Complex-Valued B-Spline Neural Network and its Application to Iterative Frequency-Domain Decision Feedback Equalization for Hammerstein Communication Systems

Sheng Chen[1,2], Xia Hong[3], Emad Khalaf[2], Fuad E. Alsaadi[2], Chris J. Harris[1]

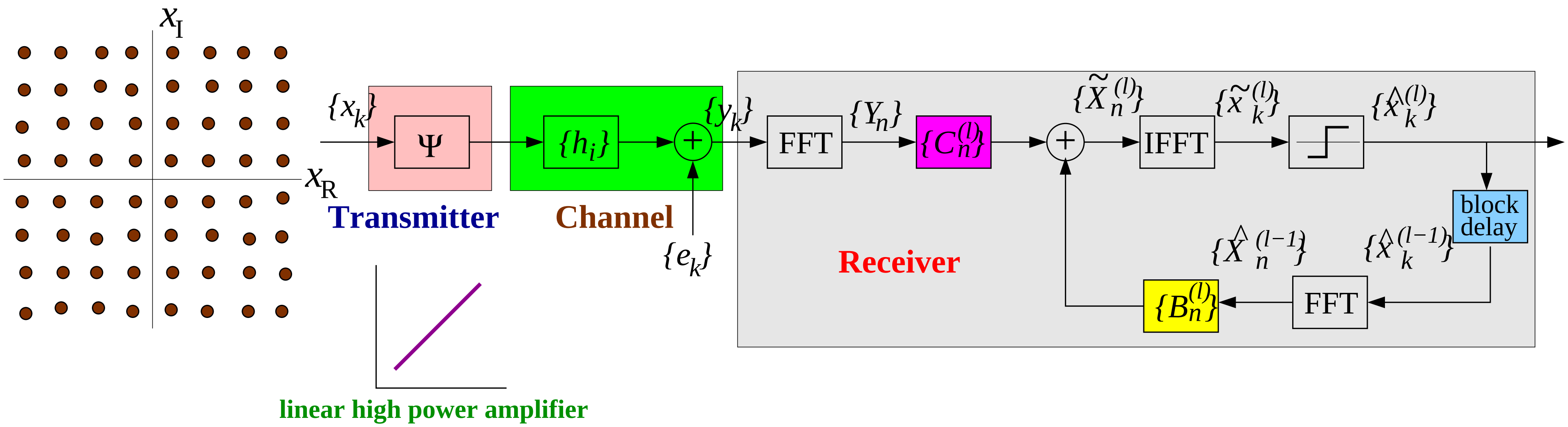
[1]School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK

[2]Electrical and Computer Engineering Department, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia

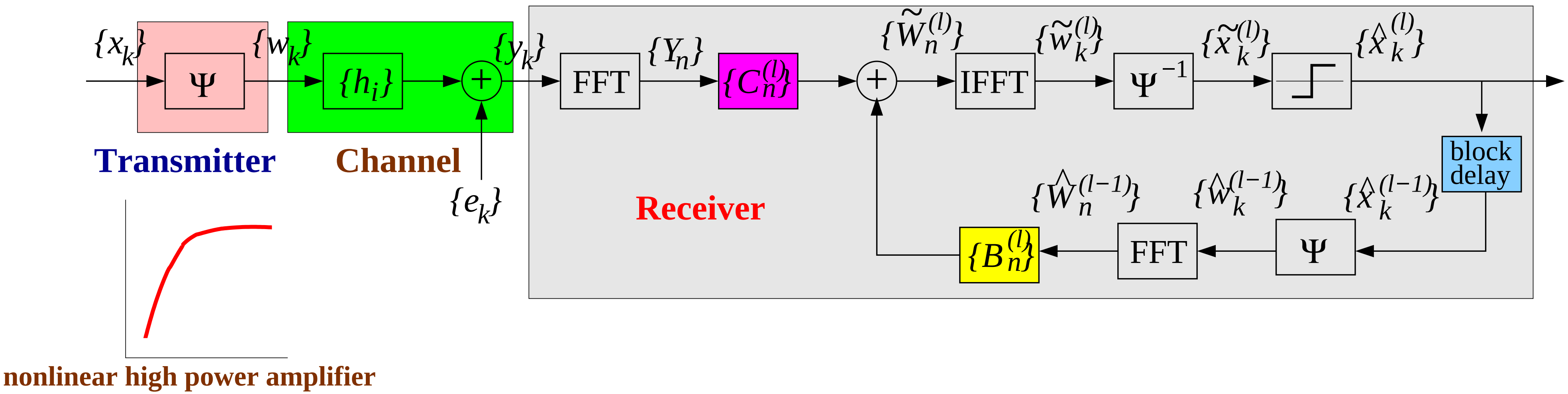
[3]Department of Computer Science, School of Mathematical and Physical Sciences, University of Reading, Reading, RG6 6AY, UK

Application

- Single-carrier black transmission with iterative frequency domain decision feedback equalization



- High-rate power efficient SC black transmission with iterative FD DFE would look like



