# A RADIAL BASIS FUNCTION EQUALIZER FOR OPTICAL FIBER COMMUNICATIONS SYSTEMS

*Wei Wang*

University of Maryland Baltimore County
Department of Computer Science and Electrical Engineering
1000 Hilltop Circle, Baltimore, MD 21250

## ABSTRACT

A radial basis function (RBF) equalizer is introduced for mitigation of intersymbol interference in optical communications systems. It is shown that prior information on the noise and channel characteristics can be effectively incorporated into the structure of an RBF equalizer. A training algorithm for tracking time varying statistics of the input is presented and the proposed equalizer is applied for mitigation of polarization mode dispersion in optical communications channel with dominating amplified spontaneous emission noise.

## 1. INTRODUCTION

Nonlinear channel equalization has become the subject of research interest during the past few years. The application of neural network techniques has resulted in considerable advancement. Multilayer perceptron (MLP) and radial basis function (RBF) equalizers have shown significant performance gain over conventional transversal and decision feedback equalizers in nonlinear channel equalization due to their nonlinear structures [1, 2, 3]. The price paid for the performance improvement, however, is an increase in complexity and long training time, which are the major limiting factors of applications of neural networks for high speed transmission systems.

The complexity of RBF equalizer can be substantially reduced by incorporating prior information about channel characteristics. RBF equalizer is a linear combination of basis functions. This structure is closely connected with the Bayesian method [4], which is the optimal solution that achieves the minimum decision error probability. The close relationship with Bayesian approach provides valuable insights on how to design the RBF equalizer by taking into account the physical properties of transmission channels.

In this paper, a radial basis function (RBF) equalizer is presented to mitigate polarization mode dispersion (PMD) for optical fiber communications systems. The proposed equalizer can effectively adapt to the characteristics of the optical channel, which is nonlinear, time-varying and corrupted by non-Gaussian and signal dependent noises. we derive a recursive learning algorithm to track channel changes and design the RBF equalizer by incorporating the prior information about the channel distortion. Simulation results are presented to demonstrate its successful application.

## 2. OPTICAL COMMUNICATIONS CHANNEL

The optical fiber communications channel is a time varying nonlinear system. Polarization mode dispersion (PMD), the primary source of inter-symbol (ISI), is well known to severely impair the signal quality in high bit rate long haul optical fiber communications systems. PMD is caused when the light polarized in one axis travels faster than light polarized in the orthogonal axis because of the birefringence of optical fiber. The gap between the arrival times of the two components, defined as the differential group delay (DGD), $\tau$, leads to signal pulse broadening, hence ISI. PMD can be characterized by the polarization dispersion vector, $\Omega$, whose direction determines the two principle states of polarization and whose magnitude is equal to $\tau$ [5]. The birefringence of optical fiber results from intrinsic factors, such as geometric irregularities of the fiber core or internal stresses, or external factors, such as bending, twisting and environment temperature changing. Since all these mechanisms exist to some extent in any field-installed fiber, birefringence varies randomly along its length, which leads to time varying optical channels.

The first order PMD effect can be characterized by the channel response

$$h(t) = \gamma\delta(t) + (1-\gamma)\delta(t-\tau) \tag{1}$$

where $\gamma$ is the power splitting factor representing the ratio of signal strengths in the two principal states of polarization, which is uniformly distributed in $[0, 1]$. The effect of high order PMD, which dominates for large DGD values, is frequency dependent, hence cannot be modeled as a linear system as in Eq. (1). Even for the first order PMD, however, the linearity of the channel is destroyed by the existence of the photodetector in the receiver, which can be modeled as a square law device, that converts optical signal power to electric current.

In optical fiber transmission systems with optical amplifiers, the amplified spontaneous emission (ASE) is the dominant source of noises that leads to asymmetric distributions of marks and spaces after passing through the photodetector. The probability density function of the detected signal $I$ is a function of energy $E$ of the transmitted signal and the power spectral density $N_0$ of the ASE noise. The received marks and spaces have different pdfs that are approximated as in [6], reproduced as follows,

$$p_1(I) = \frac{1}{N_0}(\frac{I}{E})^{M-1/2}\exp(-\frac{I+E}{N_0})I_{M-1}(2\frac{\sqrt{IE}}{N_0}) \tag{2}$$

