

Single-Carrier Frequency-Domain Equalization with Hybrid Decision Feedback Equalizer for Hammerstein Channels Containing Nonlinear Transmit Amplifier

Sheng Chen, *Fellow, IEEE*, Xia Hong, *Senior Member, IEEE*, Emad Khalaf, Ali Morfeq, Naif D. Alotaibi and Chris J. Harris

Abstract—We propose a nonlinear hybrid decision feedback equalizer (NHDFE) for single-carrier (SC) block transmission systems with nonlinear transmit high power amplifier (HPA), which significantly outperforms our previous nonlinear SC frequency-domain equalization (NFDE) design. To obtain the coefficients of the channel impulse response (CIR) as well as to estimate the nonlinear mapping and the inverse nonlinear mapping of the HPA, we adopt a complex-valued (CV) B-spline neural network approach. Specifically, we use a CV B-spline neural network to model the nonlinear HPA, and we develop an efficient alternating least squares scheme for estimating the parameters of the Hammerstein channel, including both the CIR coefficients and the parameters of the CV B-spline model. We also adopt another CV B-spline neural network to model the inversion of the nonlinear HPA, and the parameters of this inverting B-spline model can be estimated using the least squares algorithm based on the pseudo training data obtained as a natural byproduct of the Hammerstein channel identification. The effectiveness of our NHDFE design is demonstrated in a simulation study, which shows that the NHDFE achieves a signal-to-noise ratio gain of 4 dB over the NFDE at the bit error rate level of 10^{-4} .

Index Terms—Single-carrier block transmission, decision feedback equalizer, nonlinear high power amplifier, Hammerstein channel, complex-valued B-spline neural network

I. INTRODUCTION

Traditional time-domain (TD) equalization suffers from the drawbacks of poor performance and excessively slow convergence for the high-data-rate applications in fourth generation (4G) and beyond 4G (B4G) systems, because the intersymbol interference (ISI) of these wireless channels spans over tens or even hundreds of symbols. Orthogonal frequency-division multiplexing (OFDM) [1], [2], which provides a low-complexity high-performance solution for mitigating long ISI, 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), and a transmitted OFDM

signal can be seriously distorted by the high power amplifier (HPA) at transmitter, which exhibits nonlinear saturation characteristics [3]–[7]. A viable alternative solution for long ISI mitigation is single-carrier (SC) block transmission with frequency-domain equalization (FDE) [8], [9]. Although the total complexity of a SC-FDE transceiver is the same as that of an OFDM transceiver, the SC-FDE transmitter does not require the fast Fourier transform (FFT) operation. Therefore, SC block transmission is particularly suited for uplink implementation where the transmitter is a pocket-size handset with a limited hardware and computational capacity.

SC-FDE and OFDM essentially adopt linear equalization techniques, and they suffer from considerable bit error rate (BER) performance degradation for transmission over deep frequency-selective fading channels. An effective means of overcoming this problem associated with SC-FDE is to employ decision feedback equalizer (DFE) [8]–[20]. Various SC-FDE-DFE structures can be roughly divided into two categories. The first category of SC-FDE-DFE has a hybrid equalization structure, whereas SC-FDE, which acts as feedforward equalizer, is augmented with a TD transversal filter for feedback equalization [8]–[14]. Calculating the tap vector of the TD feedback filter requires a matrix inversion operation, which is computationally expensive, particularly for long TD feedback filters. We refer to this class as hybrid DFE (HDFE). The second category of SC-FDE-DFE adopts an iterative block DFE (IBDFE) in frequency domain (FD) [15]–[20]. This structure is attractive since both the feedforward and feedback equalizers operate in the FD and no matrix inversion operation is required. However, the detection complexity, particularly the detection delay, of IBDFE is higher than that of HDFE, owing to the nature of iterative block detection operation.

To achieve the required high bandwidth efficiency, 4G and B4G systems adopt high-order quadrature amplitude modulation (QAM) signaling [21]. However, a high-order QAM constellation has high PAPR, which will drive the HPA at the transmitter into the nonlinear saturation region, resulting in a significant degradation in the system's achievable BER performance. Note that key design goals of B4G systems generally include both high bandwidth efficiency and high power-efficiency [22], [23]. The energy-efficiency aspect of communication favours high power-efficiency nonlinear HPAs, which however will cause a serious problem to high bandwidth-efficiency transmission technologies, such as OFDM and SC block transmission. Furthermore, recently, millimeter-wave (mmW) communications have been attracting extensive attentions, owing to the huge amount of unlicensed bandwidth

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

X. Hong (x.hong@reading.ac.uk) is with Department of Computer Science, School of Mathematical and Physical Sciences, University of Reading, Reading, RG6 6AY, UK.

E. Khalaf, A. Morfeq and N. Alotaibi (ekhalaf@kau.edu.sa, morfeq@kau.edu.sa, ndalotabi@kau.edu.sa) are with Electrical and Computer Engineering Department, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia.

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.

offered by mmW systems [24]–[26]. SC block transmission provides a viable technology for mmW based B4G systems [26]. However, for mmW communications, the design of HPA encounters severe nonlinearity [27], [28]. With the transmitter HPA exhibits severe nonlinear distortions, existing OFDM and SC-FDE or SC-FDE-DFE techniques can no longer work. How to effectively compensate for the nonlinear distortions of the HPA in the design of B4G wireless systems is vital for achieving the design goals of high bandwidth efficiency and high power efficiency.

A classical way of avoiding the nonlinearity of the transmitter HPA is output back-off (OBO). By lowering the average transmit power sufficiently away from the saturation power level of the HPA, hopefully the peak transmitted signal may still fall in the near-linear operating region of the HPA. For high PAPR modulation signals, however, OBO must be very severe to be effective. But such a large OBO will dramatically reduce the efficiency of the HPA, and more importantly it may not meet the required link power budget, especially for the user at cell edge. A more effective approach to compensate for the nonlinear distortions of HPA is to implement a digital predistorter at the transmitter, and various predistorter techniques have been developed [29]–[35]. Implementing the predistorter is attractive for the downlink, where the base station transmitter has the sufficient hardware and software capacities to accommodate the hardware and computational requirements for implementing digital predistorter. In the uplink, however, it is challenging to implement predistorter at transmitter, because it is extremely difficult for a pocket-size handset to absorb the required hardware and computational complexity. Therefore, the predistorter option is not viable for the SC block transmission based uplink system. Consequently, the base station receiver must deal with the nonlinear distortions of the transmitter HPA. 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 equalization for the SC block transmission based CV Hammerstein channel is a challenging task. A recent work [36] successfully developed a nonlinear SC-FDE (NFDE) solution based on a CV B-spline (BS) neural network approach [37], [38] for the Hammerstein channels with nonlinear transmit HPA.

Against the above background, in this paper, we propose a new nonlinear HDFE (NHDFE) design for Hammerstein channels, which considerably outperforms the NFDE design proposed in [36]. One reason that we favour a HDFE structure, rather than an IBDFE structure, is because we consider the uplink, whereby the BS has sufficient computational capacity to perform the matrix inversion required to determine the TD feedback equalizer's taps. On the other hand, by avoiding the iterative block detection of an IBDFE structure, we reduce the data detection delay. This NHDFE design requires the knowledge of the FD channel transfer function coefficients (FDCTFCs) of the dispersive channel as well as the nonlinear mapping and the inverse nonlinear mapping of the HPA, which are generally unknown. Motivated by the previous work [36], we adopt an effective CV BS neural network approach to

estimate the CV Hammerstein channel as well as to invert the HPA's nonlinear static mapping. More specifically, we use a CV BS neural network to model the nonlinear HPA, and we develop an efficient alternating least squares (ALS) scheme for estimating the parameters of the Hammerstein channel, including both the channel impulse response (CIR) coefficients and the coefficients of the CV BS model for the nonlinear HPA. This ALS algorithm is guaranteed to converge in just one iteration. It is worth emphasizing that our ALS algorithm is different to the two schemes derived in [36]. Similar to the work [36], we also adopt another CV BS neural network to model the inversion of the nonlinear HPA. Although the HPA's output at the transmitter is unobservable at the receiver for identifying this CV BS 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 BS model using the standard least squares (LS) algorithm [36]. To the best of our knowledge, this is the first practical and effective scheme for designing NHDFE for SC block transmission systems with nonlinear HPA at transmitter. The effectiveness of our proposed CV BS neural network based NHDFE solution is demonstrated in a simulation study, and the results obtained show that the proposed NHDFE solution outperforms the NFDE design of [36] by 4 dB in the signal-to-noise ratio (SNR) at the BER level of 10^{-4} .

The remainder of this paper is organized as follows. Section II provides the Hammerstein channel model for the SC block transmission system with nonlinear transmit HPA and summarises the proposed NHDFE design for the given Hammerstein channel. Section III is devoted to our proposed CV BS neural network approach for realizing the NHDFE solution. An extensive simulation study is provided in Section IV. Our conclusion remarks are drawn in Section V.

II. HAMMERSTEIN CHANNEL AND PROPOSED NHDFE SOLUTION

In our discussions, we represents a CV number $x \in \mathbb{C}$ either by $x = x_R + jx_I$, where $j = \sqrt{-1}$, while $x_R = \Re[x]$ and $x_I = \Im[x]$ denote the real and imaginary parts of x , or by $x = |x| \exp(j\angle x)$ with $|x|$ denoting the amplitude of x and $\angle x$ its phase.

A. System signal model

We adopt the unique word (UW) structure for SC block transmission as shown in Fig. 1. Each transmit block consists of N_d data symbols and the UW of length N_{uw} . N_{uw} must be chosen to be at least as long as the length of the CIR. For notation convenience, we express each transmission block as

$$\mathbf{x}[s] = [x_0[s] \ x_1[s] \ \cdots \ x_{N-1}[s]]^T, \quad (1)$$

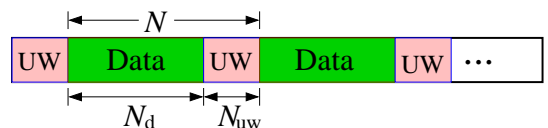


Fig. 1. UW-structure for SC block transmission.

