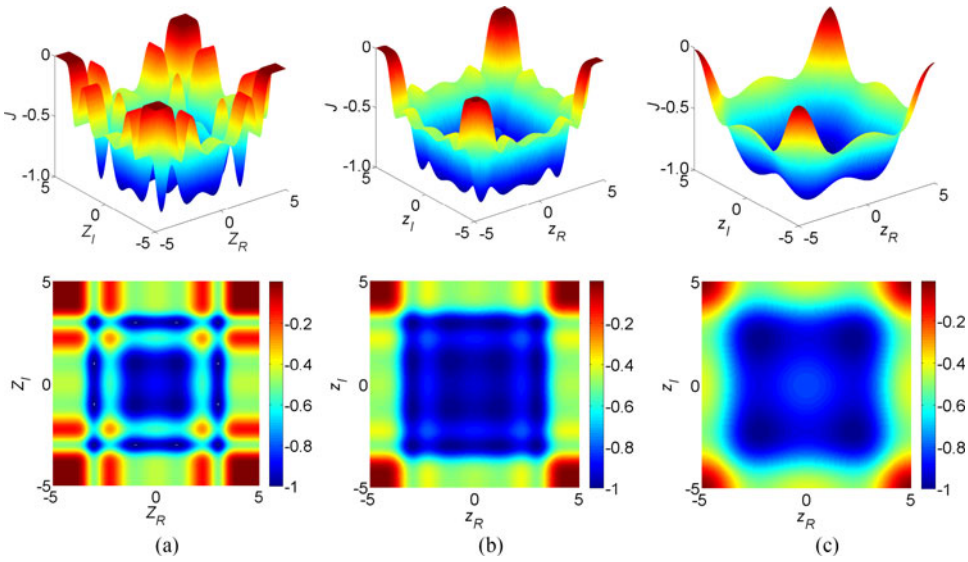
IEEE Photonics Journal
An IEEE Photonics Society Publication

Open Access

Information Entropy and Fuzzy Logic Based Equalizer for PolMux QAM Coherent Optical Communication Systems

Volume 9, Number 5, October 2017

Zhili Zhou
Yiju Zhan
Qingling Cai
Xiukai Ruan
Guihua Cui
Yuxing Dai
Haiyong Zhu



DOI: 10.1109/JPHOT.2017.2754504
1943-0655

Information Entropy and Fuzzy Logic Based Equalizer for PolMux QAM Coherent Optical Communication Systems

Zhili Zhou,[1] Yiju Zhan,[2] Qingling Cai,[2] Xiukai Ruan,[3,4] Guihua Cui,[3,4] Yuxing Dai,[3,4] and Haiyong Zhu[4]

[1]School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou 510006, China
[2]School of Engineering, Sun Yat-sen University, Guangzhou 510006, China
[3]Department of Information and Communication Engineering, Wenzhou University, Wenzhou 325035, China
[4]National-Local Joint Engineering Laboratory for Digitalized Electrical Design Technology, Wenzhou 325035, China

DOI:10.1109/JPHOT.2017.2754504

Manuscript received May 16, 2017; revised September 8, 2017; accepted September 17, 2017. Date of publication September 20, 2017; date of current version October 4, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61671329, Grant 61201426, Grant 61775170, in part by the Collaborative Innovation and Platform Environment Construction Project in Guangdong under Grant 2016A040403048, in part by the Key Projects of Public Welfare Research and Capacity Building in Guangdong under Grant 2015B010103003, in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LQ16F010010. Corresponding author: Xiukai Ruan. (e-mail: ruanxiukai@163.com)

Abstract: In polarization-multiplexed (PolMux) coherent optical communication systems, adaptive blind equalizer is efficient in demultiplexing and mitigating intersymbol interference (ISI). A novel blind algorithm based on Information Entropy and fuzzy logic is proposed, in which the Renyi's α entropy is adopted to measure the uncertainty of error between the desired and estimated probability density function (PDF). The nonparametric PDF estimator of Parzen window method is employed to estimate the PDF of symbols. Meantime, a fuzzy-logic tuning unit is designed to adjust the kernel size of Parzen window, which leads to fast convergence rate and small steady mean-square error. By simulation in PolMux-16 quadrature amplitude modulation (QAM) coherent systems, the correctness and effectiveness of the proposed algorithm are verified.

Index Terms: Information entropy, fuzzy logic, equalizer, optical communication.

1. Introduction

With the increasing demands for the capacity of optical communication networks, many techniques that can improve transmission capacity and spectrum efficiency have been proposed. For instance, digital coherent detection combined with digital signal processing (DSP) which can mitigate the transmission impairments effectively, has completely changed the way in design of optical communication systems. Due to the advantage of avoiding transmission pilot data and high bandwidth efficiency, numerous blind equalization algorithms [1]–[7] have been proposed in the past decades. Meantime, PolMux coherent technology doubles the spectral efficiency, and is widely employed in coherent optical communication systems. Together with quadrature phase shift keying (QPSK), 100 Gps optical coherent transmission is available for commercial systems [8].

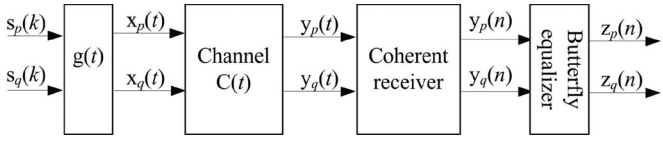


Fig. 1. Signal model.

In addition, to meet the growing demand of larger capacity in optical communication systems, two techniques involved: higher order multilevel modulation (e.g., MPSK and MQAM) and optical subcarrier multiplexing(e.g., Co-OFDM, Nyquist WDM). However, as the modulation constellation size increasing, the ISI will be more serious accordingly, which is a new challenge for DSP algorithms to mitigate ISI caused by the transmission impairments such as the Chromatic Dispersion (CD) and Polarization Mode Dispersion (PMD). To mitigate ISI, lots of valuable research findings [9]–[19] about blind equalization for high-order QAM coherent transmission systems have been proposed, in which the constant modulus algorithm (CMA) and multimodulus algorithm (MMA) are the best known algorithms. CMA and MMA typically minimize a cost function that is able to indirectly extract the higher order statistics of the signal or the current level of ISI from the equalizer output. Unfortunately, CMA and MMA are only effective for the constant modulus modulation and they exhibit large errors for high-order QAM systems. In addition, convergence performance of the CMA and MMA are closely related to the selection of step size. Motivated by the aforementioned problems, therefore the improving performance could be expected by finding some better cost functions.

From the knowledge of Information Theory, it is known that the data distribution contains more information than the evaluations based on simple statistics [20]. It is predictable that the performance of an equalizer tested based on data distribution will be better than that only based on simple statistics (e.g., CMA, MMA). Blind equalization techniques based on Information Theoretic criteria and PDF estimation of transmitted data [6], [7] have been applied to blind deconvolute linear channel and achieved satisfied performance. In this kind of equalizer, the Parzen window method [21] is employed to estimate the data PDF. However, it lacks the ability to overcome the problem of phase ambiguity which a carrier phase rotator is needed to produce the correct constellation orientation. In addition, the perhaps most important issue in Parzen window is the selection of window size, it is hard to determine the optimal kernel size of Parzen window. For different applications, different strategies may be taken.

In this paper, we propose a novel blind algorithm named INF-FL hereafter, which uses the Renyi's α entropy to measure the uncertainty of error between the desired and estimated PDF. A nonparametric PDF estimator of Parzen window method is employed to estimate the PDF of symbols, in which two one-dimensional Gaussian distributions (one for each of the real and imaginary parts) are assumed, hence it can preserve phase recovery like the MMA. The kernel size of Parzen window in the proposed algorithm can be adjusted adaptively using fuzzy logic method. Simulation results show that the proposed algorithm can obtain faster convergence rate and smaller steady MSE than CMA and MMA.