$$p_0(I) = \frac{1}{N_0} \frac{(I/N_0)^{M-1} \exp(-I/N_0)}{(M-1)!} \tag{3}$$

where $M = B_0/B_e$ is the number of modes per polarization state in the received optical spectrum, $B_0$ and $B_e$ are the optical bandwidth and the electrical bandwidth of the system at the detector, respectively, and $I_n$ denotes the $n$th modified Bessel function of the first kind. The means and variances of the received marks and spaces, $\mu_1$, $\sigma_1^2$, $\mu_0$ and $\sigma_0^2$, can be obtained from Eq. (2) and (3) as $\mu_1 = MN_0+E$, $\sigma_1^2 = MN_0^2 + 2EN_0$, $\mu_0 = MN_0$, $\sigma_0^2 = MN_0^2$. The two terms of $\sigma_1^2$ are often referred to as "noise/noise beat" and "signal/noise beat", respectively. We can see that the marks have a noncentral chi-square distribution, and the spaces have a central chi-square distribution.

For the purpose of simplicity, however, Gaussian distributions with the same mean and variance are commonly used as

$$p_1(I) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp(\frac{-(I-\mu_1)^2}{2\sigma_1^2}) \tag{4}$$

$$p_0(I) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp(\frac{-(I-\mu_0)^2}{2\sigma_0^2}). \tag{5}$$

Note that $I$ is a sum of $2M$ independent random variables, hence it is reasonable to assume Gaussian pdf from the central limit theorem for large $M$.

## 3. CONSTRUCTION OF RBF EQUALIZER

Adaptive electrical equalization has shown to be an effective technique to mitigate ISI due to PMD in optical communications systems [7, 8]. Most equalizers are based on Wiener filters, which give the least mean-squared error (MSE) of equalizer output. These equalizers assume a stationary linear channel model and achieve optimum only when signals are corrupted by additive Gaussian noise [9]. As discussed in the above section, all these assumptions do not hold in the optical channel. In addition, Wiener filters are usually trained by a supervised method, which implies that "desired data" is required in the learning process. However, this is not practical in optical communications systems due to the nonzero mean "noise/noise beat" and signal dependent "signal/noise beat". Without the exact knowledge of noises, one can not determine what the "desired data" is. Also, the fact that the transmitted sequence is drawn from a finite alphabet, {0,1}, is not exploited in these equalizers. The RBF equalizer, on the other hand, is advantageous with regards to all the above mentioned points.

The output of the RBF network is a linear combination of basis functions,

$$f(\mathbf{y}) = \sum_{i}^{K} w_i \phi_i(\|\mathbf{y} - \mathbf{c}_i\|), \tag{6}$$

where $\mathbf{y}$ is the equalizer input and also the channel output, $\mathbf{c}_i$'s are the centers and $K$ is the number of basis functions. The structure of RBF network holds the exact frame for the Bayesian approach, which is represented as

$$E[\mathbf{x}|\mathbf{y}] = \frac{\sum_{i=1}^{2^N} \mathbf{x}_i p(\mathbf{y}|\mathbf{x}_i) p(\mathbf{x}_i)}{\sum_{i=1}^{K} p(\mathbf{y}|\mathbf{x}_i) p(\mathbf{x}_i)}, \tag{7}$$

where $\mathbf{x}_i$ denotes the transmitted sequences, i.e., the channel input, of length $N$, $\mathbf{x}_i = [x_1, x_2, \ldots, x_N]$, $x_j \in \{0, 1\}$.

We can see that the basis function, $\phi_i$, in Eq. (6) corresponds to the normalized conditional distribution of channel output, $p(\mathbf{y}|\mathbf{x}_i)/\sum_{i=1}^{K} p(\mathbf{y}|\mathbf{x}_i)$, in Eq. (7) since all the possible transmitted sequences are equally likely. The centers $\mathbf{c}_i$'s are related to the noise-free channel outputs of different transmitted sequences. The outer layer weight $w_i$ corresponds to the transmitted sequence $\mathbf{x}_i$. Once the parameters of RBF equalizer are trained to play their corresponding physical roles, we can expect that it achieves the minimum decision error probability.