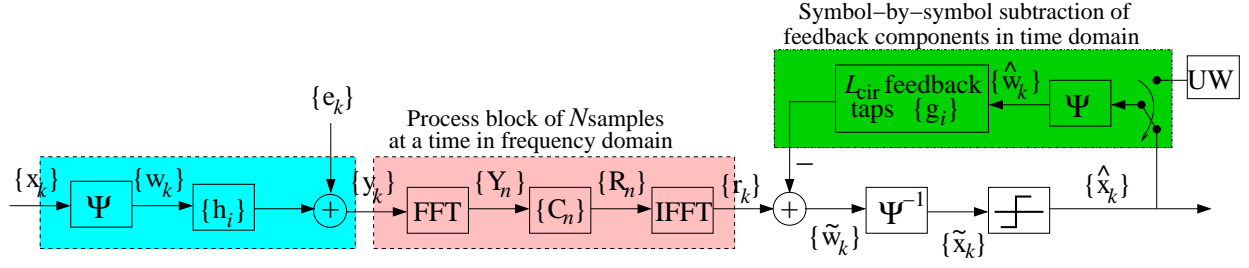


Fig. 2. Schematic diagram of the proposed NHDFFE.

where $[s]$ denotes the block index, $N = N_d + N_{uw}$ which defines the size of FFT for the SC block transmission system, and $()^T$ denotes the transpose operator. We consider M -QAM signaling and, therefore, $x_k[s]$ for $0 \leq k \leq N_d - 1$ take the values from the M -QAM constellation

$$\mathbb{X} = \{d(2l - \sqrt{M} - 1) + jd(2q - \sqrt{M} - 1), 1 \leq l, q \leq \sqrt{M}\}, \quad (2)$$

where $2d$ is the minimum distance between symbol points. For notational simplification, the block index $[s]$ is dropped in the sequel. The signal block \mathbf{x} is amplified by the transmitter HPA to yield the actually transmitted signal vector

$$\mathbf{w} = [w_0 \ w_1 \ \cdots \ w_{N-1}]^T, \quad (3)$$

where

$$w_k = \Psi(x_k), \ 0 \leq k \leq N - 1, \quad (4)$$

and $\Psi(\cdot)$ represents the CV static nonlinearity of the HPA.

The solid state power amplifier [6], [7] is typically employed. The nonlinearity $\Psi(\cdot)$ of this type of HPA is defined by the HPA's amplitude response $A(r)$ and phase response $\Upsilon(r)$, which are given respectively by

$$A(r) = \frac{g_a r}{\left(1 + \left(\frac{g_a r}{A_{\text{sat}}}\right)^{2\beta_a}\right)^{\frac{1}{2\beta_a}}}, \quad (5)$$

$$\Upsilon(r) = \frac{\alpha_\phi r^{q_1}}{1 + \left(\frac{r}{\beta_\phi}\right)^{q_2}} \text{ [degree]}, \quad (6)$$

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 [6], [7]. Given the input

$$x_k = |x_k| \exp(j\angle x_k) \quad (7)$$

to the HPA, therefore, the output of the HPA is given by

$$w_k = A(|x_k|) \exp(j(\angle x_k + \Upsilon(|x_k|))). \quad (8)$$

The operating status of the HPA may be specified by the OBO, which is defined as the ratio of the maximum output power P_{max} of the HPA to the average output power P_{aop} of the HPA output signal, given by

$$\text{OBO} = 10 \cdot \log_{10} \frac{P_{\text{max}}}{P_{\text{aop}}}. \quad (9)$$

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

The amplified signal block \mathbf{w} is transmitted through the channel whose CIR coefficient vector is expressed by

$$\mathbf{h} = [h_0 \ h_1 \ \cdots \ h_{L_{\text{cir}}}]^T. \quad (10)$$

Since the length of the UW is set to be no less than the CIR length, we have $L_{\text{cir}} \leq N_{\text{uw}}$. 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 $\Psi(\cdot)$, and the CIR coefficients are re-scaled as h_i/h_0 for $0 \leq i \leq L_{\text{cir}}$. At the receiver, the channel-impaired received signals y_k are given by

$$y_k = \sum_{i=0}^{L_{\text{cir}}} h_i w_{k-i} + e_k, \ 0 \leq k \leq N - 1, \quad (11)$$

where e_k is the channel AWGN with $E\{|e_k|^2\} = 2\sigma_e^2$, and $E\{\cdot\}$ denotes the expectation.

B. Receiver NHDFFE structure

The proposed NHDFFE structure is depicted in Fig. 2, which consists of a block processing of the N samples $\mathbf{y} = [y_0 \ y_1 \ \cdots \ y_{N-1}]^T$ at a time in the FD and a sample-by-sample subtraction of the L_{cir} feedback components in the TD.

More specifically, passing \mathbf{y} through the N -point FFT processor yields the FD received signal block $\mathbf{Y} = [Y_0 \ Y_1 \ \cdots \ Y_{N-1}]^T$ with the elements

$$Y_n = \sum_{k=0}^{N-1} y_k \exp\left(-j\frac{2\pi kn}{N}\right), \ 0 \leq n \leq N - 1. \quad (12)$$

It is well known that

$$Y_n = H_n W_n + \Xi_n, \ 0 \leq n \leq N - 1, \quad (13)$$

in which Ξ_n is the FD representation of the channel AWGN with $E\{|\Xi_n|^2\} = 2\sigma_e^2 N$, and $\mathbf{W} = [W_0 \ W_1 \ \cdots \ W_{N-1}]^T$ denotes the N -point FFT of \mathbf{w} with $E\{|W_n|^2\} = E\{|w_k|^2\}N = \sigma_w^2 N$, while the FDCTFC vector $\mathbf{H} = [H_0 \ H_1 \ \cdots \ H_{N-1}]^T$ is the N -point FFT of \mathbf{h} . Denote the coefficients of the FD feedforward equalizer by $\{C_n\}_{n=0}^{N-1}$. Then the block processing in the FD constitutes the FD feedforward equalization specified by

$$R_n = C_n Y_n, \ 0 \leq n \leq N - 1. \quad (14)$$

Converting $\{R_n\}_{n=0}^{N-1}$ by the N -point inverse FFT (IFFT) yields the TD signal block

$$r_k = \frac{1}{N} \sum_{n=0}^{N-1} R_n \exp\left(j\frac{2\pi nk}{N}\right), \ 0 \leq k \leq N - 1. \quad (15)$$

In the proposed NHDFFE, the feedback equalizer operates in the TD sample-by-sample for $0 \leq k \leq N_d - 1$. Let the coefficients of the TD feedback equalizer be $\{g_i\}_{i=1}^{L_{\text{cir}}}$. Then, the equalized signal at sample k is given by

$$\tilde{w}_k = r_k - \sum_{i=1}^{L_{\text{cir}}} g_i^* \hat{w}_{k-i}, \quad (16)$$

where \hat{w}_{k-i} denotes the estimate of w_{k-i} for $1 \leq i \leq L_{\text{cir}}$. Assume that the inverse mapping $\Psi^{-1}(\cdot)$ of the HPA is available at the receiver. Then the “soft” estimate of the TD symbol x_k can be generated according to

$$\tilde{x}_k = \Psi^{-1}(\tilde{w}_k), \quad 0 \leq k \leq N_d - 1. \quad (17)$$

Quantizing \tilde{x}_k yields the detected symbol \hat{x}_k . The detected symbol \hat{x}_k is then fed back and distorted by the nonlinear mapping $\Psi(\cdot)$ of the HPA, assuming that it is available at the receiver, to yield the input to the TD feedback equalizer

$$\hat{w}_k = \Psi(\hat{x}_k). \quad (18)$$

Note that if we denote the UW by $[u_1 \ u_2 \ \cdots \ u_{L_{\text{cir}}}]^T$, then

$$\hat{w}_{-l} = \Psi(u_{L_{\text{cir}}+1-l}), \quad 1 \leq l \leq L_{\text{cir}}. \quad (19)$$

Thus the TD feedback equalizer (16) is always causal.

Appendix provides the minimum mean square error (MMSE) solution for the coefficient vector $\mathbf{g} = [g_1 \ g_2 \ \cdots \ g_{L_{\text{cir}}}]^T$ of the TD feedback equalizer and the coefficients $\{C_n\}_{n=0}^{N-1}$ of the FD feedforward equalizer (also see [12]). Specifically, the MMSE solution of \mathbf{g} is given by

$$\begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_{L_{\text{cir}}} \end{bmatrix} = - \begin{bmatrix} v_0 & v_{-1} & \cdots & v_{-L_{\text{cir}}+1} \\ v_1 & v_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & v_{-1} \\ v_{L_{\text{cir}}-1} & \cdots & v_1 & v_0 \end{bmatrix}^{-1} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_{L_{\text{cir}}} \end{bmatrix}, \quad (20)$$

where

$$v_l = \sum_{n=0}^{N-1} \frac{\exp\left(-j\frac{2\pi nl}{N}\right)}{\frac{2\sigma_w^2}{\sigma_e^2} + |H_n|^2} \quad (21)$$

with $v_{-l} = v_l^*$. The MMSE solution of the FD feedforward equalizer's coefficients is given by

$$C_n = \frac{H_n^* G_n}{\frac{2\sigma_e^2}{\sigma_w^2} + |H_n|^2}, \quad 0 \leq n \leq N-1, \quad (22)$$

where

$$G_n = 1 + \sum_{i=1}^{L_{\text{cir}}} g_i^* \exp\left(-j\frac{2\pi in}{N}\right). \quad (23)$$

III. CV B-SPLINE NEURAL NETWORK APPROACH FOR IMPLEMENTING NHDFFE

The proposed NHDFFE requires the knowledge of the FD-CTFCs $\{H_n\}$ or the CIR coefficients $\{h_i\}$ and the CV nonlinearity $\Psi(\cdot)$ of the transmitter HPA as well as the inverse mapping $\Psi^{-1}(\cdot)$, which are unknown at the receiver. Similar to the work [36], we adopt the CV BS neural network approach to estimate $\{h_i\}$, $\Psi(\cdot)$ and $\Psi^{-1}(\cdot)$. More specifically, we use a CV BS neural network [37], [38] to represent the mapping $\hat{w} = \hat{\Psi}(x) : \mathbb{C} \rightarrow \mathbb{C}$ that is the estimate of the CV nonlinear function $\Psi(\cdot)$. We then develop an efficient ALS algorithm for jointly estimating $\{h_i\}$ and $\Psi(\cdot)$ by exploiting this BS modeling. Note that this ALS algorithm is different to the two algorithms derived in [36]. Furthermore, we utilise another CV BS neural network to model $\Psi^{-1}(\cdot)$. To estimate this inverting model requires the “input-output” training data $\{\mathbf{w}, \mathbf{x}\}$. Although \mathbf{w} is unobserved and is therefore unavailable, the pseudo training data $\hat{\mathbf{w}}$ can be constructed as a byproduct of the Hammerstein channel identification [36], and this allows us to estimate this inverting model based on the standard LS algorithm. It is worth pointing out that the pseudo training input $\hat{\mathbf{w}}$ is highly noisy, which may cause biased estimate for many models, such as the polynomial based inverse model. However, the BS model has the maximum robustness and optimal convexity property [39]–[41], and this ensures an excellent inverse modeling performance with noisy training input data.

A. Complex-valued B-spline neural network

According to the physics of real-life HPA, we point out that the nonlinearity $\Psi(\cdot)$ satisfies the following conditions.

- 1) $\Psi(\cdot)$ 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 $x = x_R + jx_I$ denotes the input to the HPA $\Psi(\cdot)$. Furthermore, the distributions of x_R and x_I are identical.

Based on property 2), we assume that $U_{\min} < x_s < U_{\max}$, where U_{\min} and U_{\max} are known finite real values, while x_s symbolically represents either x_R or x_I , namely, the subscript s is either R or I .