2. Signal Model

Throughout this paper, only the linear propagation impairments such as CD and PMD are concerned. Meanwhile, the discussion is restricted to the linear impairments generated by the optical fiber and the phase recovery is out of range of this paper as in [12]–[14]. Because in up to 100 Gps systems the symbol period is much smaller than the coherent time of the channel, all processing in optical propagation channel are considered as linear and time invariant.

As shown in Fig. 1, the modulated symbols sequences $\{s(k)\}$ pass through a shaping filter firstly

$$x_p(t) = \sum_k s_p(k) g(t - kT_s) \tag{1}$$

where $g(t)$ is the shaping filter, T_s is the symbol period, $x_p(t)$ is baseband transmitted signal on X-polarization. p (resp.q) denotes X-polarization (resp.Y-polarization). Then the baseband signal is transmitted through the optical channel, the received signal is also corrupted by additive noise

$$\mathbf{y}(t) = \sum_{n=-\infty}^{\infty} \mathbf{c}(t)\mathbf{x}(t - nT_s - t_0) + \mathbf{n}(t) \tag{2}$$

where the bivariate signal $\mathbf{y}(t) = [y_p(t), y_q(t)]^T$, $\mathbf{x}(t) = [x_p(t), x_q(t)]^T$, $\mathbf{n}(t) = [n_p(t), n_q(t)]^T$ denote the received signal, transmitted signal, additive Gaussian noise, respectively. t_0 denotes arbitrary time delay. The channel impulse response is given as

$$\mathbf{c}(t) = \begin{bmatrix} c_{p,p}(t) & c_{p,q}(t) \\ c_{q,p}(t) & c_{q,q}(t) \end{bmatrix} \tag{3}$$

where $c_{p,p}(t)$ and $c_{q,q}(t)$ denote the inter-symbol interference in X and Y polarization, respectively, $c_{p,q}(t)$ and $c_{q,p}(t)$ denote the inter-polarization interference created by PMD phenomenon.

In order to satisfy Shannon sampling theorem, on polarization p (resp.q), the received continuous-time signal is then sampled at $T_s/2$-space time intervals, the discrete-time signal is obtained from

$$\mathbf{y}\left(k\frac{T_s}{2}\right) = \sum_{n=-\infty}^{\infty} \mathbf{c}\left(k\frac{T_s}{2}\right)\mathbf{x}\left(k\frac{T_s}{2} - nT_s - t_0\right) + \mathbf{n}\left(k\frac{T_s}{2}\right). \tag{4}$$

Then the discrete-time signals pass through a length butterfly fractionally spaced equalizer (FSE), the output sample of which is retained in a decimation by two. The output sequences are given as follows

$$z_p(n) = \sum_{i=0}^{L-1} (\mathbf{w}_{p,p}\mathbf{y}_p + \mathbf{w}_{p,q}\mathbf{y}_q) \tag{5}$$

$$z_q(n) = \sum_{i=0}^{L-1} (\mathbf{w}_{q,p}\mathbf{y}_p + \mathbf{w}_{q,q}\mathbf{y}_q) \tag{6}$$

where $\mathbf{y}_p = [y_p((n-i)T_s + \frac{T_s}{2}), y_p((n-i)T_s)]^T$, $\mathbf{y}_q = [y_q((n-i)T_s + \frac{T_s}{2}), y_q((n-i)T_s)]^T$. And $\mathbf{w}_{p,p} = [w_{p,p}(2i), w_{p,p}(2i+1)]$, $\mathbf{w}_{q,p} = [w_{q,p}(2i), w_{q,p}(2i+1)]$, $\mathbf{w}_{p,q} = [w_{p,q}(2i), w_{p,q}(2i+1)]$, $\mathbf{w}_{q,q} = [w_{q,q}(2i), w_{q,q}(2i+1)]$ are the weights of butterfly FSE, and L is the length of the FSE.

3. The Proposed Blind Equalizer Algorithm

3.1. The Information Entropy Based Cost Function and Stochastic Gradient

For the sake of simplicity, we only consider single-polarization, and the situation is the same as the case of PolMux. From an Information Theoretic point of view, one would hope to extract as much information as possible from the PDF error to make z as close as possible to s in Fig. 1. The Renyi's α entropy [22] is adopted to measure uncertainty of the error

$$H_\alpha(X) = \frac{1}{1-\alpha}\log \int_{-\infty}^{+\infty} f_e^\alpha(x)dx \tag{7}$$

where $\alpha \geq 0$ and $\alpha \neq 1$, $f_e(x)$ is PDF error between the equalizer output and assumed PDF.

Renyi's entropy measure is flexible due to the parameter α, to enable several measurements of uncertainty. As illustrated in [20], when $\alpha = 2$, $H_2(X)$ implicitly uses an Euclidean distance from the point $f_e(x)$ in the simplex to the origin of the space. And $H_2(X)$ is a lower bound of Shannons entropy, it might be more efficient than Shannon's entropy for entropy maximization. Hence, in the rest of this paper, $\alpha = 2$ is chosen. Note that, the argument of the logarithm in quadratic Renyi's entropy has been called the *information potential (IP)*. When $\alpha > 1$, entropy is a monotonic decreasing function of IP. As a result, in this case, the entropy minimization is equivalent to IP maximization.

We calculate $H_2(X)$ for real and imaginary parts of the signals separately. In this scheme, the $H_2(X)$ can be formulated

$$H_2(X) = -\log\left(\int_{-\infty}^{+\infty} (f_{Z_R^{po}}(x) - f_{S_R^{po}}(x))^2 dx + \int_{-\infty}^{+\infty} (f_{Z_I^{po}}(x) - f_{S_I^{po}}(x))^2 dx \right) \tag{8}$$

where $f_X(x)$ denotes the PDF of X at x. $Z_R^{po} = \{|z_R(k)|^{po}\}$, $Z_I^{po} = |z_I(k)|^{po}$, $S_R^{po} = |s_R(k)|^{po}$, $S_I^{po} = |s_I(k)|^{po}$, and $z_R = \Re(z)$, $z_I = \Im(z)$, $s_R = \Re(s)$, $s_I = \Im(s)$, with $\Re(\cdot)$ and $\Im(\cdot)$ denote real and imaginary part, respectively.

Next, the nonparametric estimator of Parzen window method is employed to estimate the data PDFs $f_{Z_R^{po}}(x)$ and $f_{Z_R^{po}}(x)$ at time k

$$\hat{f}_{Z_R^{po}}(x) = \frac{1}{N_L} \sum_{i=0}^{N_L-1} K_{\sigma_0}(x - |z_R(k-i)|^{po}) \tag{9a}$$

$$\hat{f}_{Z_I^{po}}(x) = \frac{1}{N_L} \sum_{i=0}^{N_L-1} K_{\sigma_0}(x - |z_I(k-i)|^{po}) \tag{9b}$$

where N_L denotes the total number of previous symbols, K_{σ_0} is the Parzen window with the kernel size of σ_0. Here, Gaussian kernel is adopted

$$K_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}. \tag{10}$$

Similarly, the estimators used for the target PDF are given as follows:

$$\hat{f}_{S_R^{po}}(x) = \frac{1}{N_s} \sum_{i=1}^{N_s} K_{\sigma_0}(x - |s_R(i)|^{po}) \tag{11a}$$

$$\hat{f}_{S_I^{po}}(x) = \frac{1}{N_s} \sum_{i=1}^{N_s} K_{\sigma_0}(x - |s_I(i)|^{po}) \tag{11b}$$

where N_s is the number of the complex symbols in the constellation, and the probability of symbols are supposed to be equally likely. Taking account of Gaussian kernels, we have

$$\int_{-\infty}^{+\infty} K_\sigma(x - C_1) K_\sigma(x - C_2) dx = K_{\sigma\sqrt{2}}(C_1 - C_2). \tag{12}$$