How to Approach This Problem

- With nonlinear HPA at transmitter, channel is nonlinear
  - **Black box** learning estimates a nonlinear model $\widehat{y}_k = f(x_k, x_{k-1}, \cdots, x_{k-L})$
  - Very complicated, and little use to our application
- Nonlinear channel is **Hammerstein**: HPA static nonlinearity $w_k = \Psi(x_k)$ followed by linear dispersive channel $y_k = \sum_{i=0}^{L} h_i x_{k-i} + e_k$
  - **Grey box** approach utilizes this prior information to learn HPA's nonlinearity $\widehat{\Psi}$ and linear channel $\{\widehat{h}_i\}_{i=0}^{L}$, also to learn inversion $\widehat{\Psi}^{-1}$ but w_k is unavailable for this learning
- Standard tensor product of two sets of **polynomial** bases

$$P_l^{(\mathrm{s})}(x_{\mathrm{s}}) = x_{\mathrm{s}}^l,\ 0 \leq l \leq P_0,\ \ , \mathrm{s} = R \text{ or } I$$

  - To model Ψ

$$\widehat{w} = \widehat{\Psi}_P(x) = \sum_{l=0}^{P_0} \sum_{m=0}^{P_0} P_{l,m}(x)\theta_{l,m}^P = \sum_{l=0}^{P_0} \sum_{m=0}^{P_0} P_l^{(R)}(x_R) P_l^{(I)}(x_I)\theta_{l,m}^P$$

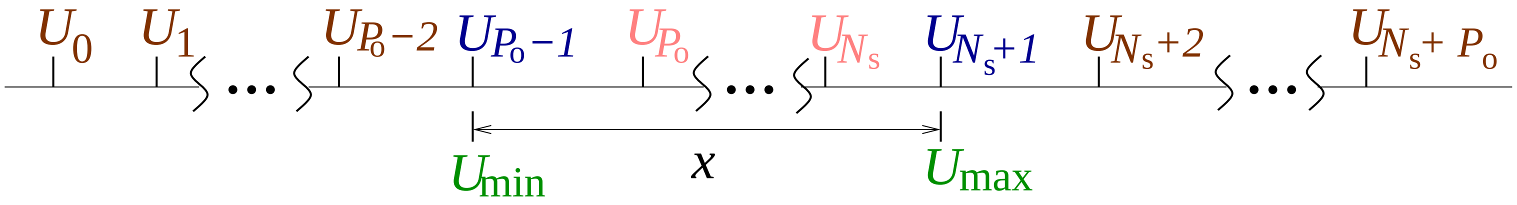
  - To model Ψ^{-1} with 'pseudo input' $\widehat{w} = \widehat{w}_R + \mathrm{j}\widehat{w}_I$

$$\widehat{x} = \widehat{\Psi}_P^{-1}(\widehat{w}) = \sum_{l=0}^{P_0} \sum_{m=0}^{P_0} P_l^{(R)}(\widehat{w}_R) P_l^{(I)}(\widehat{w}_I)\alpha_{l,m}^P$$

- Tensor product of two sets of univariate **B-spline** bases is far **better**

B-spline Basis Functions

- Knot sequence $\{U_0, U_1, \cdots, U_{N_\mathrm{s}+P_o}\}$



  - Input $U_{\min} \leq x_\mathrm{s} \leq U_{\max}$, P_o: polynomial degree, N_s: number of basis functions
  - $P_o - 1$ **external knots** and one **boundary knot** at each end, $N_\mathrm{s} + 1 - P_o$ **internal knots**
- **De Boor recursion**

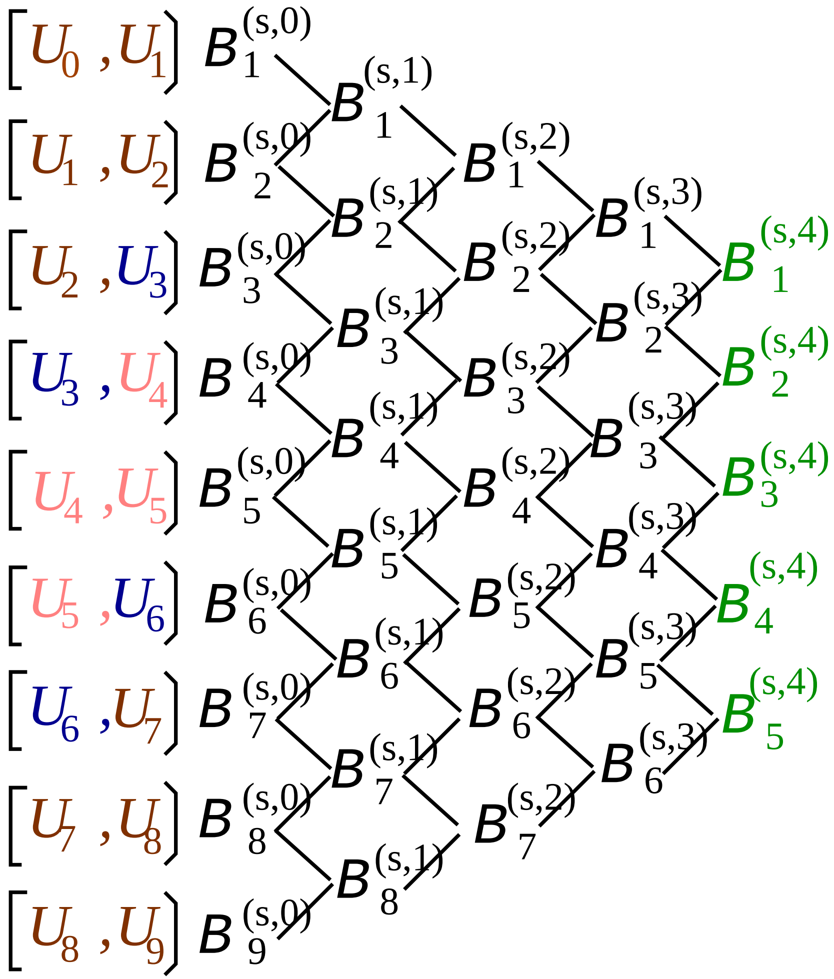
$$B_l^{(\mathrm{s},0)}(x_\mathrm{s}) = \begin{cases} 1, & \text{if } U_{l-1} \leq x_\mathrm{s} < U_l, \\ 0, & \text{otherwise,} \end{cases}$$

