding the cyclic prefix (CP) of length $N_{\rm cp}$ to $\boldsymbol{x}$ yields

$$\bar{\boldsymbol{x}} = \left[x_{-N_{\rm cp}}\ x_{-N_{\rm cp}+1}\cdots x_{-1} \mid \boldsymbol{x}^{\rm T}\right]^{\rm T}, \tag{3}$$

in which $x_{-k} = x_{N-k}$ for $1 \le k \le N_{\rm cp}$. The signal block $\bar{\boldsymbol{x}}$ is amplified by the HPA to yield the actually transmitted signal vector

$$\begin{aligned}\bar{\boldsymbol{w}} =& \left[w_{-N_{\rm cp}}\ w_{-N_{\rm cp}+1}\cdots w_{-1} \mid w_0\ w_1\cdots w_{N-1}\right]^{\rm T}\\ =& \left[w_{-N_{\rm cp}}\ w_{-N_{\rm cp}+1}\cdots w_{-1} \mid \boldsymbol{w}^{\rm T}\right]^{\rm T}\end{aligned} \tag{4}$$

where

$$w_k = \Psi\left(x_k\right),\ -N_{\rm cp} \le k \le N-1, \tag{5}$$

in which $\Psi(\ )$ represents the CV static nonlinearity of the transmitter HPA, and $w_{-k} = w_{N-k}$ for $1 \le k \le N_{\rm cp}$. The most widely used HPA is 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

$$A(r) = \frac{g_a r}{\left(1+\left(\frac{g_a r}{A_{\rm sat}}\right)^{2\beta_a}\right)^{\frac{1}{2\beta_a}}}, \tag{6}$$

$$\Upsilon(r) = \frac{\alpha_\phi r^{q_1}}{1+\left(\frac{r}{\beta_\phi}\right)^{q_2}}, \tag{7}$$

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 $A_{\rm sat}$ 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 recent wireless standards [6], [7] for example has the parameter set

$$\begin{aligned}&g_a = 19,\ \beta_a = 0.81,\ A_{\rm sat} = 1.4;\\ &\alpha_\phi = -48000,\ \beta_\phi = 0.123,\ q_1 = 3.8,\ q_2 = 3.7.\end{aligned} \tag{8}$$

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

$$w_k = A(|x_k|) \cdot e^{{\rm j}\cdot\left(\angle x_k + \Upsilon(|x_k|)\right)}. \tag{9}$$

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 $P_{\rm max}$ of the HPA to the average output power $P_{\rm aop}$ of the HPA output signal, given by

$${\rm OBO} = 10 \cdot \log_{10}\frac{P_{\rm max}}{P_{\rm aop}}. \tag{10}$$

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

The amplified signal $\bar{\boldsymbol{w}}$ is transmitted through the channel whose CIR coefficient vector is expressed by

$$\boldsymbol{h} = \left[h_0\ h_1\cdots h_{L_{\rm cir}}\right]^{\rm T}. \tag{11}$$

The CIR length satisfies $L_{\rm cir} \le N_{\rm cp}$. Without loss of generality, we assume that $h_0 = 1$. This is 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 $0 \le i \le L_{\rm cir}$. At the receiver, after the CP removal, the channel-impaired received signals y_k are given by

$$y_k = \sum_{i=0}^{L_{\rm cir}} h_i w_{k-i} + e_k,\ 0 \le k \le N-1, \tag{12}$$

in which $w_{k-i} = w_{N+k-i}$ for $k < i$, where $e_k = e_{k_R} + {\rm j}\cdot e_{k_I}$ is the channel AWGN with $E\{e_{k_R}^2\} = E\{e_{k_I}^2\} = \sigma_e^2$. Passing $\boldsymbol{y} = \left[y_0\ y_1\cdots y_{N-1}\right]^{\rm T}$ through the N-point FFT processor yields the FD received signal vector

$$\boldsymbol{Y} = \left[Y_0\ Y_1\cdots Y_{N-1}\right]^{\rm T} = \boldsymbol{F}\boldsymbol{y}, \tag{13}$$

where

$$\boldsymbol{F} = \frac{1}{\sqrt{N}}\begin{bmatrix}1 & 1 & \cdots & 1\\ 1 & e^{-{\rm j}2\pi/N} & \cdots & e^{-{\rm j}2\pi(N-1)/N}\\ \vdots & \vdots & \vdots & \vdots\\ 1 & e^{-{\rm j}2\pi(N-1)/N} & \cdots & e^{-{\rm j}2\pi(N-1)(N-1)/N}\end{bmatrix}, \tag{14}$$

is the FFT matrix which has the orthogonal property of $\boldsymbol{F}^{\rm H}\boldsymbol{F} = \boldsymbol{F}\boldsymbol{F}^{\rm H} = \boldsymbol{I}$. The elements of $\boldsymbol{Y}$ are given by

$$Y_n = H_n W_n + \Xi_n,\ 0 \le n \le N-1, \tag{15}$$

where $\Xi_n = \Xi_{n_R} + {\rm j}\cdot\Xi_{n_I}$ is the FD representation of the channel 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 $0 \le n \le N-1$ are given by the N-point FFT of $\boldsymbol{h}$

$$\left[H_0\ H_1\cdots H_{N-1}\right]^{\rm T} = \boldsymbol{F}\boldsymbol{h}, \tag{16}$$

while

$$\boldsymbol{W} = \left[W_0\ W_1\cdots W_{N-1}\right]^{\rm T} = \boldsymbol{F}\boldsymbol{w} \tag{17}$$

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 $\boldsymbol{\Xi} = \left[\Xi_0\ \Xi_1\cdots\Xi_{N-1}\right]^{\rm T}$, the FD received signal (15) can be expressed concisely as

$$\begin{aligned}\boldsymbol{Y} =& {\rm diag}\{H_0, H_1, \cdots, H_{N-1}\}\boldsymbol{W} + \boldsymbol{\Xi}\\ =& {\rm diag}\{H_0, H_1, \cdots, H_{N-1}\}\boldsymbol{F}\boldsymbol{w} + \boldsymbol{\Xi}.\end{aligned} \tag{18}$$

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

$$\widetilde{W}_n = \frac{Y_n}{H_n},\ 0 \le n \le N-1. \tag{19}$$

Performing the N-point inverse FFT (IFFT) on $\widetilde{\boldsymbol{W}} = \left[\widetilde{W}_0\ \widetilde{W}_1\cdots\widetilde{W}_{N-1}\right]^{\rm T}$ yields

$$\widetilde{\boldsymbol{w}} = \left[\widetilde{w}_0\ \widetilde{w}_1\cdots\widetilde{w}_{N-1}\right]^{\rm T} = \boldsymbol{F}^{\rm H}\widetilde{\boldsymbol{W}} = \boldsymbol{\Psi}(\boldsymbol{x}) + \boldsymbol{F}^{\rm H}\widetilde{\boldsymbol{\Xi}}, \tag{20}$$

where $\widetilde{\boldsymbol{\Xi}} = {\rm diag}\{H_0^{-1}, H_1^{-1}, \cdots, H_{N-1}^{-1}\}\boldsymbol{\Xi}$, and

$$\begin{aligned}\boldsymbol{\Psi}(\boldsymbol{x}) =& \left[\Psi(x_0)\ \Psi(x_1)\cdots\Psi(x_{N-1})\right]^{\rm T}\\ =& \left[w_0\ w_1\cdots w_{N-1}\right]^{\rm T}.\end{aligned} \tag{21}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2333555, IEEE Transactions on Signal Processing