Finally, substituting Eq. (9), Eq. (10) and Eq. (11) in Eq. (8), and together with Eq. (12), we have

$$\begin{aligned} H_2(X) = -\log\Bigg(&\frac{1}{N_L^2}\sum_{j=0}^{N_L-1}\sum_{i=0}^{N_L-1} K_\sigma(|z_I(k-i)|^{po} - |z_I(k-j)|^{po})\Bigg) \\ &+ \frac{1}{N_L^2}\sum_{j=0}^{N_L-1}\sum_{i=0}^{N_L-1} K_\sigma(|z_R(k-i)|^{po} - |z_R(k-j)|^{po})) \\ &+ \frac{1}{N_S^2}\sum_{j=1}^{N_S}\sum_{i=1}^{N_S} K_\sigma(|s_I(i)|^{po} - |s_I(j)|^{po})) \\ &+ \frac{1}{N_S^2}\sum_{j=1}^{N_S}\sum_{i=1}^{N_S} K_\sigma(|s_R(i)|^{po} - |s_R(j)|^{po})) \\ &- \frac{2}{N_L N_s}\sum_{j=1}^{N_s}\sum_{i=0}^{N_L-1} K_\sigma(|z_R(k-i)|^{po} - |s_R(j)|^{po}) \\ &- \frac{2}{N_L N_s}\sum_{j=1}^{N_s}\sum_{i=0}^{N_L-1} K_\sigma(|z_I(k-i)|^{po} - |s_I(j)|^{po}) \end{aligned} \tag{13}$$

where $\sigma = \sqrt{2}\sigma_0$.

The computational complexity of PDF for INF-FL is $O(N_L)$, so it is necessary to make a tradeoff between performance and computational complexity. As illustrated in [20], it still converges well with $N_L = 1$. As a result, the first and second terms on the right side equal zero. And the third and fourth terms on the right side are independent with $\mathbf{w}$. Finally, the fifth and sixth terms pull the output of the equalizer toward the desired PDFs. It is noteworthy that the negative logarithm is a monotonic function and therefore does not affect the location of the extremum of the cost function, so it is removed in the cost function, which is defined as

$$J(\mathbf{w}) = -\frac{2}{N_s}\sum_{j=1}^{N_s} K_\sigma(|z_R(k)|^{po} - |s_R(j)|^{po}) - \frac{2}{N_s}\sum_{j=1}^{N_s} K_\sigma(|z_I(k)|^{po} - |s_I(j)|^{po}). \tag{14}$$

When $\sigma = 2$, the 3D surface and contour diagrams of the cost function $J(\mathbf{w})$ for 16QAM with $po = 1$, 2, 3 have been shown in Fig. 2. It is extremely clear that a smaller value of po will lead to a smoother surface, and vice versa. For the case of $po = 2$, in these points of ideal 16QAM constellation [e.g., $(\pm1, \pm1)$, $(\pm1, \pm3)$, $(\pm3, \pm1)$, $(\pm3, \pm3)$], there exist corresponding global minima of the cost function $J(\mathbf{w})$, which can be found by stochastic gradient descent schemes. Hence, in the rest of this paper, we will choose $po = 2$. As a contrast, the cost functions of the CMA and MMA are of the forms

$$J_{CMA}(\mathbf{w}) = E\left[(z^2(k) - R_2)^2\right] \tag{15}$$

$$J_{MMA}(\mathbf{w}) = E\left[(z_R^2(k) - R_{2,R})^2\right] + E\left[(z_I^2(k) - R_{2,I})^2\right] \tag{16}$$

where $E[\cdot]$ indicates statistical expectation. R_2, $R_{2,R}$ and $R_{2,I}$ are given by $R_2 = E[s^4(k)]/E[s^2(k)]$, $R_{2,R} = E[s_R^4(k)]/E[s_R^2(k)]$ and $R_{2,I} = E[s_I^4(k)]/E[s_I^2(k)]$, respectively.

Fig. 3 shows the 3D surface diagrams of the cost functions for CMA and MMA. From Figs. 2 and 3, one can find that the cost function of information entropy based is better than those of CMA and MMA. Hence, there are reasons to believe that the information entropy based algorithm can achieve better performance.

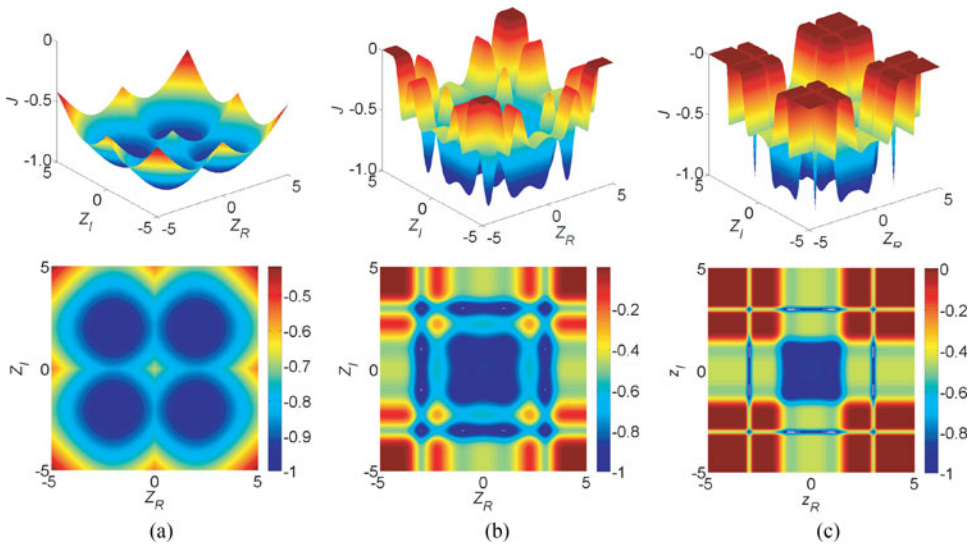


Fig. 2. 3D surface and contour diagram of cost function with $\sigma = 2$. (a) $po = 1$. (b) $po = 2$. (c) $po = 3$.

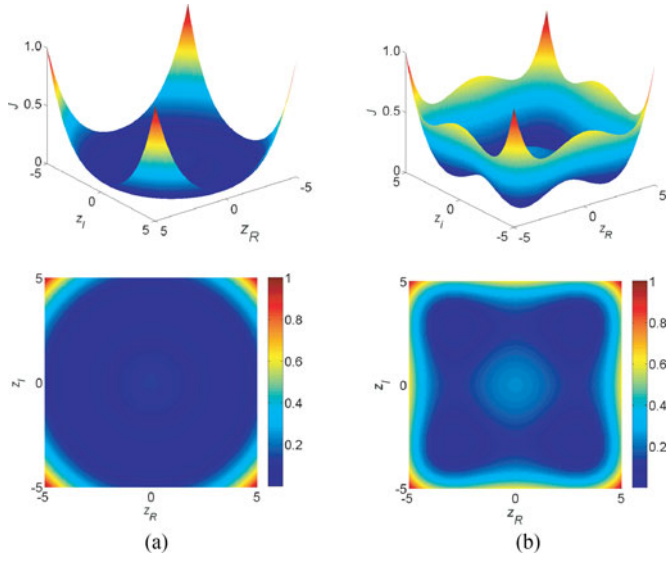


Fig. 3. 3D surface diagram of cost function (a) CMA (b) MMA.