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

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

- **Polynomial** degree $P_o = 3$ or 4 sufficient, number of **basis functions** $N_\mathrm{s} = 6$ to 10 sufficient
  - Two boundary knots on $U_{\min}$ and $U_{\max}$, internal knots uniformly distributed in $[U_{\min},\ U_{\max}]$, external knots offer potential extrapolation capability

B-spline Model



Visualisation of De Boor recursion for $P_O = 4$ and $N_{\rm s} = 5$, where $U_{\rm min} = U_3$ and $U_{\rm max} = U_6$

1. Tensor product **B-spline** **modeling** of Ψ

$$\widehat{w} = \widehat{\Psi}_B(x) = \sum_{l=1}^{N_R} \sum_{m=1}^{N_I} B_{l,m}^{(P_0)}(x)\theta_{l,m}^B$$

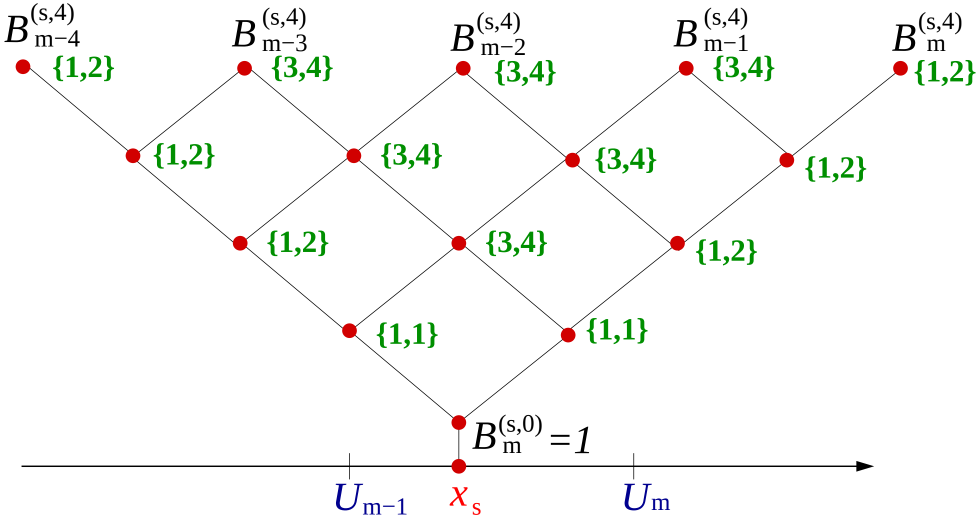
- $B_{l,m}^{(P_0)}(x) = B_l^{(R,P_0)}(x_R) B_m^{(I,P_0)}(x_I)$
- $N_R = N_I = N_{\rm s}$, $N_B = N_R N_I$
- $\boldsymbol{\theta}_B = \left[\theta_{1,1}^B \ \theta_{1,2}^B \cdots \theta_{l,m}^B \cdots \theta_{N_R,N_I}^B\right]^{\rm T} \in \mathbb{C}^{N_B}$
- Task is to estimate coefficients $\boldsymbol{\theta}_B$ given training data $\{x_k, y_k\}$

2. Tensor product **B-spline** **inverting** of Ψ

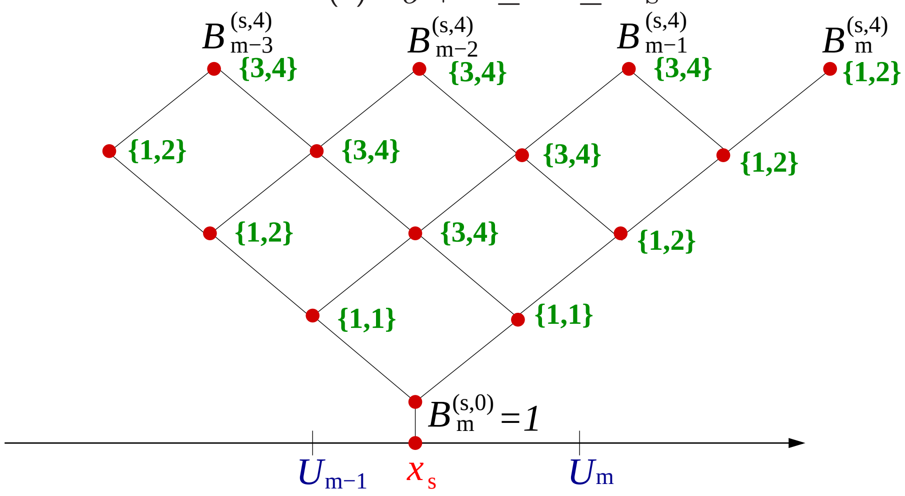
$$\widehat{x} = \widehat{\Psi}_B^{-1}(\widehat{w}) = \sum_{l=1}^{N_R} \sum_{m=1}^{N_I} B_{l,m}^{(P_0)}(\widehat{w})\alpha_{l,m}^B$$