To model a nonlinearity in the univariate dimension of x_s , we employ a univariate BS model with a piecewise polynomial degree of P_o and N_s basis functions. Such a univariate BS neural network is parametrized by the knot sequence specified by $(N_s + P_o + 1)$ knot values, $\{U_0, U_1, \dots, U_{N_s+P_o}\}$, with

$$U_0 < U_1 < \cdots < U_{P_o-2} < U_{P_o-1} = U_{\min} < U_{P_o} < \cdots < U_{N_s} < U_{N_s+1} = U_{\max} < U_{N_s+2} < \cdots < U_{N_s+P_o}. \quad (24)$$

At each end, there are $P_o - 1$ so-called “external” knots that are outside the input region $[U_{\min}, U_{\max}]$ and one boundary knot. As a result, the number of “internal” knots is given by $N_s + 1 - P_o$. Given the set of predetermined knots (24), the set of N_s BS basis functions are formed using the famous De Boor recursion [42], yielding for $1 \leq l \leq N_s + P_o$,

$$B_l^{(s,0)}(x_s) = \begin{cases} 1, & \text{if } U_{l-1} \leq x_s < U_l, \\ 0, & \text{otherwise,} \end{cases} \quad (25)$$

as well as for $l = 1, \dots, N_s + P_o - p$ and $p = 1, \dots, P_o$,

$$B_l^{(s,p)}(x_s) = \frac{x_s - U_{l-1}}{U_{p+l-1} - U_{l-1}} B_l^{(s,p-1)}(x_s) + \frac{U_{p+l} - x_s}{U_{p+l} - U_l} B_{l+1}^{(s,p-1)}(x_s). \quad (26)$$

The polynomial degree $P_o = 3$ or 4 is often sufficient for modeling the nonlinearities that can be encountered in most practical applications. The number of BS basis functions N_s 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. Specifically, $N_s = 6$ to 10 is usually sufficient for accurately modeling in the finite and known interval $[U_{\min}, U_{\max}]$. The two boundary knots are obviously related to the known values U_{\min} and U_{\max} , respectively. The $N_s + 1 - P_o$ internal knots may be uniformly spaced in the interval $[U_{\min}, U_{\max}]$. Note that there exist no data for $x_s < U_{\min}$ and $x_s > U_{\max}$ in identification but it is desired that the BS model has certain extrapolating capability outside the interval $[U_{\min}, U_{\max}]$. The external knots may be set empirically to give the BS model a required extrapolation capability. In fact, since no data appears outside $[U_{\min}, U_{\max}]$, the precise choice of these external knots does not really matter, in terms of modeling accuracy. Also note that for QAM signals, the distribution of x_s is naturally symmetric and, therefore, the knot sequence should be chosen to be symmetric too.

Using the tensor product between the two sets of univariate BS basis functions [43], $B_l^{(R,P_o)}(x_R)$ for $1 \leq l \leq N_R$ and $B_m^{(I,P_o)}(x_I)$ for $1 \leq m \leq N_I$, a set of new B-spline basis functions $B_{l,m}^{(P_o)}(x)$ can be formed and used in the CV BS neural network, giving rise to

$$\begin{aligned} \hat{w} &= \hat{\Psi}(x) = \sum_{l=1}^{N_R} \sum_{m=1}^{N_I} B_{l,m}^{(P_o)}(x) \theta_{l,m} \\ &= \sum_{l=1}^{N_R} \sum_{m=1}^{N_I} B_l^{(R,P_o)}(x_R) B_m^{(I,P_o)}(x_I) \theta_{l,m}, \end{aligned} \quad (27)$$

where $\theta_{l,m} = \theta_{l,m_R} + j\theta_{l,m_I} \in \mathbb{C}$, $1 \leq l \leq N_R$ and $1 \leq m \leq N_I$, are the CV weights. Denote the weight vector of the BS model (27) as

$$\theta = [\theta_{1,1} \ \theta_{1,2} \ \dots \ \theta_{l,m} \ \dots \ \theta_{N_R,N_I}]^T \in \mathbb{C}^{N_B}, \quad (28)$$

where $N_B = N_R N_I$. The task of identifying the nonlinearity $\Psi(\cdot)$ is turned into one of estimating the parameter vector θ .

Remark 1: A conventional way of modeling a nonlinearity in the univariate dimension of x_s is to use a univariate polynomial model with the polynomial degree P_o , which has $P_o + 1$ basis functions given by

$$1, x_s, x_s^2, \dots, x_s^{P_o}. \quad (29)$$

Thus, the tensor-product polynomial model for modeling the CV Ψ has $(P_o + 1)^2$ basis functions for any given input x , and the complexity of the polynomial model is obviously on the order of $\mathcal{O}((P_o + 1)^2)$. The tensor-product BS model (27) on the other hand has N_s^2 basis functions, where $N_s = N_R = N_I$. Therefore, the complexity of this BS model would appear to

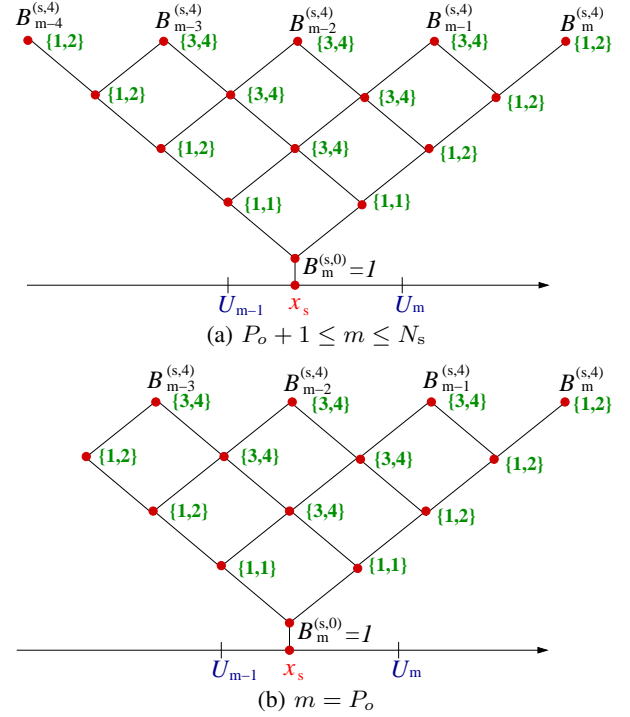


Fig. 3. Complexity of the B-spline model with $P_o = 4$ using the De Boor recursion, where $\{a, b\}$ beside a node indicates that it requires a additions and b multiplications to compute the basis function value at this node. Note that the case of $m = N_s + 1$ is similar to (b).

be on the order of $\mathcal{O}(N_s^2)$. Since typically $N_s > P_o$, this would be much higher than the complexity of the polynomial model. But this is in fact not the case. Given a value $x_s \in [U_{\min}, U_{\max}]$, there are only $P_o + 1$ basis functions with nonzero values at most as illustrated in Fig. 3, which shows the complexity of generating the BS 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_s employed. For the BS model with $P_o = 4$, the total computational requirements are 26 additions and 38 multiplications at most. Thus, in the tensor-product BS model (27), there are only $(P_o + 1)^2$ non-zero basis functions at most for any given input. The complexity of computing the BS model (27) is therefore also on the order of $\mathcal{O}((P_o + 1)^2)$. In particular, the upper bound complexity for $P_o = 4$ is listed in Table I, which includes generating the two sets of BS basis functions for real and imaginary parts, respectively, and the output of the tensor-product BS model (27).

Remark 2: The BS basis functions are optimally stable bases and they have the maximum numerical robustness [39]–[41]. This optimal robustness property of the BS model is due to the convexity of its model bases, i.e., they are all positive and sum to one. In fact, it is straightforward to illustrate this optimality of the BS model theoretically by comparing it with the polynomial model. Let us consider the univariate

TABLE I
UPPER BOUND COMPLEXITY OF TENSOR-PRODUCT B-SPLINE MODEL (27)
FOR $P_o = 4$.

| Computation | Multiplications | Additions |
|---------------------------------|-----------------|---------------|
| Two sets of 1-D basis functions | 2×38 | 2×26 |
| Output of (27) | 3×25 | 24 |
| Total | 151 | 76 |

nonlinearity that can be represented by the polynomial model of degree P_o exactly as

$$y_s = \sum_{i=0}^{P_o} a_i x_s^i, \quad (30)$$

as well as by the following BS model exactly as

$$y_s = \sum_{i=1}^{N_s} b_i B_i^{(s, P_o)}(x_s). \quad (31)$$

Because of the noisy training data, the estimated model coefficients are perturbed from their true values to $\hat{a}_i = a_i + \varepsilon_i$ for the polynomial model, and to $\hat{b}_i = b_i + \varepsilon_i$ for the BS model. Assume that all the estimation noises ε_i are bounded by $|\varepsilon_i| < \varepsilon_{\max}$. The upper bound of $|y_s - \hat{y}_s|$ for the BS model is given by

$$\begin{aligned} |y_s - \hat{y}_s| &= \left| \sum_{i=1}^{N_s} b_i B_i^{(s, P_o)}(x_s) - \sum_{i=1}^{N_s} \hat{b}_i B_i^{(s, P_o)}(x_s) \right| \\ &< \varepsilon_{\max} \left| \sum_{i=1}^{N_s} B_i^{(s, P_o)}(x_s) \right| = \varepsilon_{\max}. \end{aligned} \quad (32)$$

This upper bound of the BS model output perturbation only depends on the upper bound of the perturbation noise, and it does not depend on the input value x_s , the number of basis functions N_s or the polynomial degree P_o . Hence, the B-spline model enjoys the maximum numerical robustness. By contrast, the upper bound of $|y_s - \hat{y}_s|$ for the polynomial model is given by

$$|y_s - \hat{y}_s| = \left| \sum_{i=0}^{P_o} a_i x_s^i - \sum_{i=0}^{P_o} \hat{a}_i x_s^i \right| < \varepsilon_{\max} \left| \sum_{i=0}^{P_o} x_s^i \right|. \quad (33)$$

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_s and the polynomial degree P_o . The higher the polynomial degree P_o , the more serious the polynomial model may be perturbed.