With $z(k) = \mathbf{w}(k)^T \mathbf{y}(k)$, it can be rewritten as

$$\begin{aligned} z(n) &= z_R(k) + jz_I(k) \\ &= (\mathbf{w}_R^T \mathbf{y}_R(k) - \mathbf{w}_I^T \mathbf{y}_I(k)) + j(\mathbf{w}_R^T \mathbf{y}_I(k) + \mathbf{w}_I^T \mathbf{y}_R(k)). \end{aligned} \tag{17}$$

It means that, $\partial z_R(k)/\partial \mathbf{w}_R = \mathbf{y}_R(k)$, $\partial z_R(k)/\partial \mathbf{w}_I = -\mathbf{y}_I(k)$, $\partial z_I(k)/\partial \mathbf{w}_R = \mathbf{y}_I(k)$, $\partial z_I(k)/\partial \mathbf{w}_I = \mathbf{y}_R(k)$. Therefore, with $N_L = 1$, $po = 2$, the derivative of $J(\mathbf{w})$ with respect to the equalizer weights is given

as follows

$$\begin{aligned}\nabla_{\mathbf{w}} J(\mathbf{w}) &= \frac{\partial J(\mathbf{w})}{\partial \mathbf{w}_R} + j\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}_I} \\ &= \frac{1}{\sqrt{2\pi}N_s\sigma^3}\sum_{j=1}^{N_s}(sgn(z_R(k))|z_R(k)|(|z_R(k)|^2 - |s_R(j)|^2)e^{\frac{-(|z_R(k)|^2-|s_R(j)|^2)^2}{2\sigma^2}} \\ &\quad + jsgn(z_I(k))|z_I(k)|(|z_I(k)|^2 - |s_I(j)|^2)e^{\frac{-(|z_I(k)|^2-|s_I(j)|^2)^2}{2\sigma^2}})\mathbf{y}^*(k) \end{aligned} \tag{18}$$

where $sgn(\cdot)$ is the sign function, and superscript $*$ denotes complex conjugation operator.

Denote R_R and R_I by $|s_R(j)|^2$ and $|s_I(j)|^2$, respectively, where $R_R = E[s_R^4(k)]/E[s_R^2(k)]$, $R_I = E[s_I^4(k)]/E[s_I^2(k)]$. Then the stochastic gradient of the cost function with Gaussian kernel can be rewritten as

$$\begin{aligned}\nabla_{\mathbf{w}} J(\mathbf{w}) = \Bigg(&\frac{(|z_R(k)|^2 - R_R)}{\sqrt{2\pi}\sigma^3} sgn(z_R(k))|z_R(k)|e^{-\frac{(|z_R(k)|^2-R_R)^2}{2\sigma^2}} \\ &+ j\frac{(|z_I(k)|^2 - R_I)}{\sqrt{2\pi}\sigma^3} sgn(z_I(k))|z_I(k)|e^{-\frac{(|z_I(k)|^2-R_I)^2}{2\sigma^2}}\Bigg)\mathbf{y}^*(k). \end{aligned} \tag{19}$$

To compensate the $1/\sigma^3$ term that appears in the derivative of Gaussian kernel, a normalized step size $\mu_\sigma = \mu\sigma^3$ has been introduced. In the following discussion, μ will be considered as the step size [7]. And the equalizer coefficients weights are adapted by

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu_\sigma \nabla_{\mathbf{w}} J(\mathbf{w}). \tag{20}$$

3.2. The Kernel Size of Parzen Window for The Proposed Approach

Based on aforementioned discussion, before calculation of cost function, the PDFs estimation has to be done in nonparametric way as an intermediate step. As a famous nonparametric PDF estimator, it is well-known that the most important parameter in Parzen window method is the kernel size σ. The quality of the estimator is normally quantified by the mean integrated squared error ($MISE$), which is a convenient measure for analyzing the $\hat{f}(x)$ (estimated PDF) from the $f(x)$ (true PDF) as the sum of the integrated square bias and the integrated variance [23]

$$MISE(\hat{f}) = \int E[\hat{f}(x) - f(x)]^2 dx = \int (E[\hat{f}(x)] - f(x))^2 dx + \int var\hat{f}(x) dx. \tag{21}$$

The bias and the variance term are given as follows:

$$\int (E[\hat{f}(x)] - f(x))^2 dx \approx \frac{1}{2}\sigma^4 k_2^2 \int f''(x)^2 dx \tag{22}$$

$$\int var\hat{f}(x) dx \approx N^{-1}\sigma^{-1} \int K(x)^2 dx \tag{23}$$

where $k_2 = \int x^2 K(x) dx$, N denotes the number of observation samples. Approximations of (22) and (23) show that the kernel size has to be chosen in a trade-off between estimation bias and variance, whatever method of estimation is involved.

In application, the choice of kernel size σ usually depends on the degree of sensitive to the bias or variance. When it is applied to equalizer, a large value of σ means interaction of each equalized symbol with more symbols in the constellation which leads to fast convergence, while a small one will win better accuracy of the final solution. Fig. 4 shows the 3D surface and contour diagrams of cost function with $\sigma = 2, 3, 6$ at the condition of $po = 2$. It is definitely clear that a large value of σ means fast global convergence and less accuracy, and vice versa. Hence, the σ should be adjusted adaptively in the process of convergence.

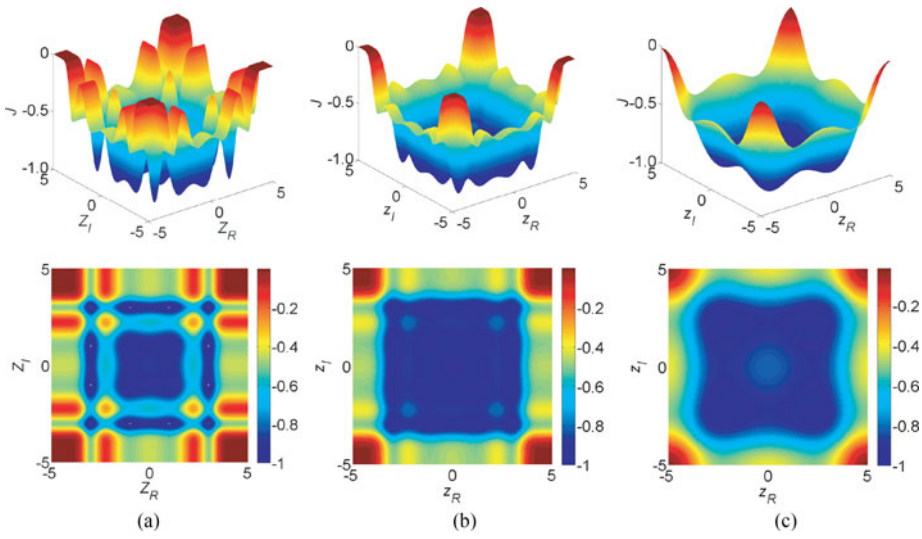


Fig. 4. 3D surface and contour diagram of cost function with $po = 2$. (a) $\sigma = 2$. (b) $\sigma = 3$. (d) $\sigma = 6$.

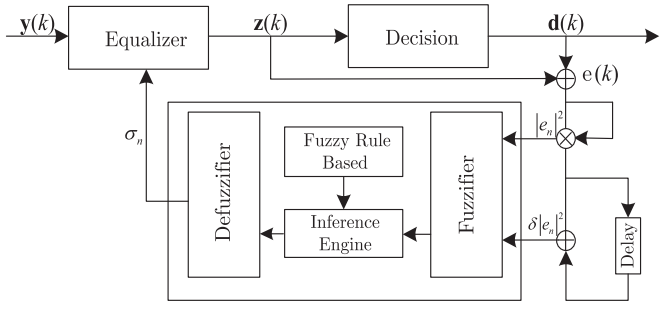


Fig. 5. Schematic of the fuzzy inference system for tuning the kernel size of Parzen window.