- $\boldsymbol{\alpha}_B = \left[\alpha_{1,1}^B \ \alpha_{1,2}^B \cdots \alpha_{l,m}^B \cdots \alpha_{N_R,N_I}^B\right]^{\rm T} \in \mathbb{C}^{N_B}$
- Task is to estimate coefficients $\boldsymbol{\alpha}_B$, given **pseudo** training data $\{\widehat{w}_k, x_k\}$
- $\widehat{w}_k$ generated based on model identified in 1. as $\widehat{w}_k = \widehat{\Psi}_B(x_k)$

Complexity



(a) $P_O + 1 \leq m \leq N_S$



(b) $m = P_O$ or $m = N_S + 1$

Complexity of B-spline model with $P_O = 4$ using De Boor recursion

1. **Polynomial**: Complexity of $\widehat{\Psi}_P$ is obviously on order of $\text{O}\big((1 + P_0)^2\big)$
2. **B-spline**: Complexity of $\widehat{\Psi}_B$ on order of $\text{O}\big(N_s^2\big)$? As $N_s > 1 + P_0$, complexity of $\widehat{\Psi}_B$ much higher than complexity of $\widehat{\Psi}_P$?
   - Given x_s, only $P_o + 1$ basis functions with nonzero values at most
   - Complexity $\widehat{\Psi}_B$ is on order of $\text{O}\big((1+P_0)^2\big)$

Complexity of polynomial model for $P_O = 4$

Computation	Multiplications	Additions
Two sets of 1-D basis functions	2×4	0
Tensor product output	3×25	2×24
Total	83	48

Upper bound complexity of B-spline model for $P_O = 4$

Computation	Multiplications	Additions
Two sets of 1-D basis functions	2×38	2×26
Tensor product output	3×25	2×24
Total	151	100

Lower bound complexity of B-spline model for $P_O = 4$

Computation	Multiplications	Additions
Two sets of 1-D basis functions	2×36	2×25
Tensor product output	3×16	2×15
Total	120	80

Optimal Property

1. **Convexity** of B-spline model bases: they are all positive and sum to one
2. B-spline have **best approximation capability**, as the basis function is complete
3. B-spline has **maximum numerical stability/robustness** compared with other polynomial forms

- True system is represented exactly by the polynomial model

$$y_{\mathrm{s}} = \sum_{i=0}^{P_O} a_i x_{\mathrm{s}}^i$$

Same system can also be represented exactly by the B-spline model

$$y_{\mathrm{s}} = \sum_{i=1}^{N_{\mathrm{S}}} b_i B_i^{(\mathrm{s},P_O)}(x_{\mathrm{s}})$$

- As identification data are noisy, the estimated model coefficients are perturbed from their true values to

$$\widehat{a}_i = a_i + \varepsilon_i$$

for the polynomial model, and for the B-spline model to

$$\widehat{b}_i = b_i + \varepsilon_i$$

Optimal Property (2)

- Assume that all estimation noises ε_i are bounded, $|\varepsilon_i| < \varepsilon_{\max}$

- The **upper bound** of $|y_{\mathrm{s}} - \widehat{y}_{\mathrm{s}}|$ for the **B-spline** model

$$|y_{\mathrm{s}} - \widehat{y}_{\mathrm{s}}| = \left| \sum_{i=1}^{N_{\mathrm{s}}} b_i B_i^{(\mathrm{s},P_o)}(x_{\mathrm{s}}) - \sum_{i=1}^{N_{\mathrm{s}}} \widehat{b}_i B_i^{(\mathrm{s},P_o)}(x_{\mathrm{s}}) \right| < \varepsilon_{\max} \left| \sum_{i=1}^{N_{\mathrm{s}}} B_i^{(\mathrm{s},P_o)}(x_{\mathrm{s}}) \right| = \varepsilon_{\max}$$