Here, we use the simple example given in [44] to demonstrate the excellent numerical stability of the BS model over the polynomial model. Fig. 4(a) depicts a quadratic polynomial function $y_s = 0.001x_s^2 - 0.02x_s + 0.1$ defined over $x_s \in [0, 20]$ in solid curve. With the knot sequence $\{-5, -4, 0, 20, 24, 25\}$, this function is modeled as a quadratic BS model of $y_s = 0.14B_1^{(s, 2)}(x_s) - 0.10B_2^{(s, 2)}(x_s) + 0.14B_3^{(s, 2)}(x_s)$, which is plotted in Fig. 4(b) in solid curve. We draw three noises from a uniformly distributed random number (UDRN) in $[-0.0001, 0.0001]$, and add them to the three parameters in the two models, respectively. Fig. 4(a) and (b) show the ten sets of the perturbed functions in dashed curve generated by perturbing the two models' parameters. It can be seen from Fig. 4(a) that the polynomial model is seriously perturbed, but there is no noticeable change at all in Fig. 4(b) for the BS model. Then we draw three perturbation noises from a UDRN in $[-0.001, 0.001]$, and add them to the three parameters of the BS model. Again, the BS model is hardly affected, as can be observed from Fig. 4(c). Finally we draw three perturbation noises from a

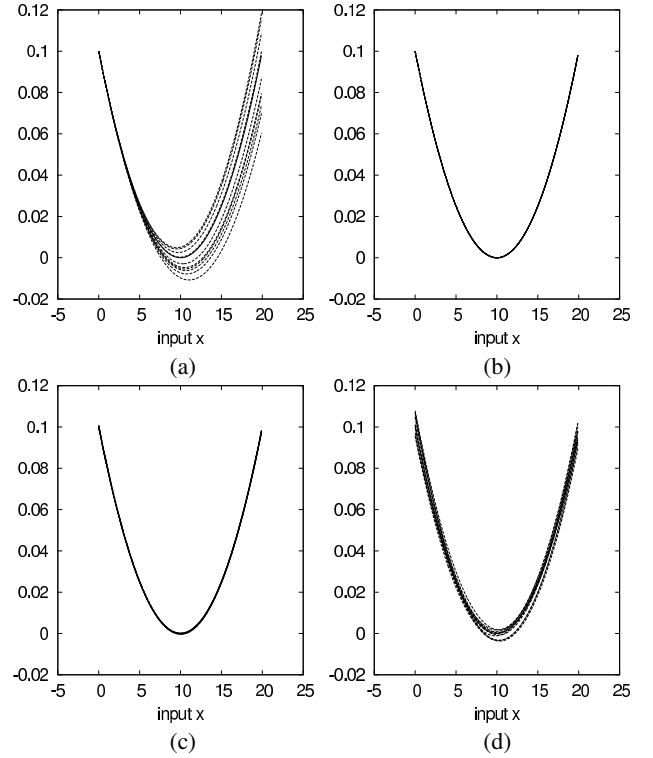


Fig. 4. (a) Polynomial model with UDRN perturbation noises drawn from $[-0.0001, 0.0001]$, (b) B-spline model with UDRN perturbation noises drawn from $[-0.0001, 0.0001]$, (c) B-spline model with UDRN perturbation noises drawn from $[-0.001, 0.001]$, and (d) B-spline model with UDRN perturbation noises drawn from $[-0.01, 0.01]$. Cited from [44].

UDRN in $[-0.01, 0.01]$ to add to the three BS parameters, and the results obtained are shown in Fig. 4(d). Comparing Fig. 4(d) with Fig. 4(a), we observe that, despite of the fact that the strength of the perturbation noise added to the BS model coefficients is 100 times larger than that added to the polynomial model coefficients, the BS model is still much less seriously perturbed than the polynomial model.

Moreover, the better performance of the BS based NFDE (BS-NFDE) over the polynomial based NFDE (P-NFDE) is shown in [36], while it is demonstrated in [45] that the BS based nonlinear OFDM receiver outperforms the polynomial based one.

B. Identification of the Hammerstein channel

The identification of the Hammerstein channel involves estimating the parameter vector θ of the CV BS neural network (27) that represents the HPA nonlinearity $\Psi(\cdot)$ as well as the CIR coefficient vector \mathbf{h} , based on a block of N training data, $\{x_k, y_k\}_{k=0}^{N-1}$. The output \hat{y}_k of the nonlinear model for 'predicting' the desired output y_k is given by

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

where $x_{k-i} = u_{L_{\text{cir}}+1-(k-i)}$ if $k < i$. The model (34) is linear with respect to \mathbf{h} when fixing θ , and it is linear with respect to θ given a fixed \mathbf{h} . According to [46], [47], the estimates of θ and \mathbf{h} are unbiased, irrespective the estimation algorithm, and the work [45] has derived two algorithms for estimating

\mathbf{h} and $\boldsymbol{\theta}$. By noting that $h_0 = 1$, we propose a different ALS procedure to estimate both $\boldsymbol{\theta}$ and \mathbf{h} , which we detail below.

Initialisation. Define the amalgamated parameter vector as

$$\boldsymbol{\omega} = [\boldsymbol{\theta}^T \ h_1 \boldsymbol{\theta}^T \ h_2 \boldsymbol{\theta}^T \ \cdots \ h_{L_{\text{cir}}} \boldsymbol{\theta}^T]^T \in \mathbb{C}^{(L_{\text{cir}}+1)N_B}. \quad (35)$$

Clearly the model (34) is linear with respect to $\boldsymbol{\omega}$. Further define the regression matrix $\mathbf{P} \in \mathbb{R}^{N \times (L_{\text{cir}}+1)N_B}$ as

$$\mathbf{P} = \begin{bmatrix} \boldsymbol{\phi}^T(0) & \boldsymbol{\phi}^T(-1) & \cdots & \boldsymbol{\phi}^T(-L_{\text{cir}}) \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\phi}^T(k) & \boldsymbol{\phi}^T(k-1) & \cdots & \boldsymbol{\phi}^T(k-L_{\text{cir}}) \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\phi}^T(N-1) & \boldsymbol{\phi}^T(N-2) & \cdots & \boldsymbol{\phi}^T(N-1-L_{\text{cir}}) \end{bmatrix}, \quad (36)$$

in which the BS basis function vector $\boldsymbol{\phi}(k) \in \mathbb{R}^{N_B}$ for the input x_k is given by

$$\boldsymbol{\phi}(k) = [\phi_{1,1}(k) \ \phi_{1,2}(k) \ \cdots \ \phi_{l,m}(k) \ \cdots \ \phi_{N_R,N_I}(k)]^T \quad (37)$$

with

$$\phi_{l,m}(k) = B_{l,m}^{(P_o)}(x_k), \quad 1 \leq l \leq N_R, 1 \leq m \leq N_I. \quad (38)$$

Assuming that $N \geq (L_{\text{cir}} + 1)N_B$, then the closed-form LS estimate of $\boldsymbol{\omega}$ is readily given by

$$\hat{\boldsymbol{\omega}} = (\mathbf{P}^T \mathbf{P} + \lambda \mathbf{I})^{-1} \mathbf{P}^T \mathbf{y}, \quad (39)$$

where λ is a very small regularization parameter, e.g., $\lambda = 10^{-5}$, and \mathbf{I} denotes the identity matrix of appropriate dimension. The regularization in (39) is necessary because the huge matrix \mathbf{P} is typically very ill-conditioned. Obviously, $\hat{\boldsymbol{\omega}}$ is a unique and unbiased estimate of $\boldsymbol{\omega}$ for $\lambda \rightarrow 0$. Therefore, the first N_B elements of $\hat{\boldsymbol{\omega}}$ provide a unique and unbiased LS estimate for the weight vector of the CV BS neural network $\boldsymbol{\theta}$, which will be denoted as $\hat{\boldsymbol{\theta}}^{(0)}$.

Alternating LS estimation. For $1 \leq \tau \leq \tau_{\text{max}}$, where τ_{max} is the maximum number of iterations, perform:

a) Given the fixed $\hat{\boldsymbol{\theta}}^{(\tau-1)}$, calculate the LS estimate $\hat{\mathbf{h}}^{(\tau)}$. Specifically, define the regression matrix $\mathbf{Q} \in \mathbb{C}^{N \times (L_{\text{cir}}+1)}$ as

$$\mathbf{Q} = \begin{bmatrix} \hat{w}_0 & \hat{w}_{-1} & \cdots & \hat{w}_{-L_{\text{cir}}} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{w}_k & \hat{w}_{k-1} & \cdots & \hat{w}_{k-L_{\text{cir}}} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{w}_{N-1} & \hat{w}_{N-2} & \cdots & \hat{w}_{N-1-L_{\text{cir}}} \end{bmatrix}, \quad (40)$$

in which

$$\hat{w}_k = \hat{\Psi}(x_k) = \sum_{l=1}^{N_R} \sum_{m=1}^{N_I} B_{l,m}^{(P_o)}(x_k) \hat{\theta}_{l,m}^{(\tau-1)}. \quad (41)$$

Then the closed-form LS estimate $\hat{\mathbf{h}}^{(\tau)}$ is readily given by

$$\hat{\mathbf{h}}^{(\tau)} = (\mathbf{Q}^H \mathbf{Q})^{-1} \mathbf{Q}^H \mathbf{y}, \quad (42)$$

$$\hat{h}_i^{(\tau)} = \hat{h}_i^{(\tau)} / \hat{h}_0^{(\tau)}, \quad 0 \leq i \leq L_{\text{cir}}, \quad (43)$$

where $(\cdot)^H$ denotes the conjugate transpose operator. Since $\hat{\boldsymbol{\theta}}^{(\tau-1)}$ is a unique and unbiased estimate of $\boldsymbol{\theta}$, the LS estimate $\hat{\mathbf{h}}^{(\tau)}$ is guaranteed to be a unique and unbiased estimate of \mathbf{h} .

b) Given the fixed $\hat{\mathbf{h}}^{(\tau)}$, calculate the LS estimate $\hat{\boldsymbol{\theta}}^{(\tau)}$. Specifically, introduce

$$\varphi_{l,m}(k) = \sum_{i=0}^{L_{\text{cir}}} \hat{h}_i^{(\tau)} B_{l,m}^{(P_o)}(x_{k-i}) \in \mathbb{C}. \quad (44)$$

Now introduce the regressor vector $\boldsymbol{\varphi}(k) \in \mathbb{C}^{N_B}$ given by

$$\boldsymbol{\varphi}(k) = [\varphi_{1,1}(k) \ \varphi_{1,2}(k) \ \cdots \ \varphi_{l,m}(k) \ \cdots \ \varphi_{N_R,N_I}(k)]^T, \quad (45)$$

and define the regression matrix

$$\mathbf{S} = [\boldsymbol{\varphi}(0) \ \boldsymbol{\varphi}(1) \ \cdots \ \boldsymbol{\varphi}(N-1)]^T \in \mathbb{C}^{N \times N_B}. \quad (46)$$

Then the closed-form LS estimate $\hat{\boldsymbol{\theta}}^{(\tau)}$ is readily given by

$$\hat{\boldsymbol{\theta}}^{(\tau)} = (\mathbf{S}^H \mathbf{S})^{-1} \mathbf{S}^H \mathbf{y}. \quad (47)$$

Since $\hat{\mathbf{h}}^{(\tau)}$ is a unique and unbiased estimate of \mathbf{h} , the LS estimate $\hat{\boldsymbol{\theta}}^{(\tau)}$ must be a unique and unbiased estimate of $\boldsymbol{\theta}$.

A few iterations, in fact no more than two, are sufficient for this estimation procedure to arrive at a highly accurate and joint unbiased estimate of \mathbf{h} and $\boldsymbol{\theta}$.