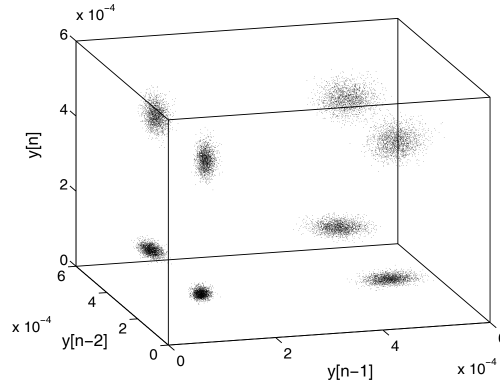


**Fig. 1**. 3D constellation of PMD channel. The data is from the simulation of a first order PMD channel with parameters given by: return-to-zero (RZ) Gaussian pulse, Peak power=2mW, Bit duration=100ps, Full width at half maximum (FWHM)=50ps, DGD=50ps, $\gamma$=0.5 and the signal to noise ratio (SNR)=5.7dB.

Our goal is to construct the RBF equalizer to match the optical communications channel. It can be seen in Fig. 1, the channel outputs are almost fully separated in a 3D constellation, which implies that ISI basically stays within 3 bits. This is true because the effect of PMD is negligible beyond its adjacent bits. Hence the optimum choice for the length of the equalizer input vector, $N$, is 3, and as a result, $2^3 = 8$ basis functions are needed for the RBF equalizer.

Note that the noise distribution at each cluster is different due to the asymmetric pdfs of the output of optical channel, which suggests that multivariate Gaussian functions are good candidates for the basis functions.

Hence we construct the RBF equalizer for optical PMD channel as

$$\begin{aligned} f(\mathbf{y}(n)) &= E[x(n-1)|\mathbf{y}(n)] = \sum_{i=1}^{8} w_i \phi_i(\mathbf{y}(n)) && (8) \\ &= \sum_{i=1}^{8} w_i \frac{g_i(\mathbf{y}(n))}{\sum_{i=1}^{8} g_i(\mathbf{y}(n))}, && (9) \end{aligned}$$

where $\mathbf{y}(n) = [y(n), y(n-1), y(n-2)]^T$, and

$$g_i(\mathbf{y}) = \frac{1}{\sqrt{|\Sigma_i|}} \exp[-\frac{1}{2}(\mathbf{y} - \mathbf{c}_i)^T \Sigma_i^{-1} (\mathbf{y} - \mathbf{c}_i)], \tag{10}$$

where $\Sigma_i$ is a $3 \times 3$ diagonal covariance matrix. Note that a decision delay of 1 bit is introduced in Eq. (8) because all the PMD effect of the central bit, not the first one, is contained in the three bits.

## 4. RBF NETWORK TRAINING

Generally, the training of RBF network consists of two stages. In the first stage, the parameters governing the basis functions are estimated. The second stage involves the learning of weights of the output layer. There are different learning strategies in the design of RBF network depending on how the network is specified [10].

### 4.1. Learning algorithm

In this paper, we present an unsupervised recursive learning strategy to estimate basis function parameters. Considering the time varying property of the optical channel, the learning algorithm should be able to track varying channel response, i.e., the estimation of basis function centers should be updated with channel changes.

We derive our learning algorithm under the following conditions:

- When channel is stationary, the estimation is unbiased.
- When channel varies, the estimation is biased to the most recent data.

We estimate mean and variance for each basis function based on a data block with length $L$. When new equalizer input enters the block, the statistics are updated and part of the previous information is discarded.

The recursive learning algorithm is presented as follows:

$$\bar{y}(n) = r\bar{y}(n-1) + (1-r)y(n), \tag{11}$$

$$\sigma^2(n) = r\sigma^2(n-1) + (1-r)\alpha(n)[y(n) - \bar{y}(n)]^2, \tag{12}$$

where $r = 1 - 1/L, \alpha(n) = \frac{1+r}{r^2[(1-r)r^{2n-1}+2]}, n = 1, 2, \ldots$, and the initial estimations are given by $\bar{y}(0) = \frac{1}{L}\sum_{k=1}^{L} y(k)$ and $\sigma^2(0) = \frac{1}{L-1}\sum_{k=1}^{L}(y(k) - \bar{y}(0))^2$. The proof is shown in the appendix. The quantity $r$ can be considered as the *forgetting factor*. If channel varies rapidly, small $L$ is preferred since the statistics "forget" faster; if channel changes slowly, large $L$ should be used since it can provide more accurate estimations. Note that, when $n$ gets large, $\alpha(n)$ converges to a constant, then Eq. (12) can be approximated as

$$\sigma^2(n) = r\sigma^2(n-1) + \frac{1}{2}(\frac{1}{r^2} - 1)[y(n) - \bar{y}(n)]^2. \tag{13}$$

### 4.2. Training of basis functions

If the training sequence, $\mathbf{x}(n)$, is available, we can easily determine the cluster to which the current input vector, $\mathbf{y}(n)$, belongs, thus update the corresponding center with the learning algorithm described above. When $\mathbf{y}(n)$ corresponds to the $i$-th basis function, then let $\mathbf{y}_i(n) = \mathbf{y}(n)$. The $i$-th center is updated by

$$\mathbf{c}_i = \bar{\mathbf{y}}_i(n). \tag{14}$$

where $\bar{\mathbf{y}}_i(n) = [\bar{y}_{i,1}, \bar{y}_{i,2}, \bar{y}_{i,3}]$ with each component obtained by Eq. (11).