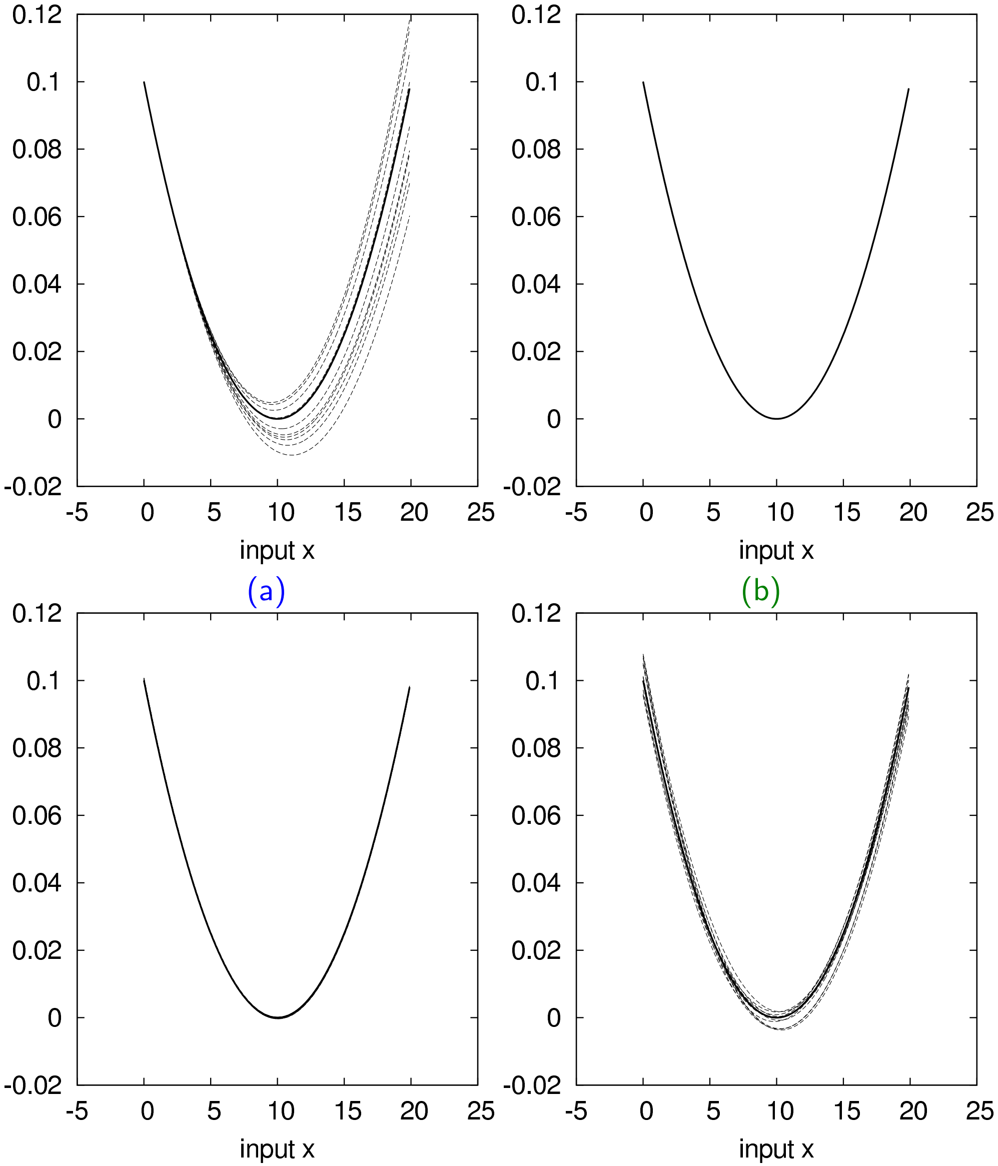
  – Only depends on $\varepsilon_{\max}$, thus has maximum robustness

- The **upper bound** of $|y_{\mathrm{s}} - \widehat{y}_{\mathrm{s}}|$ for the **polynomial** model

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

  – Depends on $\varepsilon_{\max}$, input value x_{s} and polynomial degree P_o

Numerical Stability Example



(a) (b) (c) (d)

- Quadratic polynomial

$$y = 0.001x^2 - 0.02x + 0.1$$

defined over $x \in [0,\ 20]$ in solid line

- Quadratic B-spline

$$y = 0.14B_1^{(2)}(x) - 0.10B_2^{(2)}(x) + 0.14B_3^{(2)}(x)$$

with knot sequence $\{-5, -4, 0, 20, 24, 25\}$ in solid line

- 10 set of perturbed functions in dashed line

(a) Polynomial, ε_i uniformly randomly drawn in $[-0.0001,\ 0.0001]$
(b) B-spline, ε_i uniformly randomly drawn from $[-0.0001,\ 0.0001]$
(c) B-spline, ε_i uniformly randomly drawn from $[-0.001,\ 0.001]$
(d) B-spline, ε_i uniformly randomly drawn from $[-0.01,\ 0.01]$

- Despite $\varepsilon_{\max}$ added to B-spline coefficients is 100 times larger than added to polynomial coefficients, B-spline model is much less seriously perturbed than polynomial model

Alternating Least Squares for Ψ and h

- $h_0 = 1$: $h_i = h_i/h_0$, $0 \leq i \leq L$, $\Psi = h_0 * \Psi$, and given training data $\{x_k, y_k\}_{k=1}^N$
- **Initialization**. For linear model in $\boldsymbol{\omega} = \left[\boldsymbol{\theta}^{\mathrm{T}}\ h_1\boldsymbol{\theta}^{\mathrm{T}}\ h_2\boldsymbol{\theta}^{\mathrm{T}} \cdots h_L\boldsymbol{\theta}^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{(L+1)N_B}$
  - Assume $N \geq (L+1)N_B$, we have closed-form unbiased **regularized** LS estimate $\widehat{\boldsymbol{\omega}}$
  - First N_B elements of $\widehat{\boldsymbol{\omega}}$ provides initial unbiased estimate for $\boldsymbol{\theta}$, denoted as $\widehat{\boldsymbol{\theta}}^{(0)}$
- **ALS**. For $1 \leq \tau \leq \tau_{\max}$, e.g., $\tau_{\max} = 2$, perform
  1. Given $\widehat{\boldsymbol{\theta}}^{(\tau-1)}$, for linear model in $\boldsymbol{h}$, we have closed-form unbiased LS estimate $\widehat{\boldsymbol{h}}^{(\tau)}$, and scale it according to $\widehat{h}_i^{(\tau)} = \widehat{h}_i^{(\tau)}/\widehat{h}_0^{(\tau)}$, $0 \leq i \leq L$
  2. Given $\widehat{\boldsymbol{h}}^{(\tau)}$, for linear model in $\boldsymbol{\theta}$, we have closed-form unbiased LS estimate $\widehat{\boldsymbol{\theta}}^{(\tau)}$