Remark 3: This ALS scheme is different from the two ALS algorithms given in [36]. In fact, it is clear that this ALS procedure guarantees to converge in one iteration. A few iterations, typically 2, are adopted simply to improve the estimation accuracy or to reduce the estimation variances. More specifically, since $\hat{\boldsymbol{\theta}}^{(0)}$ is a unique and unbiased estimate of $\boldsymbol{\theta}$, $\hat{\mathbf{h}}^{(1)}$ is guaranteed to be a unique and unbiased estimate of \mathbf{h} . On the other hand, the unique and unbiased LS estimate $\hat{\boldsymbol{\omega}}$ of the high-dimensional amalgamated parameter vector $\boldsymbol{\omega} \in \mathbb{C}^{(L_{\text{cir}}+1)N_B}$ may have relative high estimation variances, owing to the low ratio of the available training data N over the dimension $(L_{\text{cir}} + 1)N_B$. Consequently, The first N_B elements of $\hat{\boldsymbol{\omega}}$, i.e. $\hat{\boldsymbol{\theta}}^{(0)}$, may have a relatively poor estimation accuracy. With the fixed $\hat{\mathbf{h}}^{(1)}$, the LS estimate $\hat{\boldsymbol{\theta}}^{(1)}$ has lower estimation variance than $\hat{\boldsymbol{\theta}}^{(0)}$. In the second iteration, the estimation accuracy of $\hat{\mathbf{h}}^{(2)}$ and $\hat{\boldsymbol{\theta}}^{(2)}$ will be further enhanced, in comparison to the estimation accuracy of $\hat{\mathbf{h}}^{(1)}$ and $\hat{\boldsymbol{\theta}}^{(1)}$, and $\hat{\mathbf{h}}^{(2)}$ and $\hat{\boldsymbol{\theta}}^{(2)}$ provide highly accurate and joint unbiased estimate of \mathbf{h} and $\boldsymbol{\theta}$.

With the estimated CIR tap vector $\hat{\mathbf{h}}$, the estimated FDCT-FCs $\hat{\mathbf{H}}$ can readily be obtained by performing the N -point FFT on $\hat{\mathbf{h}}$. With the estimated HPA's nonlinearity $\hat{\Psi}(\cdot)$ and the CIR tap vector $\hat{\mathbf{h}}$, an estimate of the noise power $2\sigma_e^2$ is readily given by

$$2\hat{\sigma}_e^2 = \frac{1}{N} \sum_{k=0}^{N-1} |y_k - \hat{y}_k|^2, \quad (48)$$

while an estimate for the power σ_w^2 of the unobserved w_k is given by

$$\hat{\sigma}_w^2 = \frac{1}{N} \sum_{k=0}^{N-1} |\hat{w}_k|^2 = \frac{1}{N} \sum_{k=0}^N |\hat{\Psi}(x_k)|^2. \quad (49)$$

Both $2\sigma_e^2$ and σ_w^2 are needed in calculating the coefficients of the NHDFF.

C. Identification of the inverse nonlinear mapping of the HPA

Implementing the NHDFE of Fig. 2 also requires the inverse mapping of the HPA's CV nonlinearity defined by

$$x_k = \Psi^{-1}(w_k) = \Phi(w_k). \quad (50)$$

We utilise another CV BS neural network to model $\Psi^{-1}(\cdot)$. To estimate this inverting model $\Phi(\cdot)$ requires the "input-output" training data $\{w_k, x_k\}$ but w_k is unobserved and therefore unavailable for this modeling. As pointing out in [36], the pseudo training data $\{\hat{w}_k, x_k\}_{k=0}^{N-1}$ can be constructed as a byproduct of the Hammerstein channel identification. More specifically, given the estimated HPA's nonlinearity $\hat{\Psi}(\cdot)$, we can calculate $\hat{w}_k = \hat{\Psi}(x_k)$. This allows us to estimate the inverting model $\Phi(\cdot)$ based on the standard LS algorithm. However, the training input \hat{w}_k is highly noisy and this may introduce potentially serious bias in the estimate.

For notational convenience, define two knots sequences similar to (24) for w_R and w_I , respectively. Similar to (27), we construct the inverting BS neural network

$$\begin{aligned} \hat{x} &= \hat{\Phi}(w) = \sum_{l=1}^{N_R} \sum_{m=1}^{N_I} B_{l,m}^{(P_o)}(w) \alpha_{l,m} \\ &= \sum_{l=1}^{N_R} \sum_{m=1}^{N_I} B_l^{(R,P_o)}(w_R) B_m^{(I,P_o)}(w_I) \alpha_{l,m}, \end{aligned} \quad (51)$$

where $B_l^{(R,P_o)}(w_R)$ and $B_m^{(I,P_o)}(w_I)$ are respectively calculated based on (25) and (26), while

$$\alpha = [\alpha_{1,1} \ \alpha_{1,2} \ \cdots \ \alpha_{l,m} \ \cdots \ \alpha_{N_R,N_I}]^T \in \mathbb{C}^{N_B} \quad (52)$$

is the parameter vector of this inverting BS neural network. Here for notational simplicity, we assume that the same number of basis functions and polynomial degree are used for the two BS neural networks $\Psi(x_k)$ and $\Phi(w_k)$.

Over the pseudo training data set $\{\hat{w}_k, x_k\}_{k=0}^{N-1}$, the regression matrix $\tilde{\mathbf{B}} \in \mathbb{R}^{N \times N_B}$ can be formed as

$$\tilde{\mathbf{B}} = \begin{bmatrix} B_{1,1}^{(P_o)}(\hat{w}_0) & B_{1,2}^{(P_o)}(\hat{w}_0) & \cdots & B_{N_R,N_I}^{(P_o)}(\hat{w}_0) \\ B_{1,1}^{(P_o)}(\hat{w}_1) & B_{1,2}^{(P_o)}(\hat{w}_1) & \cdots & B_{N_R,N_I}^{(P_o)}(\hat{w}_1) \\ \vdots & \vdots & \ddots & \vdots \\ B_{1,1}^{(P_o)}(\hat{w}_{N-1}) & B_{1,2}^{(P_o)}(\hat{w}_{N-1}) & \cdots & B_{N_R,N_I}^{(P_o)}(\hat{w}_{N-1}) \end{bmatrix}. \quad (53)$$

and the closed-form LS solution for α is readily given by

$$\hat{\alpha} = (\tilde{\mathbf{B}}^T \tilde{\mathbf{B}})^{-1} \tilde{\mathbf{B}}^T \mathbf{x}. \quad (54)$$

Remark 4: A tensor-product polynomial model, having a polynomial degree of P_o in each dimension, can also be utilized to estimate $\Psi^{-1}(\cdot)$, based on the same LS identification procedure developed here. It is well-known that since the input \hat{w}_k is noisy, the LS estimate (54) is generally biased. Thus, the maximum robustness property of the BS model as illustrated in *Remark 2* is particularly helpful in minimizing this bias. Consequently, the BS based approach yields significantly better performance than the polynomial based approach.

IV. SIMULATION STUDY

A. Simulation system set up

A 64-QAM Hammerstein system was simulated in which the HPA employed at the transmitter was described by (5) and (6). The parameters of this nonlinear HPA were set to those of the NEC GaAs power amplifier adopted in the recent wireless standards [6], [7], which are

$$\begin{aligned} g_a &= 19, \beta_a = 0.81, A_{\text{sat}} = 1.4; \\ \alpha_\phi &= -48000, \beta_\phi = 0.123, q_1 = 3.8, q_2 = 3.7. \end{aligned} \quad (55)$$

The size of the transmitted data block was $N = 2048$. We assumed a quasi-static Rayleigh multipath channel with the CIR length $L_{\text{cir}} = 9$ and a power delay profile defined by

$$\mathbb{E}\{|h_l|\} = \begin{cases} \exp\left(-\frac{l}{\eta}\right), & 0 \leq l \leq 4, \\ \exp\left(-\frac{l-5}{\eta}\right), & 5 \leq l \leq L_{\text{cir}} = 9, \end{cases} \quad (56)$$

with the channel degradation factor $\eta = 3$. The CIR coefficients h_l for $0 \leq l \leq L_{\text{cir}}$ remained constant during the communication session. The effective system throughput is given by

$$\text{Effective throughput} = \frac{N_d}{N}, \quad (57)$$

where the effective data block length $N_d = N - L_{\text{cir}} = N - 9$. The larger the data block length N is, the more bandwidth efficient the system is.

We used a full data block with $N = 2048$ training samples in the joint estimation of the CV CIR coefficient vector \mathbf{h} and the CV parameter vector θ of the BS model for $\Psi(\cdot)$ as well as the estimation of the CV parameter vector α of the BS model for $\Psi^{-1}(\cdot)$. The piecewise quartic polynomial of $P_o = 4$ was chosen as the BS basis function, since $P_o = 4$ is sufficient for most practical applications. The number of BS basis functions was set to $N_R = N_I = 8$, as 8 basis functions is sufficient to partitioning or covering the input interval $[-d(\sqrt{M}-1), d(\sqrt{M}-1)]$. 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 SNR was defined as $\text{SNR} = E_w/N_o$, where $E_w = \mathbb{E}\{|w_k|^2\} = \sigma_w^2$ was the average transmitted signal power and $N_o = 2\sigma_e^2$ was the channel AWGN's power.