If the HPA $\Psi(\ )$ at the transmitter were linear, $\widetilde{w}_k$ would be an estimate of the transmitted data symbol x_k. But $\Psi(\ )$ is nonlinear, and the linear equalisation (19) 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, an estimate of the transmitted data vector $\boldsymbol{x}$ is given by

$$\widehat{\boldsymbol{x}} = \boldsymbol{\Psi}^{-1}(\widetilde{\boldsymbol{w}}) = \left[\Psi^{-1}(\widetilde{w}_0)\ \Psi^{-1}(\widetilde{w}_1)\cdots\Psi^{-1}(\widetilde{w}_{N-1})\right]^{\mathrm{T}}. \tag{22}$$

III. Nonlinear SC-FDE of Hammerstein System

The reliable detection of the transmitted 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 $\Psi(\ )$. We adopt the CV B-spline neural network [25], [26] to represent the mapping $\widehat{w} = \widehat{\Psi}(x) : \mathbb{C} \to \mathbb{C}$ that is the estimate of the CV nonlinear function $\Psi(\ )$. We then propose two efficient algorithms for jointly estimating h_i and $\Psi(\ )$ based on this CV B-spline modelling of $\Psi(\ )$. Furthermore, we utilise another CV B-spline neural network to model $\Psi^{-1}(\ )$, the inversion of the HPA's CV nonlinearity. To estimate this inverting model requires the "input-output" training data $\{\boldsymbol{w}, \boldsymbol{x}\}$, but $\boldsymbol{w}$ is unobserved. Fortunately, as a byproduct of the Hammerstein channel identification, we can construct the pseudo training data $\widetilde{\boldsymbol{w}}$, and this allows us to estimate the inverting model. Before we proceed, we point out that the HPA $\Psi(\ )$ of (6) and (7) 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 $U_{\min} < x_{\mathrm{od}} < U_{\max}$, where $U_{\min}$ and $U_{\max}$ are known finite real values, while x_{od} symbolically represents either x_R or x_I.

A. Complex-valued B-spline neural network

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

$$U_0 < U_1 < \cdots < U_{P_o-2} < U_{P_o-1} = U_{\min} < U_{P_o} < \cdots < U_{N_{\mathrm{od}}} < U_{N_{\mathrm{od}}+1} = U_{\max} < U_{N_{\mathrm{od}}+2} < \cdots < U_{N_{\mathrm{od}}+P_o}. \tag{23}$$

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 $N_{\mathrm{od}} + 1 - P_o$. Given the set of predetermined knots (23), the set of N_{od} B-spline basis functions can be formed by using the De Boor recursion [29], yielding for $1 \le l \le N_{\mathrm{od}} + P_o$,

$$B_l^{(\mathrm{od},0)}(x_{\mathrm{od}}) = \begin{cases} 1, & \text{if } U_{l-1} \le x_{\mathrm{od}} < U_l, \\ 0, & \text{otherwise,} \end{cases} \tag{24}$$

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

$$B_l^{(\mathrm{od},p)}(x_{\mathrm{od}}) = \frac{x_{\mathrm{od}} - U_{l-1}}{U_{p+l-1} - U_{l-1}} B_l^{(\mathrm{od},p-1)}(x_{\mathrm{od}}) + \frac{U_{p+l} - x_{\mathrm{od}}}{U_{p+l} - U_l} B_{l+1}^{(\mathrm{od},p-1)}(x_{\mathrm{od}}). \tag{25}$$

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

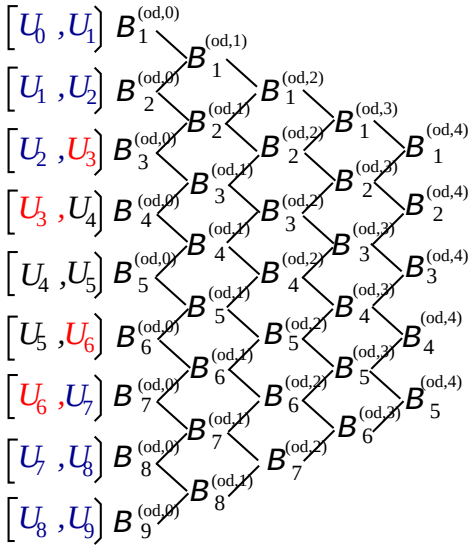


Fig. 1. Visualisation of the De Boor recursion for $P_o = 4$ and $N_{\mathrm{od}} = 5$, where $U_{\min} = U_3$ and $U_{\max} = U_6$.

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 $[U_{\min},\ U_{\max}]$, where $U_{\min}$ and $U_{\max}$ are known. The extrapolation capability of the B-spline model is influenced by the choice of the external knots. Note that there exist no data for $x_{\mathrm{od}} < U_{\min}$ and $x_{\mathrm{od}} > U_{\max}$ in identification but it is desired that the B-spline model has certain extrapolating capability outside the interval $[U_{\min},\ U_{\max}]$. The external knots can be set empirically to meet the required extrapolation capability.

Using the tensor product between the two sets of univariate B-spline basis functions [30], $B_l^{(R,P_o)}(x_R)$ 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

$$\widehat{w} = \widehat{\Psi}(x) = \sum_{l=1}^{N_R}\sum_{m=1}^{N_I} B_{l,m}^{(P_o)}(x)\theta_{l,m} = \sum_{l=1}^{N_R}\sum_{m=1}^{N_I} B_l^{(R,P_o)}(x_R) B_m^{(I,P_o)}(x_I)\theta_{l,m}, \tag{26}$$

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

Consider now using the CV B-spline neural network (26) to approximate the HPA nonlinearity $\Psi(\ )$ over one data symbol

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2333555, IEEE Transactions on Signal Processing

block $\boldsymbol{x}$. 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 (26) as

$$\boldsymbol{\theta} = \left[\theta_{1,1}\ \theta_{1,2} \cdots \theta_{l,m} \cdots \theta_{N_R,N_I}\right]^{\mathrm{T}}, \tag{27}$$

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

$$\mathbf{B} = \begin{bmatrix} B_{1,1}^{(P_o)}(x_0) & B_{1,2}^{(P_o)}(x_0) & \cdots & B_{N_R,N_I}^{(P_o)}(x_0) \\ B_{1,1}^{(P_o)}(x_1) & B_{1,2}^{(P_o)}(x_1) & \cdots & B_{N_R,N_I}^{(P_o)}(x_1) \\ \vdots & \vdots & \vdots & \vdots \\ B_{1,1}^{(P_o)}(x_{N-1}) & B_{1,2}^{(P_o)}(x_{N-1}) & \cdots & B_{N_R,N_I}^{(P_o)}(x_{N-1}) \end{bmatrix}. \tag{28}$$

Then the B-spline model (26) over $\boldsymbol{x}$ can be represented concisely by

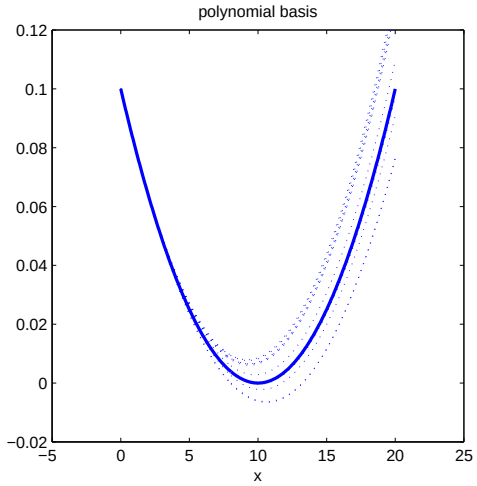
$$\widehat{\boldsymbol{w}} = \mathbf{B}\boldsymbol{\theta} \tag{29}$$

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

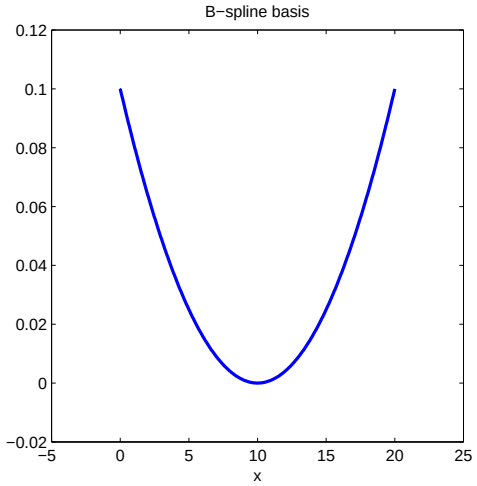
Because of the piecewise nature of B-spline functions, given a value $x \in \mathbb{C}$, there are only $P_o + 1$ basis functions with nonzero values at most for each of the real and imaginary parts. This is advantageous as P_o can be set to a quite low value, e.g. $P_o = 4$ is often sufficient. The complexity of the De Boor recursion is, therefore, on the order of P_o^2, denoted by $\mathcal{O}(P_o^2)$. Thus the computational cost of evaluating (26) scales up to about three times of the De Boor recursion, including evaluation of both real and imaginary parts as well as the tensor product calculation.