Remark: ALS guarantees converging to unbiased estimate of $\boldsymbol{h}$ and $\boldsymbol{\theta}$

Least Squares for Ψ^{-1}

- To estimate Ψ^{-1}, we need input w_k and desired output x_k, but w_k unavailable
- With estimated $\widehat{\Psi}$, generate **pseudo** training data $\{\widehat{w}_k, x_k\}_{k=1}^N$ with $\widehat{w}_k = \widehat{\Psi}(x_k)$
- For linear model in $\boldsymbol{\alpha}$, we have closed-form LS estimate $\widehat{\boldsymbol{\alpha}}$

Remark: Pseudo training input $\widehat{w}_k$ is highly noisy, which will serious affect polynomial inverse model but not B-spline inverse model (B-spline has maximum robustness property)

Performance Evaluation

- Simulation system set up: block size $N = 2048$, 64-QAM
  - HPA $w = \Psi(x)$ **amplitude** and **phase** response

$$A(r) = \frac{g_a r}{\left(1 + \left(\frac{g_a r}{A_{\text{sat}}}\right)^{2\beta_a}\right)^{\frac{1}{2\beta_a}}}, \quad \Upsilon(r) = \frac{\alpha_\phi r^{q_1}}{1 + \left(\frac{r}{\beta_\phi}\right)^{q_2}} \text{ [degree]}, \quad r = |x|$$

  Maximum and average powers of w: $P_{\max}$ and P_{aop}, then operating status specified by **output back off**

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

  - 10 tap channel $L = 9$
  - Signal to noise ratio: SNR $= E_x/N_o$, with E_x average symbol energy, and N_o noise power
- **Polynomial** model: polynomial degree $P_o = 4$
- **B-spline** model: $P_o = 4$, $N_R = N_I = 8$