To demonstrate the optimal performance of the proposed BS neural network based NHDFE (BS-NHDFE), in the simulation, two tensor-product polynomial models, both having a polynomial degree of $P_o = 4$ in each dimension, were also be utilized to estimate the CV HPA's static nonlinearity $\Psi(\cdot)$ and its inversion $\Psi^{-1}(\cdot)$, respectively, based on the same identification procedures developed in Sections III-B and III-C, yielding a polynomial based NHDFE (P-NHDFE). The

TABLE II
EMPIRICALLY DETERMINED KNOT SEQUENCES.

| |
|---|
| Knot sequence for x_R and x_I |
| -10.0, -9.0, -1.0, -0.9 , -0.06, -0.04, 0.0, 0.04, 0.06, 0.9 , 1.0, 9.0, 10.0 |
| Knot sequence for w_R and w_I |
| -20.0, -18.0, -3.0, -1.4 , -0.8, -0.4, 0.0, 0.4, 0.8, 1.4 , 3.0, 18.0, 20.0 |

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

| Tap No. | True parameters | $E_w/N_o = 25$ dB | | $E_w/N_o = 30$ dB | |
|------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| | | B-spline | Polynomial | B-spline | Polynomial |
| OBO = 3 dB | | | | | |
| h_0 | 1 | 1 | 1 | 1 | 1 |
| h_1 | $-0.3732 - j0.6123$ | $-0.3732 - j0.6122$ | $-0.3741 - j0.6128$ | $-0.3732 - j0.6122$ | $-0.3741 - j0.6128$ |
| h_2 | $0.3584 + j0.3676$ | $0.3578 + j0.3682$ | $0.3564 + j0.3686$ | $0.3581 + j0.3679$ | $0.3567 + j0.3683$ |
| h_3 | $0.3052 + j0.2053$ | $0.3039 + j0.2058$ | $0.3051 + j0.2061$ | $0.3045 + j0.2056$ | $0.3057 + j0.2059$ |
| h_4 | $0.2300 + j0.1287$ | $0.2300 + j0.1289$ | $0.2308 + j0.1292$ | $0.2300 + j0.1288$ | $0.2308 + j0.1291$ |
| h_5 | $0.7071 + j0.7071$ | $0.7071 + j0.7083$ | $0.7074 + j0.7087$ | $0.7071 + j0.7077$ | $0.7073 + j0.7081$ |
| h_6 | $0.6123 - j0.3732$ | $0.6136 - j0.3723$ | $0.6137 - j0.3720$ | $0.6130 - j0.3727$ | $0.6131 - j0.3724$ |
| h_7 | $-0.3584 + j0.3676$ | $-0.3590 + j0.3663$ | $-0.3601 + j0.3662$ | $-0.3588 + j0.3668$ | $-0.3598 + j0.3668$ |
| h_8 | $-0.2053 - j0.3052$ | $-0.2053 - j0.3053$ | $-0.2052 - j0.3037$ | $-0.2053 - j0.3052$ | $-0.2052 - j0.3037$ |
| h_9 | $0.1287 - j0.2300$ | $0.1279 - j0.2299$ | $0.1274 - j0.2299$ | $0.1283 - j0.2299$ | $0.1278 - j0.2300$ |
| OBO = 5 dB | | | | | |
| h_0 | 1 | 1 | 1 | 1 | 1 |
| h_1 | $-0.3732 - j0.6123$ | $-0.3731 - j0.6121$ | $-0.3741 - j0.6119$ | $-0.3736 - j0.6124$ | $-0.3742 - j0.6120$ |
| h_2 | $0.3584 + j0.3676$ | $0.3579 + j0.3682$ | $0.3575 + j0.3694$ | $0.3584 + j0.3668$ | $0.3577 + j0.3691$ |
| h_3 | $0.3052 + j0.2053$ | $0.3039 + j0.2059$ | $0.3050 + j0.2055$ | $0.3061 + j0.2044$ | $0.3055 + j0.2052$ |
| h_4 | $0.2300 + j0.1287$ | $0.2299 + j0.1289$ | $0.2309 + j0.1286$ | $0.2304 + j0.1292$ | $0.2309 + j0.1285$ |
| h_5 | $0.7071 + j0.7071$ | $0.7072 + j0.7081$ | $0.7073 + j0.7079$ | $0.7077 + j0.7072$ | $0.7071 + j0.7074$ |
| h_6 | $0.6123 - j0.3732$ | $0.6136 - j0.3723$ | $0.6132 - j0.3728$ | $0.6111 - j0.3741$ | $0.6126 - j0.3731$ |
| h_7 | $-0.3584 + j0.3676$ | $-0.3591 + j0.3664$ | $-0.3597 + j0.3666$ | $-0.3589 + j0.3671$ | $-0.3594 + j0.3671$ |
| h_8 | $-0.2053 - j0.3052$ | $-0.2053 - j0.3053$ | $-0.2052 - j0.3041$ | $-0.2053 - j0.3054$ | $-0.2052 - j0.3041$ |
| h_9 | $0.1287 - j0.2300$ | $0.1280 - j0.2296$ | $0.1281 - j0.2294$ | $0.1285 - j0.2307$ | $0.1285 - j0.2296$ |

tensor-product polynomial model of degree $P_o = 4$ had 25 basis functions, which was comparable to the tensor-product BS model of at most 25 non-zero basis functions. Thus, the complexities of the two approaches were similar.

B. Estimation results

The ALS algorithm was used to identify this Hammerstein channel, specifically, to provide both the estimates of the CIR vector $\hat{\mathbf{h}}$ and the BS neural network's or polynomial

model's weight vector $\hat{\boldsymbol{\theta}}$. It was observed that $\tau_{\max} = 2$ iterations were sufficient for the algorithm to obtain the highly accurate estimation results as summarized in Table III as well as illustrated in Figs. 5 to 8, under four different operational conditions. It can be seen from Table III that the identification of CIR tap vector in the nonlinear Hammerstein channel was achieved with high precision under the adverse operational condition of OBO= 3 dB and $E_w/N_o = 25$ dB for both the BS based and polynomial based approaches. Note that

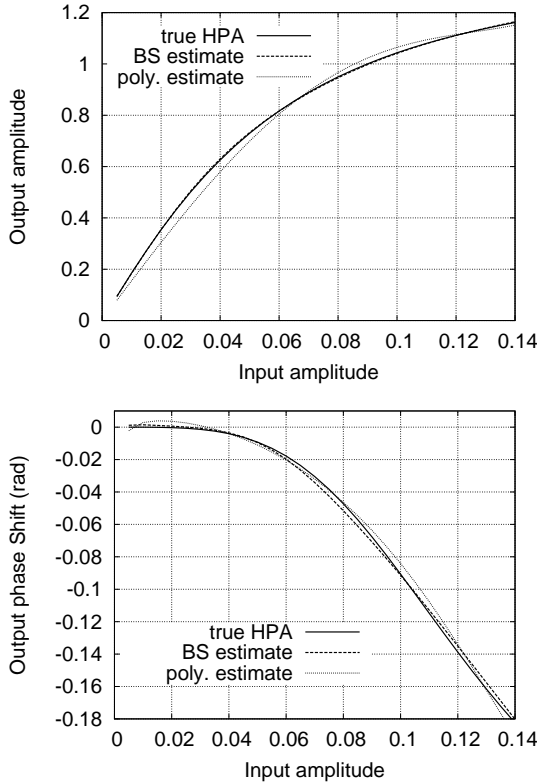


Fig. 5. Comparison of the HPA's static nonlinearity $\Psi(\cdot)$ and the B-spline estimated HPA nonlinearity as well as the polynomial estimated HPA nonlinearity under OBO= 3 dB and $E_w/N_o = 25$ dB.

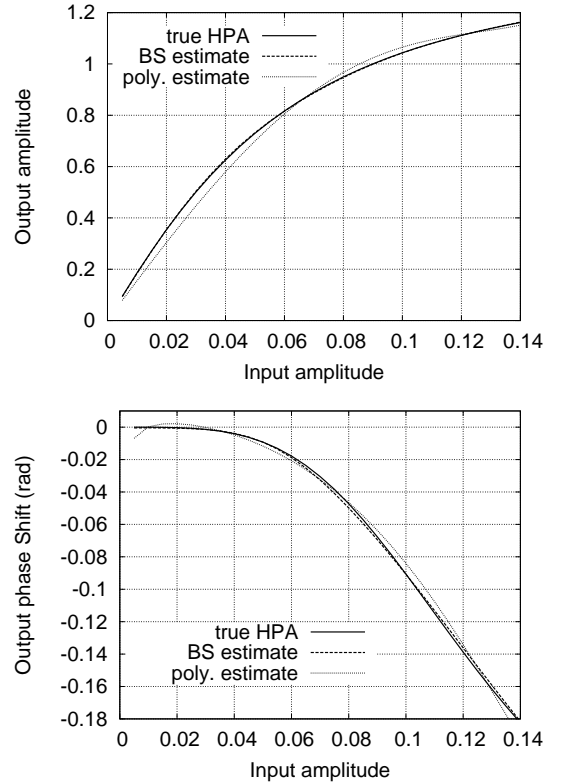


Fig. 6. Comparison of the HPA's static nonlinearity $\Psi(\cdot)$ and the B-spline estimated HPA nonlinearity as well as the polynomial estimated HPA nonlinearity under OBO= 3 dB and $E_w/N_o = 30$ dB.

under the HPA operational condition of OBO= 5 dB, the peak amplitude of $|x_k|$ was less than 0.09, while under the condition

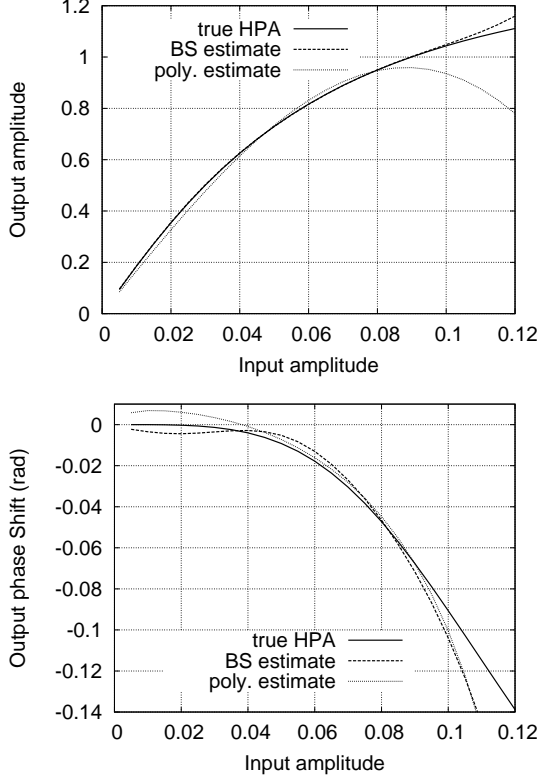


Fig. 7. Comparison of the HPA's static nonlinearity $\Psi(\cdot)$ and the B-spline estimated HPA nonlinearity as well as the polynomial estimated HPA nonlinearity under OBO= 5 dB and $E_w/N_o = 25$ dB.

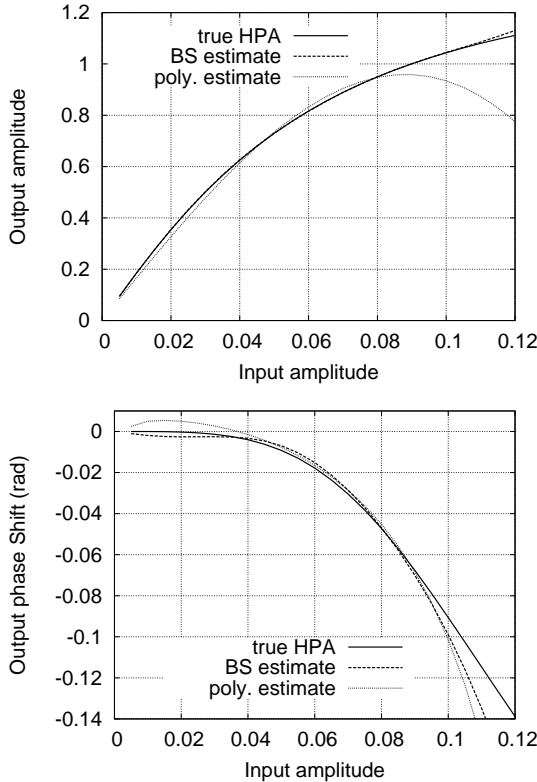


Fig. 8. Comparison of the HPA's static nonlinearity $\Psi(\cdot)$ and the B-spline estimated HPA nonlinearity as well as the polynomial estimated HPA nonlinearity under OBO= 5 dB and $E_w/N_o = 30$ dB.

of OBO= 3 dB, the peak amplitude of $|x_k|$ was less than 0.14. The results of Figs. 5 to 8 clearly demonstrate the capability of the proposed CV BS neural network to accurately model the HPA's static nonlinearity, within the HPA's input range. The results of Figs. 5 to 8 also show that the polynomial estimate of the HPA nonlinearity, although is sufficient accurate, is less accurate than the BS based estimate.