Remarks: B-splines have been widely studied in the subjects of approximation theory and numerical analysis, owing to their many excellent properties, including numerical stability. B-spline basis functions as model basis have the best approximation capability according to the Stone Weierstrass Approximation Theorem. Although any polynomial function can also be used to approximate a continuous function, the B-spline functions are proven to be optimally stable bases [31]–[33]. Specifically, a critical aspect to consider in the evaluation of a model representation is the stability with respect to perturbation of the model parameters, and 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 [31]–[33].

The excellent numerical stability of the B-spline model is demonstrated using a simple example. Fig. 2 (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 estimated exactly as a quadratic B-spline model of $\widehat{y}_R = 0.14B_1^{(R,2)}(x_R) - 0.10B_2^{(R,2)}(x_R) + 0.14B_3^{(R,2)}(x_R)$, which is depicted in Fig. 2 (b) in solid line. In any identification, the data are inevitably noisy, which will perturb the model parameters away from their true values. To simulate this noise effect, we draw three uniformly distributed random numbers from $[-0.0001,\ 0.0001]$ and add them to the three parameters in the two models, respectively. Fig. 2 depicts the ten sets of the perturbed functions in dotted line generated by perturbing the two models, respectively, in this manner. It can be clearly seen from Fig. 2 (a) that the polynomial function $y_R = 0.001x_R^2 - 0.02x_R + 0.1$ is seriously perturbed, but there is no noticeable change in Fig. 2 (b) for the quadratic B-spline model. 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.



(a) Polynomial basis



(b) B-spline basis

Fig. 2. Illustration of the superior numerical stability of the B-spline model over the polynomial model.

B. Identification of the SC-FDE Hammerstein channel

We present two identification schemes for the SC-FDE Hammerstein channel, each involving the estimation of the CIR coefficient vector $\boldsymbol{h}$ as well as the parameter vector $\boldsymbol{\theta}$ of the CV B-spline neural network (26). Consider the joint estimation of $\boldsymbol{\theta}$ and $\boldsymbol{h}$ based on a block of K training data, $\left\{x_k, y_k\right\}_{k=0}^{K-1}$, where $K \leq N$. The identification task can be formulated as the one that minimises the cost function

$$J(\boldsymbol{h}, \boldsymbol{\theta}) = \frac{1}{K}\sum_{k=0}^{K-1}\left|\widehat{e}_k\right|^2 = \frac{1}{K}\sum_{k=0}^{K-1}\left|y_k - \widehat{y}_k\right|^2, \tag{30}$$

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

$$\widehat{y}_k = \sum_{i=0}^{L_{\mathrm{cir}}} h_i \widehat{w}_{k-i} = \sum_{i=0}^{L_{\mathrm{cir}}} h_i \sum_{l=1}^{N_R}\sum_{m=1}^{N_I} B_{l,m}^{(P_o)}(x_{k-i})\theta_{l,m}, \tag{31}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2333555, IEEE Transactions on Signal Processing

where $x_{k-i} = x_{N+k-i}$ if $k < i$.

Scheme 1: Note that (31) 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. Specifically, let $\boldsymbol{y} = \begin{bmatrix} y_0 & y_1 \cdots y_{K-1} \end{bmatrix}^{\mathrm{T}}$ and $\widehat{\boldsymbol{e}} = \begin{bmatrix} \widehat{e}_0 & \widehat{e}_1 \cdots \widehat{e}_{K-1} \end{bmatrix}^{\mathrm{T}}$. Then over the training data set, the system can be represented as

$$\boldsymbol{y} = \mathbf{P}\boldsymbol{h} + \widehat{\boldsymbol{e}} = \mathbf{Q}\boldsymbol{\theta} + \widehat{\boldsymbol{e}}, \tag{32}$$

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

$$\mathbf{P} = \begin{bmatrix} \widehat{w}_0 & \widehat{w}_{-1} & \cdots & \widehat{w}_{-L_{\mathrm{cir}}} \\ \vdots & \vdots & \vdots & \vdots \\ \widehat{w}_k & \widehat{w}_{k-1} & \cdots & \widehat{w}_{k-L_{\mathrm{cir}}} \\ \vdots & \vdots & \vdots & \vdots \\ \widehat{w}_{K-1} & \widehat{w}_{K-2} & \cdots & \widehat{w}_{K-1-L_{\mathrm{cir}}} \end{bmatrix}, \tag{33}$$

$$\mathbf{Q} = \begin{bmatrix} \varphi_{1,1}(0) & \cdots & \varphi_{l,m}(0) & \cdots & \varphi_{N_R,N_I}(0) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \varphi_{1,1}(k) & \cdots & \varphi_{l,m}(k) & \cdots & \varphi_{N_R,N_I}(k) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \varphi_{1,1}(K-1) & \cdots & \varphi_{l,m}(K-1) & \cdots & \varphi_{N_R,N_I}(K-1) \end{bmatrix}, \tag{34}$$

in which

$$\widehat{w}_k = \widehat{\Psi}(x_k) = \sum_{l=1}^{N_R} \sum_{m=1}^{N_I} B_{l,m}^{(P_o)}(x_k)\theta_{l,m}, \tag{35}$$

$$\varphi_{l,m}(k) = \sum_{i=0}^{L_{\mathrm{cir}}} h_i B_{l,m}^{(P_o)}(x_{k-i}), \tag{36}$$

with $x_k = x_{N+k}$ if $k < 0$. For the cost function (30) and the model (32), according to [34], [35], the estimates of $\boldsymbol{\theta}$ and $\boldsymbol{h}$ are unbiased, irrespective to the optimization algorithm used.

We adopt the following ALS procedure to estimate $\boldsymbol{h}$ and $\boldsymbol{\theta}$, which is a coordinate gradient descent algorithm [36], [37]. However, unlike a generic coordinate gradient descent algorithm, in our case we have the closed-form solutions for both $\boldsymbol{h}$ and $\boldsymbol{\theta}$, and 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 $\mathbf{P}$ of (33). Calculate $\boldsymbol{h}$ as the LS estimate given by

$$\widehat{\boldsymbol{h}}^{(0)} = \left(\mathbf{P}^{\mathrm{H}}\mathbf{P}\right)^{-1}\mathbf{P}^{\mathrm{H}}\boldsymbol{y}. \tag{37}$$

Then obtain $\widehat{\boldsymbol{h}}^{(0)}$ by normalising $\widehat{h}_i^{(0)} \leftarrow \widehat{h}_i^{(0)}/\widehat{h}_0^{(0)}$ for $0 \le i \le L_{\mathrm{cir}}$.