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

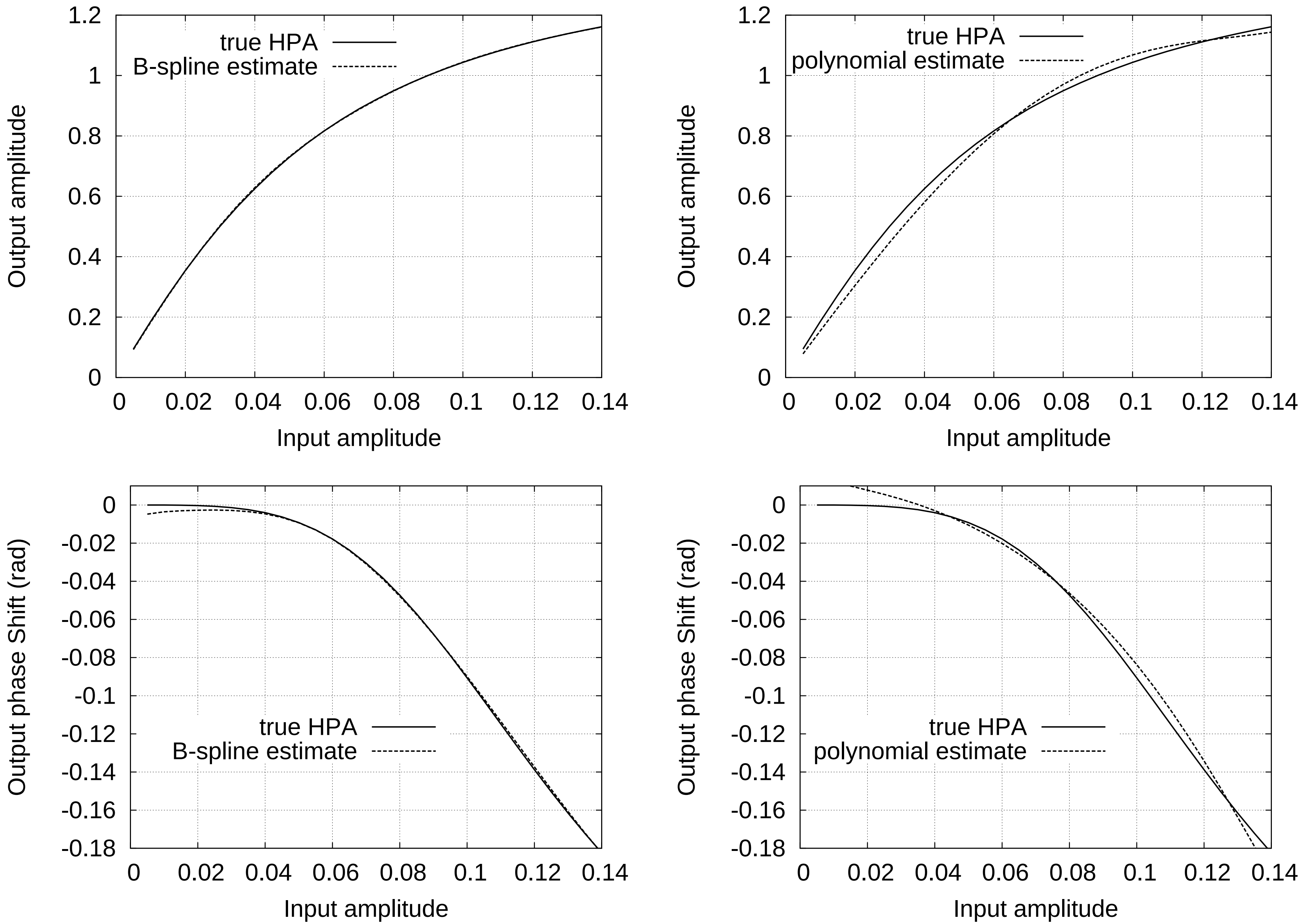
Dispersive Channel Identification

- Experiments were repeated 100 independent runs
- Channel taps were identified with very **high accuracy** for both **B-spline** and **polynomial** based approaches
- For highly nonlinear (OBO $= 3\,$dB) and highly noisy (SNR $= 5\,$dB):

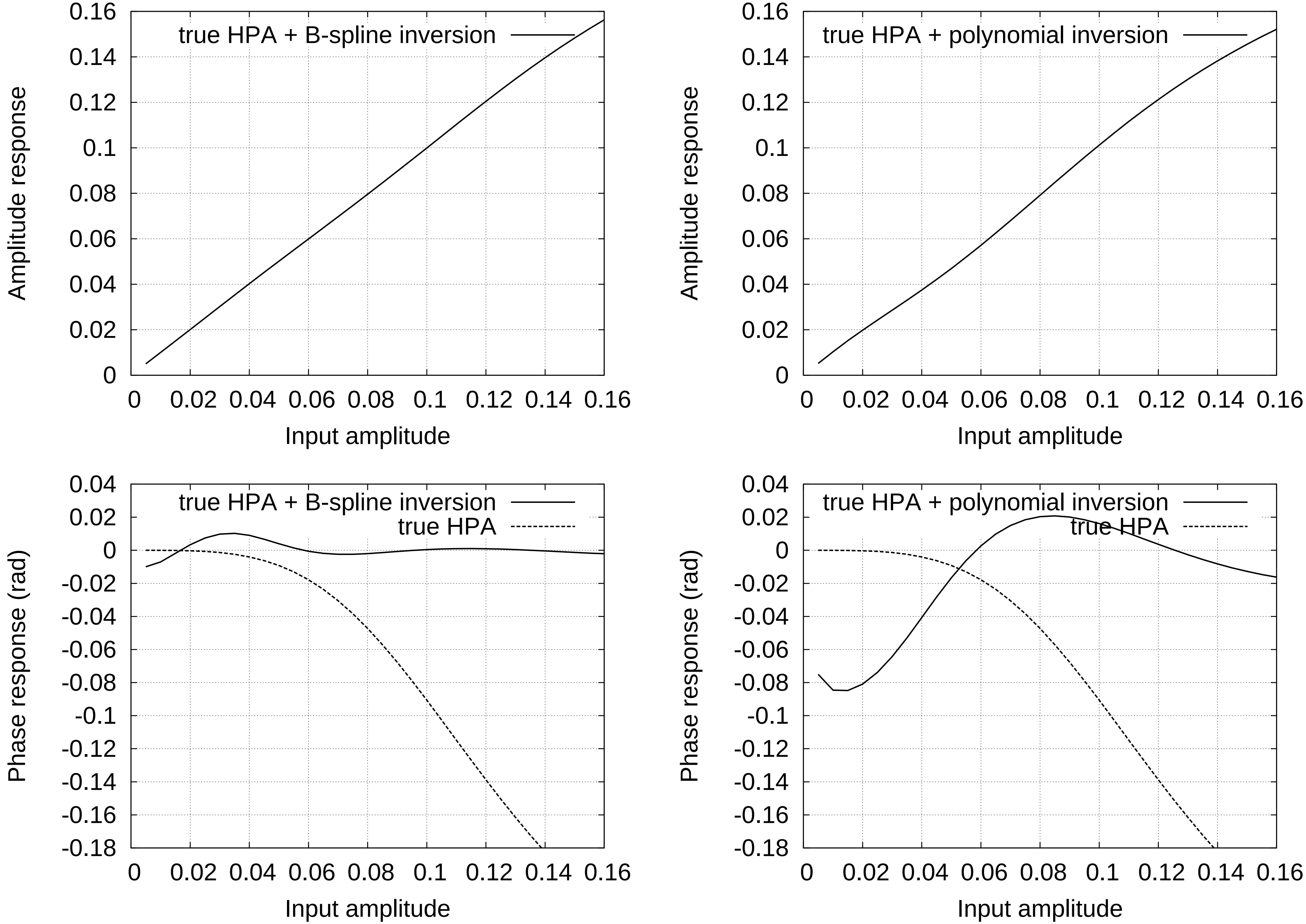
true value	BS average	BS std	Poly average	Poly std
1	1	NA	1	NA
-0.3732 -0.6123	-0.3732 -0.6122	9.152e-04 1.021e-03	-0.3735 -0.6120	9.176e-04 1.027e-03
0.3584 0.3676	0.3586 0.3676	9.702e-04 8.555e-04	0.3596 0.3680	9.723e-04 8.540e-04
0.3052 0.2053	0.3052 0.2052	9.278e-04 8.596e-04	0.3052 0.2058	9.262e-04 8.591e-04
0.2300 0.1287	0.2300 0.1286	7.806e-04 8.650e-04	0.2310 0.1277	7.786e-04 8.603e-04
0.7071 0.7071	0.7070 0.7069	1.161e-03 1.178e-03	0.7072 0.7066	1.165e-03 1.187e-03
0.6123 -0.3732	0.6122 -0.3733	1.051e-03 1.115e-03	0.6118 -0.3721	1.052e-03 1.116e-03
-0.3584 0.3676	-0.3583 0.3675	9.100e-04 1.056e-03	-0.3582 0.3689	9.077e-04 1.055e-03
-0.2053 -0.3052	-0.2054 -0.3051	9.343e-04 9.233e-04	-0.2064 -0.3052	9.327e-04 9.284e-04
0.1287 -0.2300	0.1287 -0.2299	8.017e-04 8.728e-04	0.1284 -0.2291	8.057e-04 8.615e-04