The combined responses of the HPA's true nonlinearity and its estimated inversion obtained by the proposed BS inverting scheme as well as the polynomial inverting scheme under the two operating conditions of OBO = 3 dB and SNR = 25 dB as well as OBO = 5 dB and SNR = 30 dB are depicted in Figs. 9 and 10, respectively. The results of Figs. 9 and 10 demonstrate the capability of the CV BS neural network to accurately model the inversion of the HPA's nonlinearity based only on the pseudo training data. More specifically, the results of Figs. 9 and 10 clearly show that the combined response of the true HPA's nonlinearity $\Psi(\cdot)$ and its BS inversion estimate $\hat{\Phi}(\cdot)$ satisfies

$$\hat{\Phi}(\Psi(x)) \approx x, \quad (58)$$

where x denotes the input to the HPA. That is, the magnitude of the combined response is $|\hat{\Phi}(\Psi(x))| \approx |x|$ and the phase shift of the combined response is approximately zero. In other words, $\hat{\Phi}(\cdot)$ is an accurate inversion of $\Psi(\cdot)$. By contrast, the polynomial inversion is much less accurate, as can be clearly seen from Figs. 9 and 10. Thus, the results of Figs. 9 and 10 confirm the analysis of *Remark 2* in Section III-A and *Remark 4* in Section III-C, demonstrating clearly the optimal robustness properties of the proposed BS based approach.

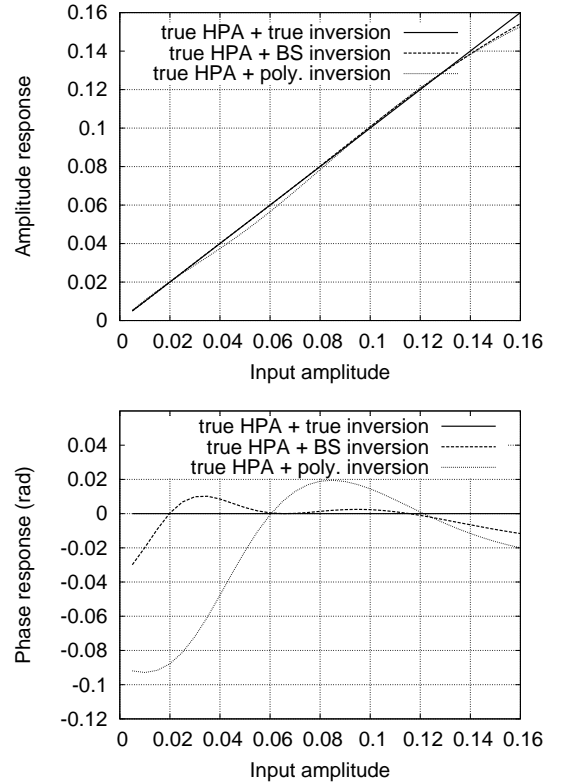


Fig. 9. Combined response of the true HPA and the estimated B-spline inversion as well as the estimated polynomial inversion under OBO= 3 dB and $E_w/N_o = 25$ dB.

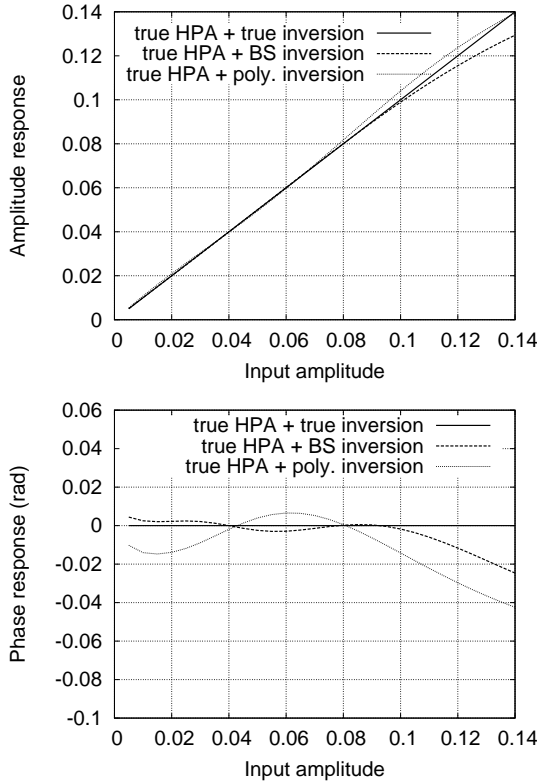


Fig. 10. Combined response of the true HPA and the estimated B-spline inversion as well as the estimated polynomial inversion under OBO= 5 dB and $E_w/N_o = 30$ dB.

C. Bit error rate performance evaluation

The achievable BER performance of the proposed NHDFE constructed based on the estimated CIR \hat{h} , the estimated BS model $\hat{\Psi}(\cdot)$ and the estimated BS inversion $\hat{\Psi}^{-1}(\cdot)$ are plotted in Fig. 11 under the two different operating conditions of the HPA. In the proposed NHDFE structure, if the TD DFE is removed, it reduces to the NFDE structure of [36], whereby the coefficient vector of the NFDE is set to the MMSE solution, namely, (22) with $G_n = 1$ for $0 \leq n \leq N - 1$. The BER performance achieved by this BS based NFDE based on the estimated \hat{h} , $\hat{\Psi}(\cdot)$ and $\hat{\Psi}^{-1}(\cdot)$ are also depicted in

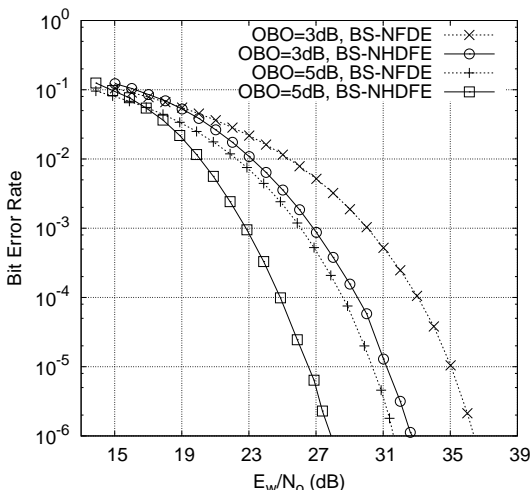


Fig. 11. Bit error rate performance comparison of the proposed B-spline based NHDFE and the B-spline based nonlinear FDE [36] under the two HPA operating conditions of OBO= 3 dB and OBO= 5 dB.

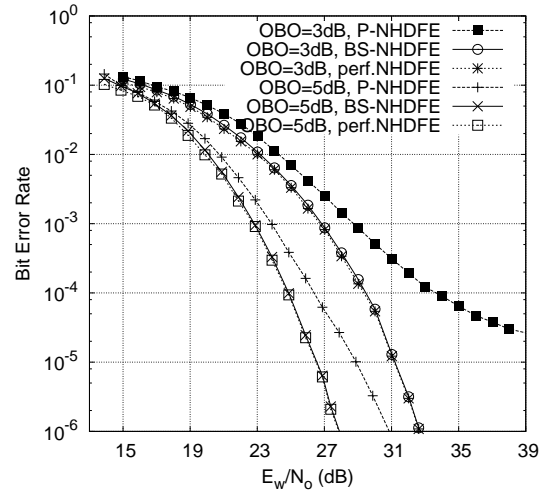


Fig. 12. Bit error rate performance comparison of the proposed B-spline based NHDFE and polynomial based NHDFE as well as the ideal NHDFE based on the perfect knowledge of \mathbf{h} , $\Psi(\cdot)$ and $\Psi^{-1}(\cdot)$ under the two HPA operating conditions of OBO= 3 dB and OBO= 5 dB.

Fig. 11 for comparison. The results of Fig. 11 confirm that the proposed BS based NHDFE outperforms the BS based NFDE significantly. In particular, at the BER level of 10^{-4} , the BS based NHDFE achieves an SNR gain of approximately 4 dB over the BS based NFDE of [36]. Also observe from Fig. 11 that under extremely poor SNR conditions, specifically, $\text{SNR} \leq 17$ dB for OBO = 5 dB and $\text{SNR} \leq 18$ dB for OBO = 3 dB, the BER of the BS-NFDE is smaller than that of the BS-NHDFE. This is due to the well known error propagation of the DFE structure at extremely poor SNR conditions.

The achievable BER performance of the proposed BS based NHDFE (BS-NHDFE) are compared with those of the polynomial based NHDFE (P-NHDFE) in Fig. 12 for the two HPA operating conditions of OBO = 3 dB and OBO = 5 dB, using the benchmark of the ideal NHDFE (perf.NHDFE) constructed based on the perfect knowledge of the CIR tap vector \mathbf{h} , the true HPA's nonlinearity $\Psi(\cdot)$ and its true inversion $\Psi^{-1}(\cdot)$. Note that we model or estimate the true HPA's inversion $\Psi^{-1}(\cdot)$ by a BS neural network based on the true input-output training data of $\{w_k, x_k\}_{k=0}^{N-1}$, which in reality are unavailable at the receiver. Not surprisingly, the performance of the P-NHDFE is considerably inferior to that of the proposed BS-NHDFE, particularly when the HPA is operating in the severe nonlinear region. This performance degradation of the P-NHDFE is mainly caused by serious bias in the estimate for $\Psi^{-1}(\cdot)$. Also unsurprisingly, the performance of the proposed BS-NHDFE is almost indistinguishable from that of the ideal NHDFE based on the perfect \mathbf{h} , $\Psi(\cdot)$ and $\Psi^{-1}(\cdot)$. This is simply owing to the ability of the proposed BS based estimation approach to very accurately identify \mathbf{h} , $\Psi(\cdot)$ and $\Psi^{-1}(\cdot)$, as clearly demonstrated in the results of Section IV-B.

V. CONCLUSIONS

A novel NHDFE design has been developed for the SC block transmission system over the Hammerstein channel, where the nonlinear distortion is caused by the high power amplifier at transmitter. We have adopted to use a CV B-spline

neural network for modeling the HPA's CV static nonlinearity as well as to use another CV B-spline neural network for modeling the inverse mapping of the HPA's nonlinearity. During training, the Hammerstein channel model parameters that include the CIR coefficients and the B-spline neural network weights can readily be estimated using a highly efficient ALS algorithm, while the weights of the B-spline inversion model can be identified using a standard LS algorithm based on the pseudo training data as a natural by-product of the Hammerstein channel model identification. Simulation results obtained have demonstrated that our proposed B-spline based NHDFFE significantly outperforms the previous state-of-the-art for nonlinear SC-FDE. The results have also been presented to show the superior performance of the proposed B-spline based NHDFFE design over the polynomial based NHDFFE design of comparable complexity.