TD 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 $\mathbf{Q}$ of (34). The LS estimate of $\widehat{\boldsymbol{\theta}}^{(\tau)}$ is readily given by

$$\widehat{\boldsymbol{\theta}}^{(\tau)} = \left(\mathbf{Q}^{\mathrm{H}}\mathbf{Q}\right)^{-1}\mathbf{Q}^{\mathrm{H}}\boldsymbol{y}. \tag{38}$$

b) For $\mathbf{P}$ of (33), fix $\widehat{w}_k$ according to (35) based on $\widehat{\boldsymbol{\theta}}^{(\tau)}$. Calculate

$$\widehat{\boldsymbol{h}}^{(\tau)} = \left(\mathbf{P}^{\mathrm{H}}\mathbf{P}\right)^{-1}\mathbf{P}^{\mathrm{H}}\boldsymbol{y}. \tag{39}$$

Then obtain $\widehat{\boldsymbol{h}}^{(\tau)}$ by normalising $\widehat{h}_i^{(\tau)} \leftarrow \widehat{h}_i^{(\tau)}/\widehat{h}_0^{(\tau)}$ 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 at least a local minimum solution for minimising the cost function (30).

Scheme 2: Our second algorithm is based on the following two linear regression models to represent the HPA and channel respectively as

$$\widetilde{\boldsymbol{w}} = \widehat{\boldsymbol{\Psi}}(\boldsymbol{x}) + \widetilde{\boldsymbol{e}} = \mathbf{B}\boldsymbol{\theta} + \boldsymbol{F}^{\mathrm{H}}\widehat{\boldsymbol{\Xi}}, \tag{40}$$

$$\boldsymbol{y} = \mathbf{P}\boldsymbol{h} + \widehat{\boldsymbol{e}}, \tag{41}$$

in which the real-valued B-spline basis function matrix $\mathbf{B}$ is given in (28). This scheme therefore requires $K = N$. Note that unlike the white $\widetilde{\boldsymbol{\Xi}}$ given in (20), $\widehat{\boldsymbol{\Xi}}$ in (40) is coloured. The joint estimation of $\boldsymbol{h}$ and $\boldsymbol{\theta}$ can be carried out by minimizing $\widehat{\boldsymbol{\Xi}}^{\mathrm{T}}\widehat{\boldsymbol{\Xi}}$ and $\widehat{\boldsymbol{e}}^{\mathrm{T}}\widehat{\boldsymbol{e}}$ using the coordinate descent algorithm given in the following.

Initialisation. Initialise $\widehat{w}_k = x_k$ in $\mathbf{P}$ of (33). Calculate $\boldsymbol{h}$ as the LS estimate given by $\widehat{\boldsymbol{h}}^{(0)}$ of (37). Then obtain $\widehat{\boldsymbol{h}}^{(0)}$ by normalising $\widehat{h}_i^{(0)} \leftarrow \widehat{h}_i^{(0)}/\widehat{h}_0^{(0)}$ for $0 \le i \le L_{\mathrm{cir}}$. Also calculate $\mathbf{D} = \left(\mathbf{B}^{\mathrm{T}}\mathbf{B}\right)^{-1}\mathbf{B}^{\mathrm{T}}$.

TD-FD ALS estimation. For $1 \le \tau \le \tau_{\max}$, where $\tau_{\max}$ is the maximum number of iterations, perform:

a) Apply the N-point FFT to $\widehat{\boldsymbol{h}}^{(\tau-1)}$ to yield the current estimate of H_n for $0 \le n \le N-1$, and then apply the FD one-tap equalisation using (19) to yield the current estimate of $\widetilde{\boldsymbol{W}}$. Next apply the N-point IFFT on the FD estimate of $\widetilde{\boldsymbol{W}}$ to yield the TD estimate of $\widetilde{\boldsymbol{w}}$, and then compute

$$\widehat{\boldsymbol{\theta}}^{(\tau)} = \mathbf{D}\widetilde{\boldsymbol{w}}. \tag{42}$$

b) For $\mathbf{P}$ of (33), fix $\widehat{w}_k$ according to (35) based on $\widehat{\boldsymbol{\theta}}^{(\tau)}$. Calculate $\widehat{\boldsymbol{h}}^{(\tau)}$ of (39). Then obtain $\widehat{\boldsymbol{h}}^{(\tau)}$ by normalising $\widehat{h}_i^{(\tau)} \leftarrow \widehat{h}_i^{(\tau)}/\widehat{h}_0^{(\tau)}$ 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.

Remarks: The two schemes differ in their step a). The TD-FD ALS estimation scheme (Scheme 2) is more efficient because there is no need of iterative matrix inversion to calculate $\widehat{\boldsymbol{\theta}}^{(\tau)}$.

Scheme 2 requires the training data in a full transmitting block. This is because in step a) the FDE is performed based on the full transmitting block. While the TD ALS estimation (Scheme 1) does not have this restriction and K can be smaller than N. However in step b) of Scheme 2, $\mathbf{P}$ can also be formed in the same way as of Scheme 1 by only using K samples, so that there can be the same computational costs for step b) of the two schemes. Table I summarises the computational complexity of these two ALS algorithms.

The solution $\widehat{\boldsymbol{\theta}}^{(\tau)}$ of the proposed TD-FD ALS estimation scheme given in (42) is a statistically suboptimal solution minimizing the equation error $\widehat{\boldsymbol{\Xi}}^{\mathrm{T}}\widehat{\boldsymbol{\Xi}}$. Alternatively the maximum likelihood (ML) estimator is given by

$$\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}^{(\tau)} = \left(\mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{B}\right)^{-1}\mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}\widetilde{\boldsymbol{w}} \tag{43}$$

in which $\boldsymbol{\Sigma} = \boldsymbol{F}^{\mathrm{H}}\mathrm{diag}\left\{H_0^{\mathrm{H}}H_0, H_1^{\mathrm{H}}H_1, \cdots, H_{N-1}^{\mathrm{H}}H_{N-1}\right\}\boldsymbol{F}$ is the inverse of the covariance matrix of the coloured noise

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2333555, IEEE Transactions on Signal Processing

TABLE I
COMPUTATIONAL COMPLEXITY OF THE TWO ALS ALGORITHMS.

Scheme 1	$\mathcal{O}\big(K^2 \cdot L_{\rm cir}\big) + \mathcal{O}\big(L_{\rm cir}^3\big) + \tau_{\max}\Big(\mathcal{O}\big(N_B \cdot K\big) + \mathcal{O}\big(N_B^2 \cdot K\big) + \mathcal{O}\big(N_B^3\big) + \mathcal{O}\big(L_{\rm cir}^2 \cdot K\big) + \mathcal{O}\big(L_{\rm cir}^3\big)\Big)$
Scheme 2	$\mathcal{O}\big(K^2 \cdot L_{\rm cir}\big) + \mathcal{O}\big(L_{\rm cir}^3\big) + \mathcal{O}\big(N_B^2 \cdot N\big) + \mathcal{O}\big(N_B^3\big) + \tau_{\max}\Big(\mathcal{O}\big(N \cdot \log_2 N\big) + \mathcal{O}\big(N \cdot N_B\big) + \mathcal{O}\big(L_{\rm cir}^2 \cdot K\big) + \mathcal{O}\big(L_{\rm cir}^3\big)\Big)$

term $\boldsymbol{F}^{\rm H}\widehat{\boldsymbol{\Xi}}$. Clearly, the optimal ML estimator $\widehat{\boldsymbol{\theta}}_{\rm ML}^{(\tau)}$ (43) is computationally more expensive than the proposed (42).