3.3. The Variable Kernel Size of Parzen Window Based on The Fuzzy Logic

Fuzzy sets are actually functions mapping a value that might be a member of the set to a number between zero and one, indicating its actual degree of membership [24]. Fuzzy logic, in the sense of fuzzy set theory, provides the underpinnings for designing and writing fuzzy models.

With fuzzy logic, one can significantly reduce development time, and model highly complex nonlinear systems. Thus, we choose the fuzzy inference system (FIS) of Fig. 5, which is based on the principle of fuzzy logic [24]–[26] to map the two input variables, $|e_n^2|$ and $\delta|e_n^2|$, into an appropriate kernel size σ_n. The two input variables are defined respectively as follows:

$$|e_n|^2 = \frac{1}{N_{sm}} \sum_{l=0}^{N_{sm}-1} |e(k-l)|^2, n = \lfloor \frac{k}{N_{sm}} \rfloor \tag{24}$$

$$\delta|e_n|^2 = |e_n|^2 - |e_{n-1}|^2 \tag{25}$$

where N_{sm} is the number of error samples processed when obtaining the short term average. $\lfloor \cdot \rfloor$ denotes the integer floor operation. Then decision error $e(k)$ can be formulated as

$$e(k) = d(k) - z(k). \tag{26}$$

Note that the FIS operates once for each N_{sm} samples, and the output σ_n is used as the kernel size of the Parezen window for the subsequent N_{sm} samples.

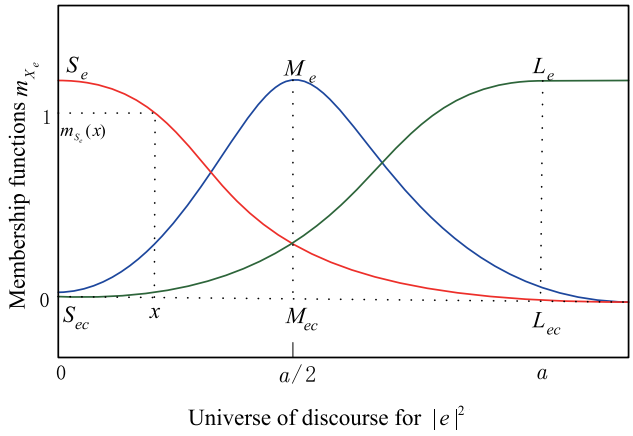


Fig. 6. Membership function against universe of discourse for $|e(n)|^2$.

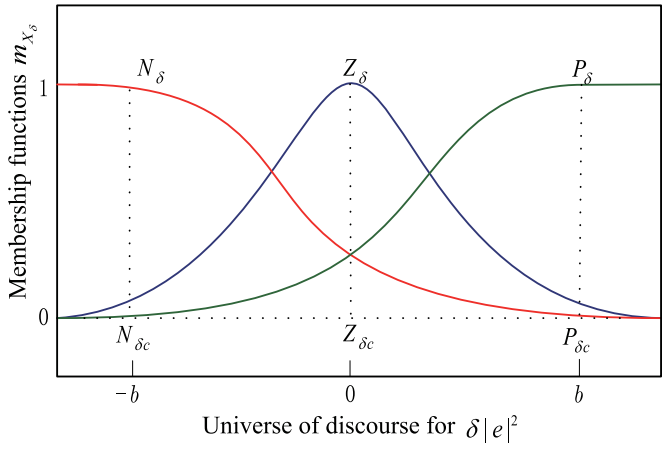


Fig. 7. Membership function over universe of discourse for $\delta|e(n)|^2$.

The fuzzifier is used to transform the two input variables of the FIS to the respective degrees, which they belong to the corresponding fuzzy sets via membership functions (MBFs), which is interpreted as a measure of the compatibility between a value from the domain and the idea underlying the fuzzy set. The Gaussian MBFs are selected to cover the universe of the input and output variables

$$m_X(x) = e^{-\frac{(x-X_c)^2}{2\rho_X}} \tag{27}$$

where, ρ_x denotes ρ_e, ρ_δ for $|e_n|^2$ and $\delta|e_n|^2$, respectively. For $|e_n|^2$, X denotes small (S_e), medium (M_e), and large (L_e), which are the fuzzy sets used to partition the universe of discourse, with the exception that $m_{L_e} = 1$ for $x \geq a$, and X_c denotes S_{ec}, M_{ec} and L_{ec} (corresponding to 0, $a/2$, a, respectively) which are the centroids of MBFs (shown in Fig. 6). For $\delta|e_n|^2$, X denotes negative (N_δ), zero (Z_δ) and positive (P_δ), with the exception that $m_{N_\delta}(x) = 1$ for $x \leq -b$ and $m_{P_\delta}(x) = 1$ for $x \geq b$, and X_c denotes $N_{\delta c}$, $Z_{\delta c}$ and $P_{\delta c}$(corresponding to $-b$, 0, b, respectively), which are the centroids of MBFs (shown in Fig. 7). For instance, in Fig. 6, the red line denotes the MBF

$$m_{S_e}(x) = e^{-\frac{(x-S_{ec})^2}{2\rho_e}}. \tag{28}$$

Thus, an actual value of x has a $m_{S_e}(x)$ degree of membership.

TABLE 1
Fuzzy Rule for The Kernel Size

σ_n		$\delta\|e_n\|^2$		
		N_δ	Z_δ	P_δ
$\|e_n\|^2$	S_e	${}^1S_\sigma$	${}^2S_\sigma$	${}^3L_\sigma$
	M_e	${}^4S_\sigma$	${}^5M_\sigma$	${}^6L_\sigma$
	L_e	${}^7M_\sigma$	${}^8L_\sigma$	${}^9L_\sigma$

TABLE 2
Comparison of the Computational Complexities of the Equalizers

Equalizers	Multiplications	Additions	Exp(·) evaluations
CMA	$32L_e + 14$	$24L_e + 2$	–
MMA	$32L_e + 24$	$24L_e + 8$	–
INF-FL	$32L_e + 32 + 44/N_{sm}$	$24L_e + 12 + 44/N_{sm}$	$2 + 12/N_{sm}$

The fuzzy inference engine constructs a set of fuzzy IF-THEN rules as shown in Table 1. The fuzzy sets used to partition kernel size σ are labelled as small (S_σ), medium (M_σ) and large (L_σ) in Table 1. Since there are 3 fuzzy sets for $|e_n|^2$ and $\delta|e_n|^2$, respectively, the number of fuzzy IF-THEN rules is 9. For example, as rule 5: IF $|e_n|^2$ is M_e AND $\delta|e_n|^2$ is Z_δ THEN σ_n is M_σ. Let $m_{X_\sigma}(\sigma_n(i))$ be the MBF value at location $\sigma_n(i)$, where $1 \leq i \leq 9$. A minimum operation is generally employed to truncate the output fuzzy set for each rule. For instance, the MBF value at $\sigma_n(5)$ is

$$m_{X_\sigma}(\sigma_n(5)) = min\{m_{M_e}(|e_n|^2), m_{Z_\delta}(\delta|e_n|^2)\}. \tag{29}$$

The defuzzifier is adopted to obtain a crisp value for the kernel size σ_n. Since decisions are made based on the testing of all of the rules in the FIS, the rules need to be combined in some manner in order to make a decision

$$\sigma_n = \frac{\sum_{i=1}^{9}(\sigma_n(i) \cdot m_{X_\sigma}(\sigma_n(i)))}{\sum_{i=1}^{9} m_{X_\sigma}(\sigma_n(i))}. \tag{30}$$