- Even higher accuracy for OBO $> 3\,$dB and/or SNR $> 5\,$dB

High Power Amplifier Identification

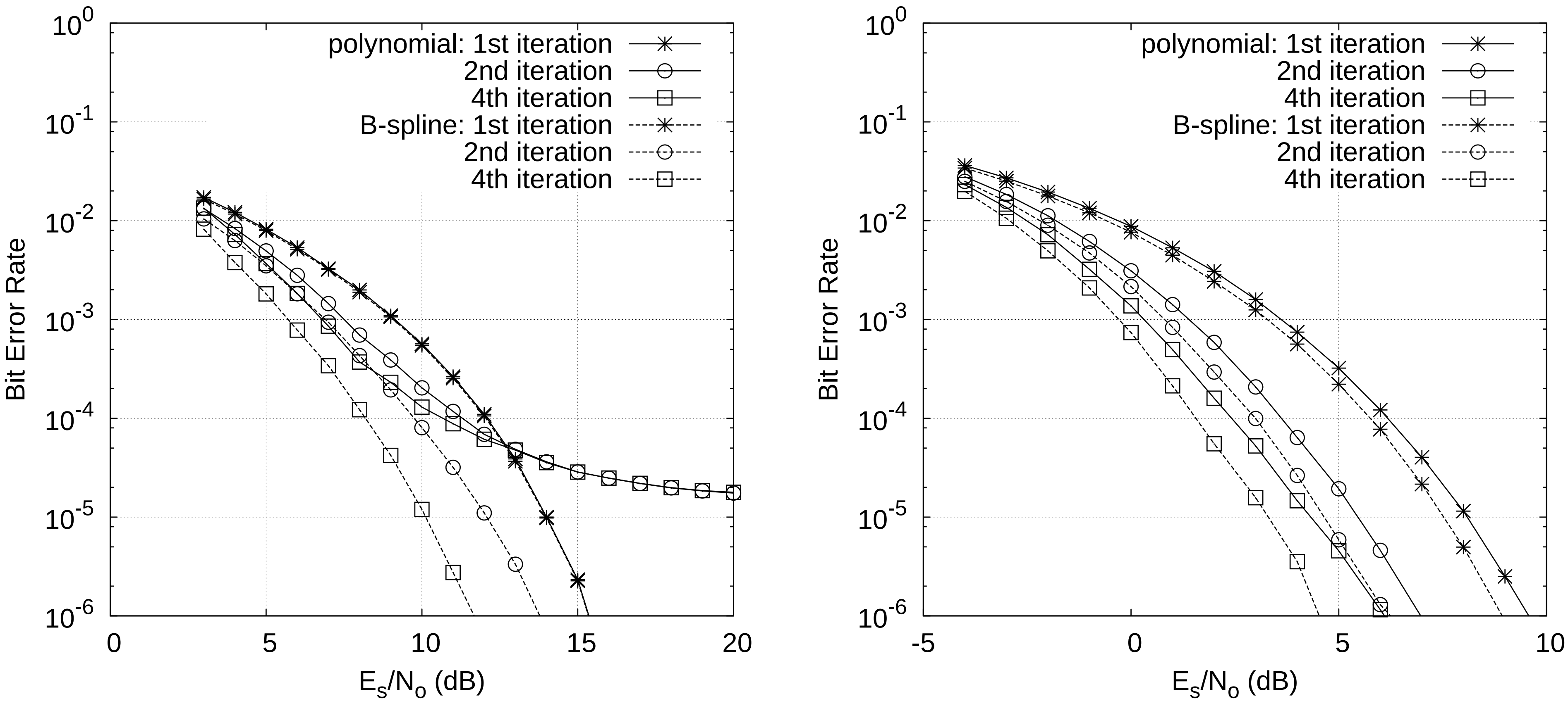


High Power Amplifier Inversion



Bit Error Rate

(a) OBO $= 3\,$dB (b) OBO $= 5\,$dB



- **B-spline** based approach **significantly outperforms** **polynomial** based counterpart

Conclusions

- **Complex-valued** B-spline neural network for a **real-world application**
  - Iterative frequency-domain decision feedback equalization for Hammerstein communication systems

- Optimal property and robustness of **B-spline** neural network
  - Particularly important for **inverting** transmitter nonlinear high power amplifier at receiver with pseudo **noisy** training **input**
  - Alternative least squares with closed-form LS estimates of linear channel and nonlinear BS model

- CV **B-spline** based approach significantly **outperforms** CV **polynomial** based counterpart