We now perform a simple analysis on the statistical properties of the estimator (42) based on an assumed "true" system given by (20) in which $\widetilde{\boldsymbol{w}}$ is obtained from the "true" H_n and $\boldsymbol{\Psi}(\boldsymbol{x})$ is parametrized by the B-spline neural network. Under this idealised condition, we have

$$\widetilde{\boldsymbol{w}} = \mathbf{B}\boldsymbol{\theta} + \boldsymbol{F}^{\rm H}{\rm diag}\{H_0^{-1}, H_1^{-1}, \cdots, H_{N-1}^{-1}\}\boldsymbol{\Xi}, \tag{44}$$

Substituting (44) into (42) leads to

$$\widehat{\boldsymbol{\theta}}^{(\tau)} = \boldsymbol{\theta} + \big(\mathbf{B}^{\rm T}\mathbf{B}\big)^{-1}\mathbf{B}^{\rm T}\boldsymbol{F}^{\rm H}{\rm diag}\{H_0^{-1}, H_1^{-1}, \cdots, H_{N-1}^{-1}\}\boldsymbol{\Xi}. \tag{45}$$

We have $E\big\{\widehat{\boldsymbol{\theta}}^{(\tau)}\big\} = \boldsymbol{\theta}$ owing to the facts that $E\{\boldsymbol{\Xi}\} = \mathbf{0}$ and $\boldsymbol{\Xi}$ is uncorrelated with the regressors, i.e. B-spline basis functions. Thus, no bias is introduced to the estimator (42). Similarly, the covariance matrix of $\widehat{\boldsymbol{\theta}}^{(\tau)}$ is given by

$$\begin{aligned} &E\Big\{\big(\widehat{\boldsymbol{\theta}}^{(\tau)} - \boldsymbol{\theta}\big)\big(\widehat{\boldsymbol{\theta}}^{(\tau)} - \boldsymbol{\theta}\big)^{\rm H}\Big\} = 2\sigma_e^2 E\Big\{\big(\mathbf{B}^{\rm T}\mathbf{B}\big)^{-1}\mathbf{B}^{\rm T}\boldsymbol{F}^{\rm H} \\ &\times {\rm diag}\big\{\big(H_0^{\rm H}H_0\big)^{-1}, \big(H_1^{\rm H}H_1\big)^{-1}, \cdots, \big(H_{N-1}^{\rm H}H_{N-1}\big)^{-1}\big\} \\ &\times \boldsymbol{F}\mathbf{B}\big(\mathbf{B}^{\rm T}\mathbf{B}\big)^{-1}\Big\} > E\Big\{\big(\widehat{\boldsymbol{\theta}}_{\rm ML}^{(\tau)} - \boldsymbol{\theta}\big)\big(\widehat{\boldsymbol{\theta}}_{\rm ML}^{(\tau)} - \boldsymbol{\theta}\big)^{\rm T}\Big\} \\ &= 2\sigma_e^2 E\Big\{\big(\mathbf{B}^{\rm T}\boldsymbol{\Sigma}\mathbf{B}\big)^{-1}\Big\}. \end{aligned} \tag{46}$$

Provided that $\mathbf{B}$ is full rank and none of H_n is zero, the covariance matrix of $\widehat{\boldsymbol{\theta}}^{(\tau)}$ converges to $\mathbf{0}$ as $N \to \infty$.

C. Inversion of the Hammerstein channel's static nonlinearity

Given the CV Hammerstein channel's static nonlinearity $\Psi(\ )$, we wish to compute its inverse defined by $x_k = \Psi^{-1}(w_k)$ in order to complete the nonlinear SC-FDE. We adopt the strategy of 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 Hammerstein channel identification presented in Section III-B, we already obtain an estimate for $\boldsymbol{w}$ as $\widehat{\boldsymbol{w}} = \mathbf{B}\boldsymbol{\theta}^{(\tau_{\max})}$. Therefore, we may construct the pseudo training data set $\big\{\widehat{w}_k, x_k\big\}_{k=0}^{N-1}$ to estimate $\boldsymbol{\alpha}$.

More specifically, define two knots sequences similar to (23) for w_R and w_I, respectively. Similar to (26), we have[1]

$$\begin{aligned} \widehat{x} =& \widehat{\Phi}(w;\boldsymbol{\alpha}) = \sum_{l=1}^{N_R}\sum_{m=1}^{N_I} B_{l,m}^{(P_o)}(w)\alpha_{l,m} \\ =& \sum_{l=1}^{N_R}\sum_{m=1}^{N_I} B_l^{(R,P_o)}(w_R)B_m^{(I,P_o)}(w_I)\alpha_{l,m}, \end{aligned} \tag{47}$$

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

where $B_l^{(R,P_o)}(w_R)$ and $B_m^{(I,P_o)}(w_I)$ are respectively calculated based on (24) and (25), while

$$\boldsymbol{\alpha} = \big[\alpha_{1,1}\ \alpha_{1,2}\cdots\alpha_{l,m}\cdots\alpha_{N_R,N_I}\big]^{\rm T}. \tag{48}$$

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

$$\widetilde{\mathbf{B}} = \begin{bmatrix} B_{1,1}^{(P_o)}(\widehat{w}_0) & B_{1,2}^{(P_o)}(\widehat{w}_0) & \cdots & B_{N_R,N_I}^{(P_o)}(\widehat{w}_0) \\ B_{1,1}^{(P_o)}(\widehat{w}_1) & B_{1,2}^{(P_o)}(\widehat{w}_1) & \cdots & B_{N_R,N_I}^{(P_o)}(\widehat{w}_1) \\ \vdots & \vdots & \vdots & \vdots \\ B_{1,1}^{(P_o)}(\widehat{w}_{N-1}) & B_{1,2}^{(P_o)}(\widehat{w}_{N-1}) & \cdots & B_{N_R,N_I}^{(P_o)}(\widehat{w}_{N-1}) \end{bmatrix}. \tag{49}$$

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

IV. SIMULATION STUDY

We considered a Hammerstein SC-FDE System in which the HPA employed was described by (6) and (7) with the parameter set given in (8). The size of the transmitted data block was set to $N = 2048$ and 64-QAM was used. We assumed a quasi-static Rayleigh multipath channel with an exponentially decreasing power delay profile, where the average gain for the lth path was given by

$$\mathrm{E}\big\{|h_l|\big\} = e^{-\frac{l}{\eta}},\ 0 \le l \le L_{\rm cir}, \tag{50}$$

with η being the channel degradation factor. In the simulation study, we set $\eta = 3$ and $L_{\rm cir} = 9$. The CIR coefficients h_l for $0 \le l \le L_{\rm cir}$ remained constant during the communication session. Note that the effective system throughput is given by

$$\text{Effective throughput} = \frac{N}{N + N_{\rm cp}}, \tag{51}$$

where the CP length $N_{\rm cp} \ge L_{\rm cir} = 9$. The larger the data block length N is, the more bandwidth efficient the system is. We also tested $N = 1024$, and the results obtained, not shown here, are very similar to the results obtained for $N = 2048$.