In this study, we mainly focus on the achievable performance of the new NHDFFE design for nonlinear SC block transmission systems and, therefore, we assume that the channel is static during the whole communication session. Note that the nonlinear HPA model obtained during the training will remain valid for the whole session since the operating status of the transmitter HPA is determined by the link budget for the whole communication session. However, the channel may vary during the transmission and, therefore, the estimated CIR obtained during the training has to be updated. Our future work will investigate how to incorporate decision-directed channel estimation with the proposed NHDFFE design.

APPENDIX

The mean square error (MSE) is defined by

$$\text{MSE} = \mathbb{E}\{|\varepsilon_k|^2\} = \mathbb{E}\{|w_k - \tilde{w}_k|^2\}. \quad (59)$$

To determine the MMSE solution of $\{C_n\}_{n=0}^{N-1}$ and $\{g_i\}_{i=1}^{L_{\text{cir}}}$, we set $\hat{w}_{k-i} = w_{k-i}$ in the TD feedback equalizer and, therefore, we have

$$\tilde{w}_k = \frac{1}{N} \sum_{n=0}^{N-1} C_n (H_n W_n + \Xi_n) \exp\left(j \frac{2\pi n k}{N}\right) - \sum_{i=1}^{L_{\text{cir}}} g_i^* w_{k-i}. \quad (60)$$

By substituting (60) into (59) and after some simplification, we obtain

$$\mathbb{E}\{|\varepsilon_k|^2\} = \frac{2\sigma_e^2}{N} \sum_{n=0}^{N-1} |C_n|^2 + \frac{\sigma_w^2}{N} \sum_{n=0}^{N-1} |C_n H_n - G_n|^2, \quad (61)$$

where

$$G_n = 1 + \sum_{i=1}^{L_{\text{cir}}} g_i^* \exp\left(-j \frac{2\pi i n}{N}\right). \quad (62)$$

By setting $\frac{\partial \mathbb{E}\{|\varepsilon_k|^2\}}{\partial C_n} = 0$, we obtain the MMSE solution for $\{C_n\}_{n=0}^{N-1}$ as

$$C_n = \frac{H_n^* G_n}{\frac{2\sigma_e^2}{N} + |H_n|^2}, \quad 0 \leq n \leq N-1. \quad (63)$$

Substituting the MMSE solution (63) into (59) yields the corresponding MSE value

$$\begin{aligned} \mathbb{E}\{|\varepsilon_k|^2\} &= \frac{2\sigma_e^2}{N} \sum_{n=0}^{N-1} \frac{|G_n|^2}{\frac{2\sigma_e^2}{N} + |H_n|^2} \\ &= \frac{2\sigma_e^2}{N} \sum_{n=0}^{N-1} \frac{\left|1 + \sum_{i=1}^{L_{\text{cir}}} g_i^* \exp\left(-j \frac{2\pi i n}{N}\right)\right|^2}{\frac{2\sigma_e^2}{N} + |H_n|^2}. \end{aligned} \quad (64)$$

Setting $\frac{\partial \mathbb{E}\{|\varepsilon_k|^2\}}{\partial g_l} = 0$ yields the set of equations

$$\sum_{i=1}^{L_{\text{cir}}} g_i \sum_{n=0}^{N-1} \frac{\exp\left(-j \frac{2\pi n(l-i)}{N}\right)}{\frac{2\sigma_e^2}{N} + |H_n|^2} = - \sum_{n=0}^{N-1} \frac{\exp\left(-j \frac{2\pi n l}{N}\right)}{\frac{2\sigma_e^2}{N} + |H_n|^2} \quad (65)$$

for $1 \leq l \leq L_{\text{cir}}$. By defining

$$v_l = \sum_{n=0}^{N-1} \frac{\exp\left(-j \frac{2\pi n l}{N}\right)}{\frac{2\sigma_e^2}{N} + |H_n|^2} \quad (66)$$

with the property $v_{-l} = v_l^*$, we obtain

$$\begin{bmatrix} v_0 & v_{-1} & \cdots & v_{-L_{\text{cir}}+1} \\ v_1 & v_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & v_{-1} \\ v_{L_{\text{cir}}-1} & \cdots & v_1 & v_0 \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_{L_{\text{cir}}} \end{bmatrix} = - \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_{L_{\text{cir}}} \end{bmatrix}. \quad (67)$$

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] N. Benvenuto, R. Dinis, F. Falconer, and S. Tomasin, "Single carrier modulation with nonlinear frequency domain equalization: An idea whose time has come – again," *Proc. IEEE*, vol. 98, no. 1, pp. 69–96, Jan. 2010.

- [11] N. Benvenuto and S. Tomasin, "On the comparison between OFDM and single carrier modulation with a DFE using a frequency-domain feedforward filter," *IEEE Trans. Communications*, vol. 50, no. 6, pp. 947–955, June 2002.
- [12] D. Falconer, S. L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, *White Paper: Frequency Domain Equalization for Single-Carrier Broadband Wireless systems*, 2002. available: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.22.1200>
- [13] H. Witschnig, M. Kemptner, R. Weigel, and A. Springer, "Decision feedback equalization for a single carrier system with frequency domain equalization – An overall system approach," in *Proc. 1st Int. Symp. Wireless communication Systems* (Mauritius), Sept. 20–22, 2004, pp. 26–30.
- [14] F. Pancaldi and G. M. Vitetta, "Block channel equalization in the frequency domain," *IEEE Trans. Communications*, vol. 53, no. 3, pp. 463–471, Mar. 2005.
- [15] N. Benvenuto and S. Tomasin, "Iterative design and detection of a DFE in the frequency domain," *IEEE Trans. Communications*, vol. 53, no. 11, pp. 1867–1875, Nov. 2005.
- [16] F. Sainte-Agathe and H. Sari, "Single-carrier transmission with iterative frequency-domain decision-feedback equalization," in *Proc. EUSIPCO 2005* (Antalya, Turkey), Sept. 4–8, 2006, pp. 1–4.
- [17] F. Sainte-Agathe and H. Sari, "New results in iterative frequency-domain decision-feedback equalization," in *Proc. EUSIPCO 2006* (Florence, Italy), Sept. 4–8, 2006, pp. 1–4.
- [18] T. J. S. Khanzada, A. R. Ali, A. Q. K. Rajput, and A. S. Omar, "A design and chronological survey of decision feedback equalizer for single carrier transmission compared with OFDM," in: D. M. A. Hussain, A. Q. K. Rajput, B. S. Chowdhry, and Q. Gee (Eds.), *Wireless Networks, Information Processing and Systems*, Springer, 2008, pp. 378–390.
- [19] C. Zhang, Z. Wang, Z. Yang, J. Wang, and J. Song, "Frequency domain decision feedback equalization for uplink SC-FDMA," *IEEE Trans. Broadcasting*, vol. 56, no. 2, pp. 253–257, June 2010.
- [20] C. Zhang, Z. Wang, C. Pan, S. Chen, and L. Hanzo, "Low-complexity iterative frequency domain decision feedback equalization," *IEEE Trans. Vehicular Technology*, vol. 60, no. 3, pp. 1295–1301, March 2011.
- [21] 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.
- [22] S. Yunas, M. Valkama, and J. Niemelä, "Spectral and energy efficiency of ultra-dense networks under different deployment strategies," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 90–100, Jan. 2015.
- [23] I. Chih-Lin, C. Rowell, S. Han, Z. Xu, G. Li, Z. Pan, "Toward green and soft: A 5G perspective," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 66–73, Feb. 2014.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] 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.
- [32] 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.
- [33] 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.
- [34] 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.
- [35] 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.
- [36] X. Hong, S. Chen, C. J. Harris, and E. Khalaf, "Single-carrier frequency domain equalization for Hammerstein communication systems using complex-valued neural networks," *IEEE Trans. Signal Processing*, vol. 62, no. 17, pp. 4467–4478, Sept. 2014.
- [37] 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.
- [38] 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.
- [39] J. M. Pena, "B-spline and optimal stability," *Mathematics of Computation*, vol. 66, no. 220, pp. 1555–1560, Oct. 1997.
- [40] T. Lyche and J. M. Pena, "Optimally stable multivariate bases," *Advances in Computational Mathematics*, vol. 20, nos. 1–3, pp. 149–159, Jan. 2004.
- [41] 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.
- [42] C. De Boor, *A Practical Guide to Splines*. New York: Springer Verlag, 1978.
- [43] C. J. Harris, X. Hong, and Q. Gan, *Adaptive Modelling, Estimation and Fusion from Data: A Neurofuzzy Approach*. Berlin: Springer-Verlag, 2002.
- [44] S. Chen, X. Hong, E. Khalaf, A. Morfeq, and N. D. Alotaibi, "Adaptive B-spline neural network based nonlinear equalization for high-order QAM systems with nonlinear transmit high power amplifier," *Digital Signal Processing*, vol. 40, pp. 238–249, May 2015.
- [45] X. Hong, S. Chen, Y. Gong, and C. J. Harris, "Nonlinear equalization of Hammerstein OFDM systems," *IEEE Trans. Signal Processing*, vol. 62, no. 21, pp. 5629–5639, Nov. 2014.
- [46] 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.
- [47] C.-F. Wu, "Asymptotic theory of nonlinear least squares estimation," *The Annals of Statistics*, vol. 9, no. 3, pp. 501–513, 1981.



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 550 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). He was elected to a Fellow of the United Kingdom Royal Academy of Engineering in 2014.



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 Department of Computer

Science, School of Mathematical and Physical Sciences, 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 200 research papers, and coauthored a research book.

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



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.



Ali Morfeq received the B.S. degree in computer engineering from the King Abdulaziz University, Saudi Arabia, in 1982, the M.S. degree in computer engineering from the Oregon state University, Corvallis, USA in 1985, and the PhD degree in computer science from the University of Colorado, Boulder, USA in 1990.

Since 1990, he has been with the King Abdulaziz University, Saudi Arabia, where currently he is an assistant professor and the chair of the Electrical & Computer Engineering Department, Faculty of

Engineering. His research interests are in software engineering, and software systems for hospital applications.



Naif D. Alotaibi completed his B.Sc. in Computer Engineering from the King Abdulaziz University in Jeddah, Saudi Arabia, in 2000, and received his Master and PhD degrees in Information Technology from Queensland University of Technology in Brisbane, Australia, in 2006 and 2012, respectively.

In 2013, he joined the Electrical and Computer Engineering Department at the Faculty of Engineering, King Abdulaziz University, Jeddah, Saudi Arabia as an assistant professor. His research interests include IT infrastructure, computer network quality and network

work QoS, and quality measurement.



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 Research 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.