4. Computational Complexity Analysis

The updating expression of equalizer coefficients weights ((20)) is that the MMA updating expression multiplied by the exponential term for real and imaginary part, respectively. Therefore, the increment of computational burden is caused by these exponential terms. For the FLS, it requires 6 exponential operations for X and Y polarization per σ_n update, respectively. Note that the σ_n is updated once for each N_{sm} symbols, so the computational complexity is divided by N_{sm} for per symbol update. The comparison of the computational complexities with CMA and MMA for per weight update is given in Table 2 where L_e is the tap number of equalizers. When $N_{sm} = 40$ and $L_e = 9$, the multiplication

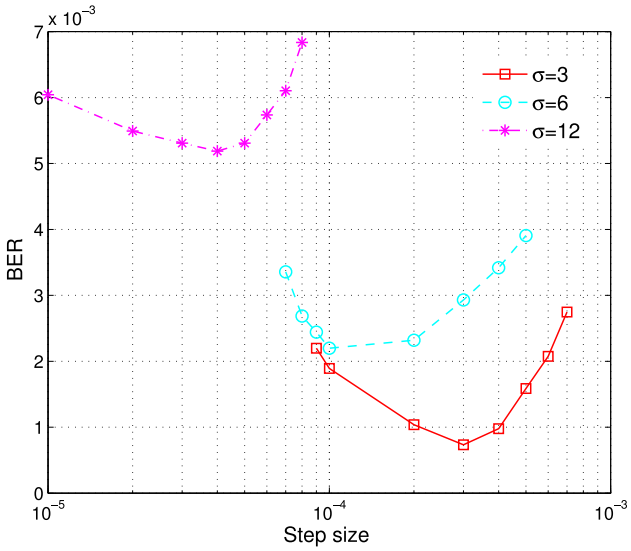


Fig. 8. BER versus the step size (residual $CD = 1000$ ps/nm, $\tau_{DGD} = 50$ ps, $\theta = \pi/4$, $OSNR = 20$ dB).

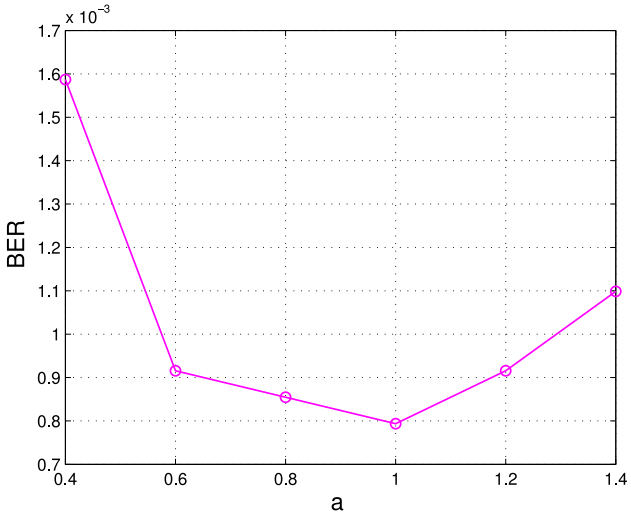


Fig. 9. BER versus the parameter a (residual $CD = 1000$ ps/nm, $\tau_{DGD} = 50$ ps, $\theta = \pi/4$, $OSNR = 20$ dB).

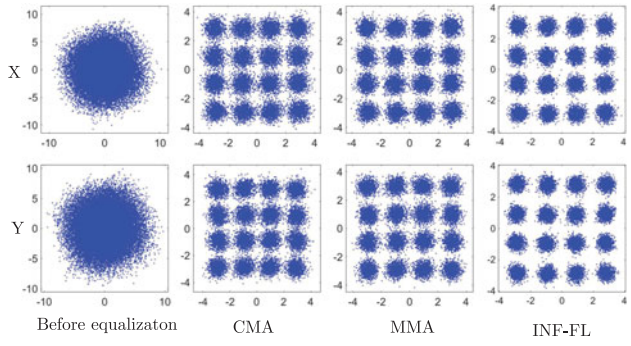


Fig. 10. PolMux-16QAM constellations diagram (residual $CD = 1000$ ps/nm, $\tau_{DGD} = 50$ ps, $\theta = \pi/4$, $OSNR = 20$ dB).

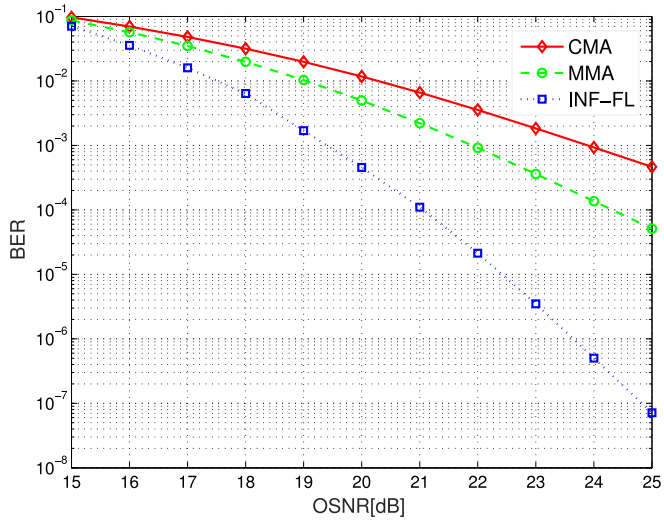


Fig. 11. BER versus OSNR(residual $CD = 1000$ ps/nm, $\tau_{DGD} = 50$ ps, $\theta = \pi/4$).

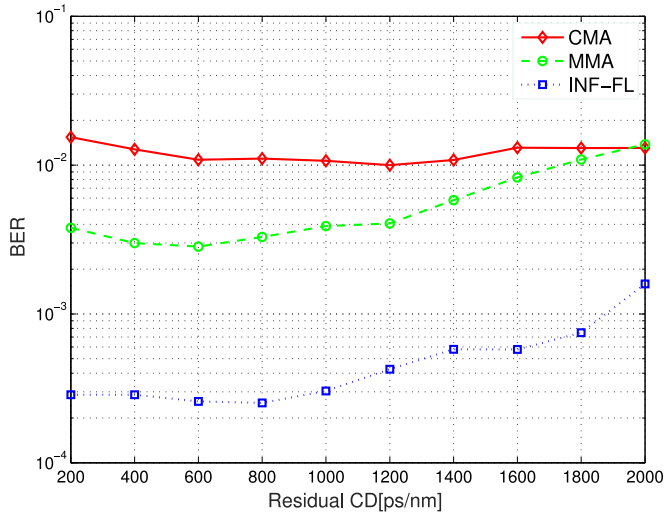


Fig. 12. BER versus residual CD($\tau_{DGD} = 50$ ps, $\theta = 0$, $OSNR = 20$ dB).

operations are 604, 624 and 644 for CMA, MMA and INF-FL, respectively. However, the INF-FL involves the additional 5 exponential operations.