We used a full data block with $K = N = 2048$ training samples in the joint estimation of the CV CIR coefficient vector $\boldsymbol{h}$ and the CV parameter vector $\boldsymbol{\theta}$ of the B-spline model for $\Psi(\ )$ as well as the estimation of the CV 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, since $P_o = 4$ is sufficient for most practical applications. The number of B-spline basis functions was set to $N_R = N_I = 8$, because 8 basis functions is sufficient to partitioning or covering the input interval $\big[-d(\sqrt{M}-1),\ d(\sqrt{M}-1)\big]$. As explained in Section III-A, the B-spline

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2333555, IEEE Transactions on Signal Processing

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

	True Parameters	Estimated parameters by Scheme 1		Estimated parameters by Scheme 2	
		$E_{\rm b}/N_{\rm o}=4$ dB OBO $=3$ dB	$E_{\rm b}/N_{\rm o}=10$ dB OBO $=2$ dB	$E_{\rm b}/N_{\rm o}=4$ dB OBO $=3$ dB	$E_{\rm b}/N_{\rm o}=10$ dB OBO $=2$ dB
h_0	1	1	1	1	1
h_1	$-0.2145-$j0.1867	$-0.2180-$j0.1814	$-0.2180-$j0.1814	$-0.2139-$j0.1871	$-0.2141-$j0.1869
h_2	$0.0399+$j0.3675	$0.0395+$j0.3678	$0.0395+$j0.3678	$0.0405+$j0.3675	$0.0403+$j0.3675
h_3	$-0.0900+$j0.4053	$-0.0923+$j0.4049	$-0.0923+$j0.4049	$-0.0899+$j0.4058	$-0.0899+$j0.4056
h_4	$-0.0893+$j0.1287	$-0.0863+$j0.1269	$-0.0862+$j0.1269	$-0.0893+$j0.1285	$-0.0893+$j0.1286
h_5	$-0.1117+$j0.3035	$-0.1145+$j0.3033	$-0.1145+$j0.3033	$-0.1118+$j0.3033	$-0.1117+$j0.3034
h_6	$-0.0766-$j0.0264	$-0.0741-$j0.0295	$-0.0741-$j0.0295	$-0.0768-$j0.0265	$-0.0767-$j0.0265
h_7	$0.0623-$j0.0668	$0.0584-$j0.0596	$0.0584-$j0.0596	$0.0627-$j0.0666	$0.0625-$j0.0667
h_8	$0.0282+$j0.0324	$0.0346+$j0.0306	$0.0346+$j0.0306	$0.0273+$j0.0323	$0.0276+$j0.0323
h_9	$-0.0395-$j0.0291	$-0.0405-$j0.0237	$-0.0405-$j0.0237	$-0.0396-$j0.0287	$-0.0396-$j0.0289

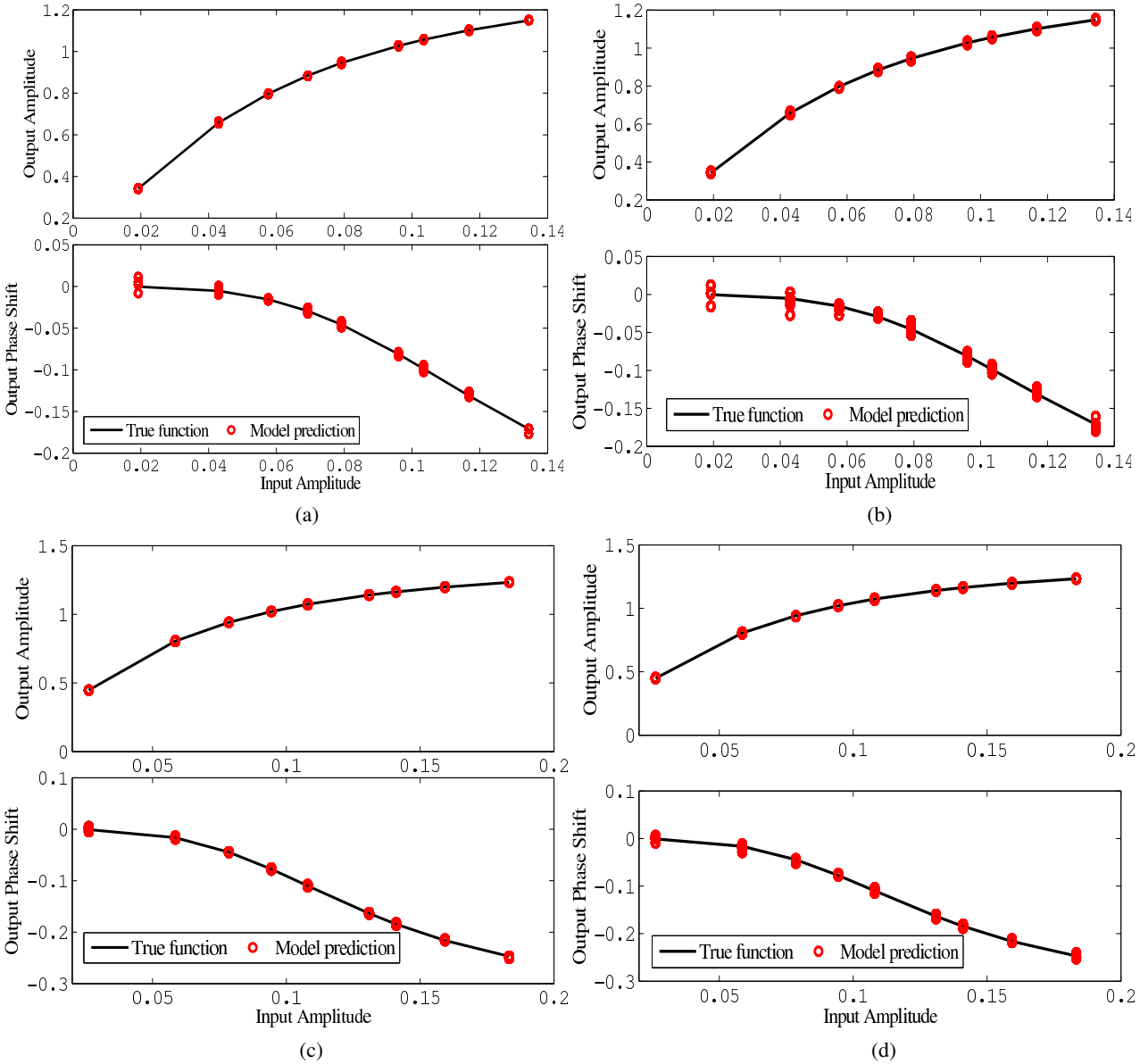


Fig. 3. Comparison of the HPA's nonlinearity $\Psi(\ )$ and the estimated nonlinearity $\widehat{\Psi}(\ )$: (a) Scheme 1 under OBO$=3$ dB, $E_{\rm b}/N_{\rm o}=4$ dB; (b) Scheme 2 under OBO$=3$ dB, $E_{\rm b}/N_{\rm o}=4$ dB; (c) Scheme 1 under OBO$=2$ dB, $E_{\rm b}/N_{\rm o}=10$ dB; and (d) Scheme 2 under OBO$=2$ dB, $E_{\rm b}/N_{\rm o}=10$ dB.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2333555, IEEE Transactions on Signal Processing

model has the desired optimal robustness property. We also tested $P_o = 3$ and $P_o = 5$ as well as $N_R = N_I = 6$ and $N_R = N_I = 10$, the corresponding results obtained, not shown here, are very similar to the results presented here. 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. II. The system's signal-to-noise ratio (SNR) was defined as 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.

For both Schemes 1 and 2, the identification experiments were conducted under two combinations of the HPA operating region and the SNR condition, which were set as OBO $= 3$ dB with SNR $= 4$ dB and OBO $= 2$ dB with SNR $= 10$ dB, respectively. The identification results of the linear subsystem in the Hammerstein channel under these two experimental conditions obtained by the two schemes are summarised in Table III, while the modelling results of the HPA static nonlinearity $\Psi(\ )$ by the B-spline neural network $\widehat{\Psi}(\ )$ for the given simulation conditions achieved by the two schemes are illustrated in Fig. 3. It can be seen from Table III that the CIR estimates obtained by the two schemes achieve high accuracy for the both system operating conditions. The results of Fig. 3 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 OBO $= 3$ dB and SNR $= 4$ dB.