As observed in Fig. 1, the variances of different basis functions are highly related since they are all extended from the same 1-D asymmetric Gaussian distributions. Hence instead of $8 \times 3$ variances, we only need two variances, as in Eq. (4) and (5), based on which the covariance matrix for each basis function can be constructed. For example, the covariance matrix of the basis function corresponding to input vector $\mathbf{x}_i = [101]$ is given by

$$\Sigma_i = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_0^2 & 0 \\ 0 & 0 & \sigma_1^2 \end{pmatrix}. \tag{15}$$

Thus the number of free parameters that need to be estimated is reduced, and so is the complexity of RBF network training.

Note that the availability of training sequence is not mandatory since "desired outputs" are not required in our learning algorithm. When transmitted signals can be detected with low decision error probability, which is the case in optical channel, we can use detected sequence for basis function training.

### 4.3. Training of output layer weights

The weights of the output layer are usually trained by a supervised process. Stochastic-gradient algorithm such as Least mean square (LMS) are used in most applications as

$$w_i(n+1) = w_i(n) + \mu_w e(n)\phi_i(\mathbf{y}(n)), \tag{16}$$

where $\mu_w$ is the step size, and the $e(n)$ is the error of the network output, $e(n) = d(n) - f(\mathbf{y}(n))$, where $d(n)$ is the "desired" output.

However, considering the relationship with Bayesian model as discussed section 3, we can avoid the weight training process by simply assigning the corresponding transmitted symbol to the weight of each basis function, i.e., $w_i = x_{i,2}$. Here $\mathbf{x}_i = [x_{i,1}, x_{i,2}, x_{i,3}]$ is the channel input vector that corresponds the $i$-th basis function.

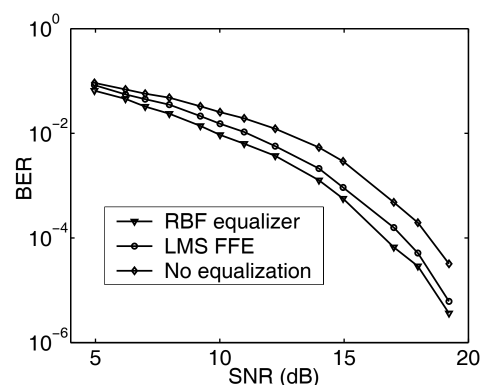


**Fig. 2**. BERs of FFE and RBF equalizer.

## 5. SIMULATION RESULTS

An all order PMD channel is simulated for transmission of 10Gbit/s RZ Gaussian pulses with 50 ps FWHM. The mean DGD of the channel is 57 ps. We construct and train the RBF equalizer by the methods described above. For the purpose of comparison, a feed forward equalizer (FFE) of length 5 is also implemented by using LMS algorithm [9]. We evaluate the performance of the equalizer by bit error rate (BER) that it can achieve. The BERs are estimated by counting the number of errors in the transmission of a pseudo-random bit string of length 8, i.e., (11101000), for $2 \times 10^6$ times through the simulation system.

BERs of RBF equalizer and FFE under different noise levels are shown in Fig. 2. We can see that obvious gain can be obtained by the application of RBF equalizer over FFE.

## 6. CONCLUSION

An RBF equalizer is proposed for mitigation of PMD induced ISI in optical communications systems. By incorporating prior information on the noise and channel characteristics, the complexity of the RBF equalizer structure and training process can be reduced without compromising its performance. An unsupervised recursive learning algorithm is presented for tracking time varying statistics of the channel. Simulation results verify the effectiveness of proposed RBF equalizer.

## 7. ACKNOWLEDGMENTS

I thank Dr. Tulay Adali for many valuable discussions.