5. Demonstration and Results

In this section, we provide simulation results to illustrate and verify the theory developed. These experiments are based on a PolMux-16QAM coherent system with symbol rate of 14GBaud. The symbol number of pseudo random sequences is 16,384. Since the phase recovery is out of range of this paper, the laster beat linewidth is set to 0. The equalizer length $L = 9$, and the initial weights are set to 0 with the exception of the central tap of $\mathbf{w}_{p,p}$ and $\mathbf{w}_{q,q}$ are 1. The considered algorithms in simulation are listed below:

1) CMA
2) MMA
3) INF-FL: the proposed information entropy and fuzzy logic based algorithm

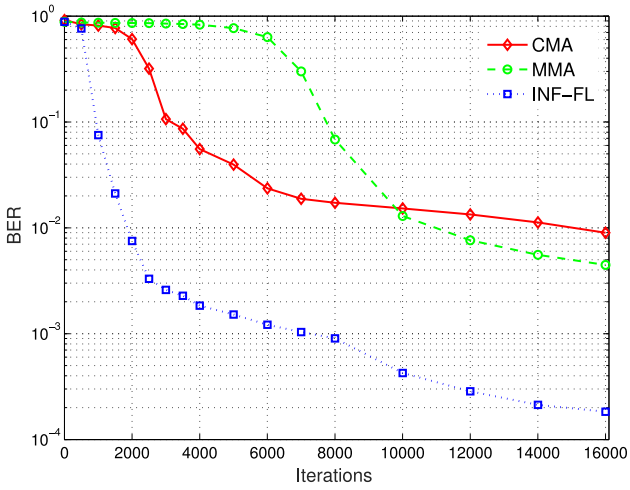


Fig. 13. BER versus the number of iterations (residual $CD = 1000$ ps/nm, $\tau_{DGD} = 50$ ps, $\theta = \pi/4$, $OSNR = 20$ dB).

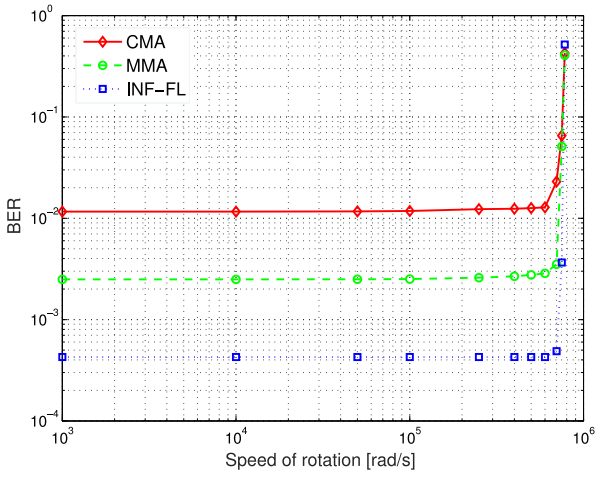


Fig. 14. BER versus speed of rotation($\tau_{DGD} = 50$ ps, $OSNR = 20$ dB).

For the sake of simplicity, the optimum step size for CMA and MMA is not discussed here. After extensive simulation studies, the step size for CMA and MMA is chosen 1×10^{-5}. For the parameters of INF-FL, we firstly adopt various constant kernel sizes σ to obtain the optimum step sizes.

Fig. 8 illustrates BER versus the step size in the condition of $\sigma = 3, 6, 12$ (2 is so small for σ, that the algorithm can not converge). One can find the optimum step sizes for $\sigma = 3, 6, 12$ are 3×10^{-4}, 1×10^{-4}, 4×10^{-5}, respectively, and the lager σ is, the smaller step size will be. As a result, the step size for INF-FL would slightly bigger than 3×10^{-4}. Then, as described in Section 3.1, a small σ will win better accuracy of the final solution, the values 2, 6, 12, are chosen for S_σ, M_σ and L_σ, respectively. And 20 is chosen for N_{sm} as reference [25]. The parameters ρ_e and ρ_δ are related with the variance of $|e_n^2|$ and $\delta|e_n^2|$, respectively. By computing with various of kernel σ, it was found that the variance of $|e_n^2|$ and $\delta|e_n^2|$ are close to the magnitude of 0.01 and 0.001, respectively. The maximum of $\delta|e_n^2|$ equal approximately 0.2 (see Fig. 15), therefore, we set the b to 0.2. Lastly, Fig. 9 illustrates the BER versus parameter a, from which we can find that the optimum value for parameter a is 1. Based on the aforementioned discussion and extensive simulation studies, the

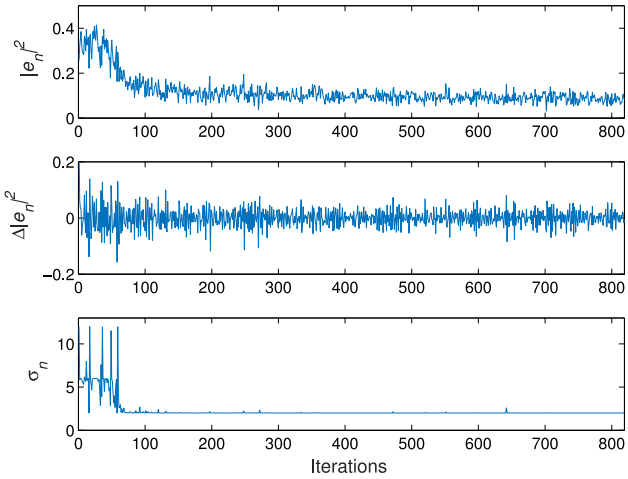


Fig. 15. The $|e_n|^2$, $\delta|e_n|^2$, and σ_n versus the number of iterations (residual $CD = 1000$ ps/nm, $\tau_{DGD} = 50$ ps, $\theta = \pi/4$, $OSNR = 20$ dB).

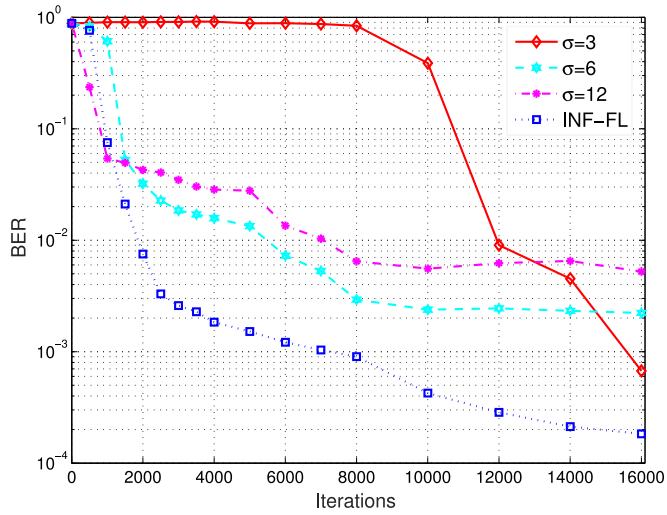


Fig. 16. BER versus the number of iterations for various constant kernel size ($CD = 1000$ ps/nm, $\tau_{DGD} = 50$ ps, $\theta = \pi/4$, $OSNR = 20$ dB).

parameters for the INF-FL are listed as follows: the step size is 4×10^{-4}, $a = 1$, $b = 0.2$, $\rho_e = 0.01$, $\rho_\delta = 0.001$, $N_{sm} = 20$. The S_σ, M_σ and L_σ are 2, 6 and 12, respectively.

Fig. 10 shows the constellations diagrams of signals before equalization and after convergence with CMA, MMA, and INF-FL, respectively. The transmission channel is set as follows: residual $CD = 1000$ ps/nm, DGD delay $\tau_{DGD} = 50$ ps, polarization rotation angle $\theta = \pi/4$ (the worst case), the $OSNR = 20$ dB(in 0.1 nm). For the INF-FL, signal constellation is the most concentrated and clear which shows that the INF-FL outperforms CMA and MMA.

Fig. 11 shows the BER as the function of the OSNR. One can find that the INF-FL is far more efficient than CMA and MMA. Meantime, it can be seen that the OSNR penalty is 2.5 dB and 4.5 dB less than that of CMA and MMA, respectively at BER=10^{-3}. Fig. 12 shows the BER as the function of the residual CD. The INF-FL ensures obviously lower BER than CMA and MMA. Fig. 13 shows the BER as the function of the number of iterations. One can find that the INF-FL can achieve faster convergence rate and smaller steady error in steady state.