The combined responses of the HPA's true nonlinearity and its estimated inversion obtained by the two schemes under the operating condition of OBO $= 3$ dB and SNR $= 4$ dB are depicted in Fig. 4. The results of Fig. 4 demonstrate the capability of the CV B-spline neural network to accurately model the inversion of the HPA's nonlinearity based only on the pseudo training data. More specifically, the results of Fig. 4 clearly show that the combined response of the HPA's nonlinearity $\Psi(\cdot)$ and its estimated inversion $\Phi(\cdot)$ satisfies

$$\Phi\big(\Psi(x)\big) \approx x, \tag{52}$$

where x denotes the input to the HPA. That is, the magnitude of the combined response is $\big|\Phi\big(\Psi(x)\big)\big| \approx |x|$ and the phase

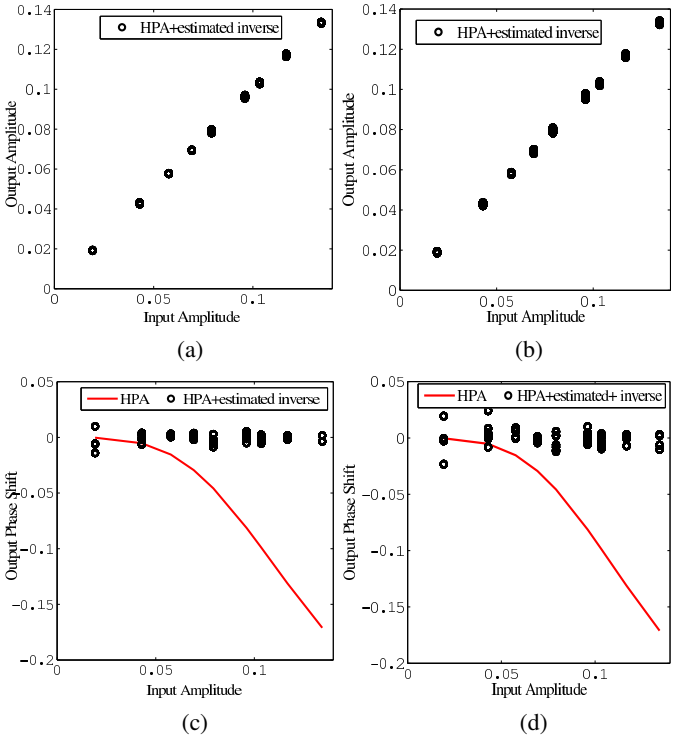


Fig. 4. Combined response of the true HPA and its estimated inversion obtained under OBO $= 3$ dB and $E_{\rm b}/N_{\rm o} = 4$ dB: (a) combined amplitude response by Scheme 1; (b) combined amplitude response by Scheme 2; (c) combined phase response by Scheme 1; and (d) combined phase response by Scheme 2.

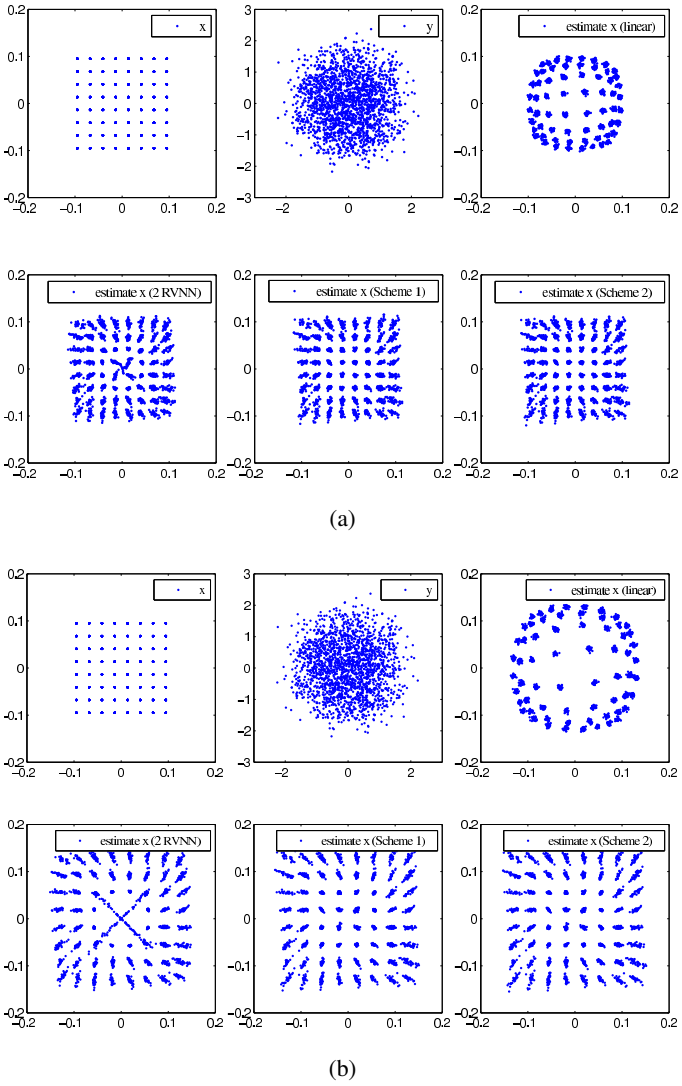


Fig. 5. Effectiveness of the proposed nonlinear SC-FDE scheme based on the estimated CIR $\widehat{\boldsymbol{h}}$ and the estimated HPA's CV static nonlinearity as well as the estimated inverse mapping for the HPA's CV nonlinearity under: (a) OBO $=$ 3 dB, $E_{\rm b}/N_{\rm o} = 4$ dB; and (b) OBO $= 2$ dB, $E_{\rm b}/N_{\rm o} = 10$ dB. The top three plots in the two sub-figures (a) and (b) depict one transmitted QAM symbol block $\boldsymbol{x}$, its received signal block $\boldsymbol{y}$, and the corresponding estimated $\widehat{\boldsymbol{x}}$ obtained by the linear SC-FDE. The bottom three plots in (a) and (b) show the estimated $\widehat{\boldsymbol{x}}$ obtained by the previous 2RV NN based SC-FDE [38] and the current CVNN based SC-FDE using the two proposed schemes, respectively.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2333555, IEEE Transactions on Signal Processing

shift of the combined response is approximately zero. In other words, $\Phi(\cdot)$ is an accurate inversion of $\Psi(\cdot)$.

The effectiveness of the proposed nonlinear SC-FDE scheme based on the CV B-spline neural network approach is illustrated in Fig. 5, where the nonlinear SC-FDE was constructed based on the estimated CIR $\widehat{\boldsymbol{h}}$ and the inverse mapping $\Phi(\ ) = \widehat{\Psi}^{-1}(\ )$ obtained under the two operating conditions. The standard linear SC-FDE scheme is also illustrated in Fig. 5 for comparison. As can be seen clearly in Fig. 5, the linear SC-FDE cannot compensate the nonlinear distortions of the Hammerstein channel. In the recent work [38], we developed a nonlinear SC-FDE approach which uses two real-valued (RV) B-spline neural networks to model the HPA's amplitude response $A(\ )$ and phase response $\Upsilon(\ )$, respectively, as well as uses another RV B-spline neural network to model the inverse mapping of the HPA's amplitude response $A^{-1}(\ )$. The computational complexity of this two RV B-spline neural networks (2RV NN) based approach is similar to our CV B-spline neural network (CVNN) based approach. The equalisation results obtained by the 2RV NN based SC-FDE scheme [38] are also depicted in Fig. 5 for the purpose of comparison. From Fig. 5, we observe that the current CVNN approach attains slightly better equalisation than this previous 2RV NN approach. This is expected. The CVNN based estimator with the LS cost function adopted in this paper is the ML estimator (MLE) under the assumption of the CV Gaussian noise. The MLE based on the amplitude and phase response (AM/PM) systems is very complicated, involving nonlinear transformations, and this MLE needs nonlinear optimisation. In this paper, we solve this nonlinear optimisation with an ALS procedure. The 2RV NN based scheme developed in [38] is simpler, in which the two LS cost functions are defined based on the AM/PM responses separately, but the resulting model residuals no longer follow Gaussian distribution. Consequently, the estimator or the optimisation of [38] is no longer statistically optimal, rather the scheme of [38] is only an approximation for the purpose of computational convenience, which can cause errors at extreme points. For example, an input data point at very low amplitude will not contribute much to the cost function in the CVNN model. But the same data point will influence the RV NN based cost function a lot more if there is a high phase error caused by the channel AWGN.