## 8. REFERENCES

[1] S. Siu, G. J. Gibson, and C. F. N. Cowan, "Decision feedback equalisation using neuralnetwork structures and performance comparison with the standard architecture," *Communications, Speech and Vision, IEE Proccedings Part I*, vol. 137, no. 4, pp. 221–225, 1990.

[2] J. Cid-Sueiro, A. Artes-Rodriguez, and A. R. Figueiras-Vidal, "Recurrent radial basis function networks for optimal symbol-by-symbol equalisation," *EURASIP Signal Processing*, vol. 40, pp. 53–63, 1994.

[3] I. Cha and S. A. Kassam, "Non-linear filtering and equalization in non-gaussian noise using radial basis function and related networks," *Signals, systems and computers*, vol. 2, pp. 1021–1025, 1993.

[4] B. Mulgrew, "Applying radial basis functions," *IEEE Signal Processing Magazine*, vol. 13, no. 2, pp. 50–65, 1996.

[5] G. P. Agrawal, *Fiber-Optic Communication Systems*. NY: Wiley, 2nd ed., 1997.

[6] P. A. Humblet and M. Azizoğlu, "On the bit error rate of lightwave systems with optical amplifiers," *J. Lightwave Technol.*, vol. 9, pp. 1576–1582, 1991.

[7] H. Bulow, F. Buchali, W. Maumert, R. Ballentin, and T. Wehren, "Pmd mitigation at 10gbit/s using linear and nonlinear integrated electronic equalizer circuits," *Electron. Lett.*, vol. 36, no. 2, pp. 163–164, 2000.

[8] T. Adalı, "Applications of signal processing to optical fiber communications," in *Proc. LEOS'02*, ThA1, 2002.

[9] S. Haykin, *Adaptive Filter Theory*. New Jersey: Prentice Hall, 2002.

[10] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford: Clarendon, 1995.

## 9. APPENDIX

Here we show that the recursive learning algorithm in Eq. (11) and (12) satisfies the two conditions described in the paper. From Eq. (11), we can obtain the direct form of $\bar{y}(n)$ as

$$\bar{y}(n) = r^n \bar{y}(0) + (1-r)\sum_{i=1}^{n} r^{n-i} y(i) \tag{17}$$

It is easy to see that the estimator for the mean value, Eq. (11), is unbiased, and more weight is put on recent data. To estimate variance, first, we need to evaluate the expectation value of $(y(n) - \bar{y}(n))^2$. For simplicity, we derive from $E[(y(n+1) - \bar{y}(n+1))^2]$.

$$\bar{y}(n+1) - y(n+1) = r(\bar{y}(n) - y(n+1)) \tag{18}$$

thus

$$E[(\bar{y}(n+1) - y(n+1))^2] = r^2[E[\bar{y}(n)^2] - \mu^2 + \sigma^2] \tag{19}$$

where $\mu$ and $\sigma^2$ are the true values of mean and variance, and

$$\begin{aligned} E[\bar{y}(n)^2] &= r^{2n} E[\bar{y}(0)^2] + 2r^n(1-r)E[\bar{y}(0)\sum_{i=1}^{n} r^{n-i} y(i)] \\ &+ (1-r)^2 E[(\sum_{i=1}^{n} r^{n-i} y(i))^2] \end{aligned} \tag{20}$$

where

$$\begin{aligned} E[\bar{y}(0)^2] &= (1-r)\sigma^2 + \mu^2 \\ E[\bar{y}(0)\sum_{i=1}^{n} r^{n-i} y(i)] &= \frac{1-r^n}{1-r}\mu^2 \\ E[(\sum_{i=1}^{n} r^{n-i} y(i))^2] &= \frac{(1-r^n)^2}{(1-r)^2}\mu^2 + \frac{1-r^{2n}}{1-r^2}\sigma^2. \end{aligned}$$

Hence,

$$E[(\bar{y}(n+1) - y(n+1))^2] = r^2[\frac{1-r}{1+r}(1 + r^{2n+1}) + 1]\sigma^2 \tag{21}$$

then

$$E[(\bar{y}(n) - y(n))^2] = r^2[\frac{1-r}{1+r}(1 + r^{2n-1}) + 1]\sigma^2 \tag{22}$$

Substituting Eq. (22) into Eq. (12), the recursive estimator for variance is unbiased when

$$\alpha(n) = \frac{1+r}{r^2[(1-r)r^{2n-1} + 2]} \tag{23}$$

It is easy to show that $r\alpha(n) > 1$, which implies that more weight is put on recent data.