To gain insight into the dynamical behavior of INF-FL, for the sake of simplicity, let the residual CD be null, then simulate an endless polarization rotation with the Jones matrix

$$\mathbf{J} = \begin{bmatrix} cos(\omega t) & sin(\omega t) \\ -sin(\omega t) & cos(\omega t) \end{bmatrix} \tag{31}$$

where ω is the angular rate of rotation.

Fig. 14 illustrates the tracking performance for different equalizers. It is clear that the behavior of the three equalizers are similar. The three equalizers are able to track angular frequencies $\sim 8 \times 10^5$ rad/s. The performance for INF-FL is found to be better than that of CMA and MMA for $\omega < 8 \times 10^5$ rad/s.

Fig. 15 shows the $|e_n|^2$, $\delta|e_n|^2$, and σ_n versus the number of iterations. Note that, as described in Section 3.3, they are updated once for each N_{sm} symbols. The length of pseudo random sequences is 16,384 and $N_{sm} = 20$, so the number of iterations is 819. One can find that in the initial stage (about from 0 to 60 iterations), when the $|e_n|^2$ and $\delta|e_n|^2$ are large, the large kernel size is chosen to reinforce convergence speed. Otherwise, when the $|e_n|^2$ and $\delta|e_n|^2$ are smaller, the small one is chosen for the sake of winning small steady MSE.

To validate the effectiveness of INF-FL further, Fig. 16 illustrates BER versus the number of iterations for various constant kernel size σ. One can find that for the large kernel size, the algorithm achieves fast convergence speed but large MSE. Otherwise, for the small kernel size, the algorithm achieves small MSE but slow convergence speed. For the INF-FL, as shown in Fig. 15, kernel size changed adaptively according $|e_n|^2$ and $\delta|e_n|^2$, which can achieve fast convergence rate and small steady MSE in steady state.

6. Conclusion

A novel blind equalizer for PolMux QAM coherent optical communication systems has been proposed. It uses the Renyi's α entropy to measure the uncertainty of error between the desired and estimated PDF. The nonparametric PDF estimator of Parzen window method is employed to estimate the PDF of symbols, while the kernel size of the Parzen window controls both the convergence speed and the accuracy of the final solution. A fuzzy-logic method using decision error is employed to adjust the kernel size of the Parzen window. Comparing with the algorithms of various constant kernel size, the fuzzy-logic method can adjust the kernel size according to the decision error, which can achieve fast convergence rate and small MSE. The comparison of performances are also made with the traditional blind algorithm CMA and MMA in various propagation conditions. The result confirms its high effectiveness and robustness for the proposed equalizer at the cost of slightly addition of computational complexity.

References

[1] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. 28, no. 11, pp. 1867–1875, Nov. 1980.
[2] J. Yang, J. J. Werner, and G. A. Dumont, "The multimodulus blind equalization and its generalized algorithms," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 5, pp. 997–1015, Jun. 2002.
[3] J. Treichler and B. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 31, no. 2, pp. 459–472, Apr. 1983.
[4] V. Sharma and V. N. Raj, "Convergence and performance analysis of Godard family and multimodulus algorithms for blind equalization," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1520–1533, Apr. 2005.
[5] J. T. Yuan and T. C. Lin, "Equalization and carrier phase recovery of CMA and MMA in blind adaptive receivers," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3206–3217, Jun. 2010.
[6] D. Erdogmus, K. E. Hild, J. C. Principe, and M. Lazaro, "Adaptive blind deconvolution of linear channels using Renyi's entropy with Parzen window estimation," *IEEE Trans. Signal Process.*, vol. 52, no. 6, pp. 1489–1498, Jun. 2004.
[7] M. Lazaro, I. Santamaria, D. Erdogmus, K. E. Hild, C. Pantaleon, and J. C. Principe, "Stochastic blind equalization based on PDF fitting using Parzen estimator," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 696–704, Feb. 2005.
[8] K. Roberts *et al.*, "Performance of dual-polarization QPSK for optical transport systems," *J. Lightw. Technol.*, vol. 27, no. 16, pp. 3546–3559, Aug. 2009.
[9] S. J. Savory, "Digital filters for coherent optical receivers," *Opt. Exp.*, vol. 16, no. 2, pp. 804–817, 2008.

[10] I. Fatadin, D. Ives, and S. J. Savory, "Blind equalization and carrier phase recovery in a 16-QAM optical coherent system," *J. Lightw. Technol.*, vol. 27, no. 15, pp. 3042–3049, Aug. 2009.
[11] S. J. Savory, "Digital coherent optical receivers: Algorithms and subsystems," *IEEE J. Sel. Topics Quantum Electron*, vol. 16, no. 5, pp. 1164–1179, Sep./Oct. 2010.
[12] M. Selmi, C. Gosset, M. Noelle, P. Ciblat, and Y. Jaoun, "Block-wise digital signal processing for PolMux QAM/PSK optical coherent systems," *J. lightw. technol.*, vol. 29, no. 20, pp. 3070–3082, Oct. 2011.
[13] E. Ip and J. M. Kahn, "Digital equalization of chromatic dispersion and polarization mode dispersion," *J. Lightw. Technol.*, vol. 25, no. 8, pp. 2033–2043, Aug. 2007.
[14] A. M. Ragheb, M. Shoaib, S. Alshebeili, and H. Fathallah, "Enhanced blind equalization for optical DP-QAM in finite precision hardware," *IEEE Photon. Technol. Lett.*, vol. 27, no. 2, pp. 181–184, Jan. 2015.
[15] J. Zhang, J. Yu, N. Chi, and H. C. Chien, "Time-domain digital pre-equalization for band-limited signals based on receiver-side adaptive equalizers," *Opt. Exp.*, vol. 22, no. 17, pp. 20515–20529, 2014.
[16] F. P. Guiomar *et al.*, "Fully blind linear and nonlinear equalization for 100G PM-64QAM optical systems," *J. Lightw. Technol.*, vol. 33, no. 7, pp. 1265–1274, Apr. 2015.
[17] M. S. Faruk and S. J. Savory, "Digital signal processing for coherent transceivers employing multilevel formats," *J. Lightw. Technol.*, vol. 35, no. 5, pp. 1125–1141, Mar. 2017.
[18] C. S. Martins *et al.*, "Distributive FIR-based chromatic dispersion equalization for coherent receivers," *J. Lightw. Technol.*, vol. 34, no. 21, pp. 5023–5032, Nov. 2016.
[19] K. Kikuchi, "Fundamentals of coherent optical fiber communications," *J. Lightw. Technol.*, vol. 34, no. 1, pp. 157–179, Jan. 2016.
[20] J. C. Principe, *Information-Theoretic Learning*. New York, NY, USA: Springer, 2010.
[21] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Statist.*, vol. 33, no. 3, pp. 1065–1076, 1961.
[22] A. Renyi, "On measures of information and entropy," *Maximum-Entropy Bayesian Methods Sci. Eng.*, vol. 1, no. 2, pp. 547–561, 1961.
[23] K. Dehnad, "Density estimation for statistics and data analysis," *Technometrics* vol. 29, no. 4, pp. 296–297, 1986.
[24] E. Cox, "The fuzzy systems handbook: A practitioner's guide to building, using, and maintaining fuzzy systems," *SIAM Rev.*, vol. 37, no. 2, pp. 91–96, 1995.
[25] S. Chen, B. L. Luk, C. J. Harris, and L. Hanzo, "Fuzzy-logic tuned constant modulus algorithm and soft decision-directed scheme for blind equalisation," *Digital Signal Process.*, vol. 20, no. 3, pp. 846–859, 2010.
[26] A. ozen, I. Kaya, and B. Soysal, "Variable step-size constant modulus algorithm employing fuzzy logic controller," *Wireless Pers. Commun.*, vol. 54, no. 2, pp. 237–250, 2010.