The achievable BER performance of the proposed CVNN based nonlinear SC-FDE are plotted in Fig. 6 under the three different operating conditions of the HPA, in comparison to the BER performance obtained by the standard linear SC-FDE and the 2RV NN based nonlinear SC-FDE [38]. Clearly, the standard linear SC-FDE is incapable of compensating the nonlinear distortions of the Hammerstein channel and its attainable BER performance is very poor even under the HPA operating condition of OBO $= 5$ dB, as can be seen from Fig. 6. By contrast, the proposed two CVNN based nonlinear SC-FDEs based on the estimated CIR and the inverse mapping of the HPA are able to compensate most of the nonlinear distortions and attain a much better BER performance. Observe that the two proposed CVNN based SC-FDEs outperform the 2RV NN based SC-FDE of [38] under the adverse conditions of OBO $= 2$ dB and 3 dB.

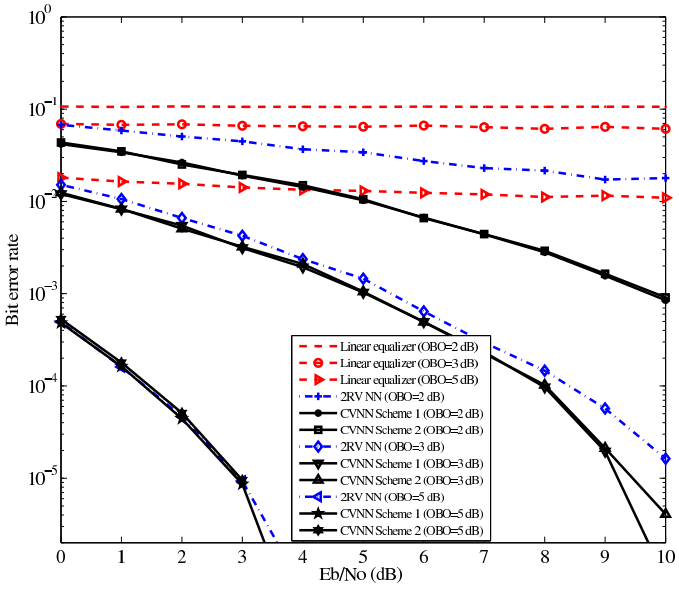


Fig. 6. The bit error rate performance comparison of the proposed two CVNN based nonlinear SC-FDE schemes with the standard linear SC-FDE as well as the 2RV NN based nonlinear SC-FDE [38].

V. Conclusions

A novel nonlinear equalisation scheme has been developed for the complex-valued Hammerstein SC-FDE 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 CV static nonlinearity as well as to use another CV B-spline neural network for modelling the inverse mapping of the HPA's nonlinearity. Our novel contribution includes deriving two highly efficient alternating least squares algorithms for estimating the CIR coefficients and the parameters of the CV 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 CV 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 CV Hammerstein channel as well as the inverse mapping of the channel's CV static nonlinearity. The results obtained also confirm the effectiveness of the proposed nonlinear equaliser constructed based on the estimated CIR and inverse B-spline mapping.

References

[1] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Communications Magazine*, 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, UK: Wiley, 2003.

[3] A. A. M. Saleh, "Frequency-independent and frequency-dependent nonlinear models of TWT amplifiers," *IEEE Trans. Communications*, 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), Sept. 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. Microwave Theory and Techniques*, 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. 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] D. Falconer, S. L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, "Frequency domain equalization for single carrier broadband wireless systems," *IEEE Communications Magazine*, vol. 40, no. 4, pp. 58–66, April 2002.

[9] F. Pancaldi, G. M. Vitetta, R. Kalbasi, N. Al-Dhahir, M. Uysal, and H. Mheidat, "Single-carrier frequency domain equalization," *IEEE Signal Processing Magazine*, vol. 25, no. 5, pp. 37–56, Sept. 2008.

[10] E. Seidel, "Progress on 'LTE Advanced' - the new 4G standard," Nomor Research GmbH: White Paper on LTE Advanced, 2008.

[11] 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.

[12] P. M. Grant, S. McLaughlin, H. Aghvami, and S. Fletcher, "Green radio – towards sustainable wireless networks," *Mobile VCE Core 5 Programme* Presentation, April 2009. Available online from http://www.ee.princeton.edu/seminars/iss/Spring2009/slides/grant.pdf

[13] S. Kato, H. Harada, R. Funada, T. Baykas, C. S. Sum, J. Wang, and M. A. Rahman, "Single-carrier transmission for multi-gigabit 60–GHz WPAN systems," *IEEE J. Selected Areas on Communications*, vol. 27, no. 8, pp. 1466–1478, Oct. 2009.

[14] S. J. Lee and W. Y. Lee, "Capacity of millimetre-wave multiple-input multiple-output channels in a conference room," *IET Communications*, vol. 6, no. 17, pp. 2879–2885, Nov. 2012.

[15] S. Rangan, T. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, March, 2014.

[16] B. Razavi, "Design of millimeter-wave CMOS radio: A tutorial," *IEEE Trans. Circuits and Systems I*, vol. 56, no. 1, pp. 4–16, Jan. 2009.

[17] S. K. Yong, P. Xia, and A. Valdes Garcia, eds., *60 GHz Technology for Gbps WLAN and WPAN: From Theory to Practice*. Chichester, UK: John Wiley & Son Ltd, 2011.

[18] 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.

[19] 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.

[20] 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, March 2008.

[21] 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, June 2009.

[22] 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, March 2011.

[23] 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.

[24] 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, June 2013.

[25] 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.

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

[27] C. Zhang, Z. Xiao, B. Gao, L. Su, and D. Jin, "Power amplifier non-linearity treatment with distortion constellation estimation and demodulation for 60 GHz single-carrier frequency-domain equalisation transmission," *IET Communications*, vol. 8, no. 3, pp. 278–286, Feb. 2014.

[28] S. Chen, X. Hong, J. Gao, and C. J. Harris, "Complex-valued B-spline neural networks for modeling and inverting Hammerstein systems," *IEEE Trans. Neural Networks and Learning Systems*, to appear, 2014.

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

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

[33] 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.

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

[35] C.-F. Wu, "Asymptotic theory of nonlinear least squares estimation," *The Annals of Statistics*, vol. 9, no. 3, pp. 501–513, 1981.

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

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

[38] X. Hong, S. Chen, and C. J. Harris, "B-Spline neural network based single-carrier frequency domain equalisation for Hammerstein channels," in *Proc. WCCI 2014* (Beijing, China), July 6-11, 2014, pp. 1-8.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2014.2333555, IEEE Transactions on Signal Processing

Xia Hong (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).

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.

Emad F. Khalaf received his BEng and MEng degrees in IT from Wroclaw University of Technology in Poland, in 1992, as one certificate, and the PhD degree in Computer networks from Wroclaw University of Technology, in Poland, in 2002.
From 2003 to 2011, he worked as an assistant professor at Computer Engineering Department, Faculty of Engineering, Philadelphia University, in Jordan. Since 2012 he is an assistant professor at Electrical and Computer Engineering Department, Faculty of Engineering, King Abdulaziz University, Jeddah, Saudi Arabia. Dr Khalaf's research interests are in network security and cryptography, speech classification and recognition.