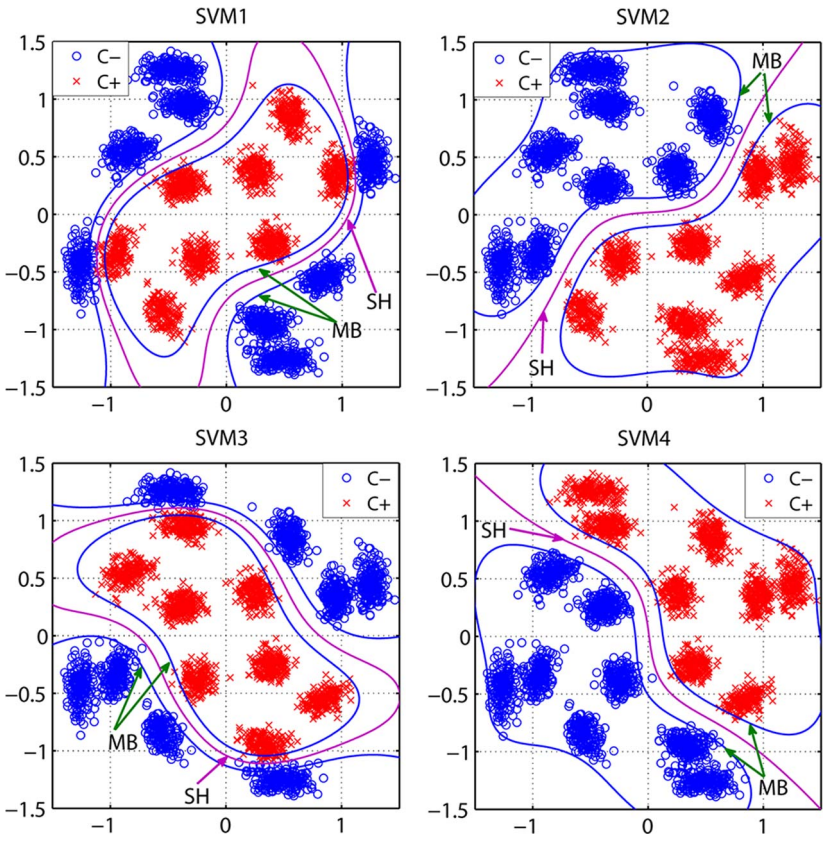
IEEE Photonics Journal
An IEEE Photonics Society Publication

Open Access

Nonparameter Nonlinear Phase Noise Mitigation by Using M-ary Support Vector Machine for Coherent Optical Systems

Volume 5, Number 6, December 2013

Minliang Li
Song Yu
Jie Yang
Zhixiao Chen
Yi Han
Wanyi Gu



DOI: 10.1109/JPHOT.2013.2287565

Nonparameter Nonlinear Phase Noise Mitigation by Using M-ary Support Vector Machine for Coherent Optical Systems

Minliang Li, Song Yu, Jie Yang, Zhixiao Chen, Yi Han, and Wanyi Gu

State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China

DOI: 10.1109/JPHOT.2013.2287565

Manuscript received August 3, 2013; revised October 7, 2013; accepted October 15, 2013. Date of publication October 28, 2013; date of current version November 1, 2013. This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2012CB315605; by the National Natural Science Foundation under Grants 61271191, 61271193, and 61072054; by the Fundamental Research Funds for the Central Universities; and by the State Key Laboratory of Information Photonics and Optical Communications (Beijing University of Posts and Telecommunications). Corresponding author: S. Yu (e-mail: yusong@bupt.edu.cn).

Abstract: The M-ary support vector machine (SVM) is introduced as a nonparameter nonlinear phase noise (NLPN) mitigation approach for the coherent optical systems. The NLPN tolerance of the system can be improved by using the M-ary SVM to conduct nonlinear detection. In this scheme, SVMs with different classification strategies are utilized to execute binary classification for signals impaired by fiber NLPN. Since the separating hyperplane of each SVM is constructed by training data, this scheme is independent from the knowledge of the transmission link. In numerical simulation, the M-ary SVM performs better than the method of amplitude-dependent phase rotation at the transmitter and receiver, as well as the maximum likelihood detection with back rotation.

Index Terms: Nonlinear phase noise, quadratic amplitude modulation (QAM), support vector machine (SVM), M-ary SVM.

1. Introduction

The fiber nonlinearities have been identified as the limiting factors for enhancing the capacity and transmission length of coherent optical system. Recently, multilevel quadratic amplitude modulation (QAM) has been applied as a promising technology to meet the growing requirement on the spectral efficiency [1]. However, in QAM system, higher launch power is required to obtain increased signal-to-noise ratio (SNR), resulting in more notable fiber nonlinear impairments [2]. Nonlinear phase noise (NLPN) is one of the major distortion factors, which is caused by the interaction between the signal and the amplified spontaneous emission (ASE) noise from inline optical amplifiers via fiber Kerr nonlinearity [3], [4]. The NLPN has attracted attention since it was firstly introduced by Gordon and Mollenauer [5].

To eliminate NLPN, many optical and electronic methods have been proposed. Compared with the optical phase conjugation schemes [6], [7], which insert hardware in transmission link, electronic methods have the advantages of being flexible and less costly. Recently, various lumped electronic processing methods have been proposed to mitigate the effect of nonlinear phase noise. The amplitude-dependent phase rotation for transmitted or received signals can reduce the NLPN variance effectively [8], [9]. Based on constructing the likelihood function, a close-form maximum likelihood-based data detection algorithm was proposed [10], as well as a maximum *a posteriori*

detector based on the Bayesian framework of factor graphs [11]. However, because of the stochastic nature and nonlinear properties of the NLPN, above approaches to suppress NLPN are all depends on the parameters of the transmission link. When the signal propagates through dynamic optical network link instead of the fixed point-to-point link, the variation of link parameters would lead these existing methods to become invalid [12], [13]. Hence, it is meaningful to investigate techniques which are independent of the parameters of the link to mitigate the NLPN.

Machine learning algorithms are efficient non-parameter methods for their properties of getting the characteristic of systems from data directly. The support vector machine (SVM) as one of powerful and widely used machine learning algorithms [14], requires a small number of parameters during establishing its model. Moreover, the priori information and heuristic assumptions are unnecessary for the SVM. In this paper, we introduce the M-ary SVM [15] as a non-parameter method to mitigate the NLPN in the signal-carrier 16-QAM coherent optical systems. By transforming the NLPN mitigation problem for the QAM signal into a hypothesis test problem, the decision for the received signal is conducted by the M-ary SVM through a series of binary classification in this work. Moreover, because the model of each SVM of the M-ary SVM is obtained by a certain amount of training data, this scheme need not any information about the transmission link. By numerical simulation, the results show that the M-ary SVM scheme can achieve better performance than the method proposed by Lau and Kahn [9].

The structure of this paper is as follows. In Section 2, we detail the mathematical model of SVM as a binary classifier, the M-ary SVM for NLPN mitigation scheme is also illustrated. The simulation scheme conducted in our study is described in the Section 3. The simulation results and discussions are presented in the Section 4. Section 5 concludes this paper.

2. Theory of M-ary SVM

Nonlinear phase noise mitigation for multilevel modulation signal by using the M-ary SVM is achieved via a number of parallel SVMs followed by a set of logical processing. Each SVM classifies every received symbol into one of the two groups by a specific decision boundary, which is obtained through a certain amount of training data. Accordingly, there is no need to know any characteristic of the transmission channel. In the following of this part, we will firstly introduce the principle of SVM, as it is the core component of the M-ary SVM. At last, a systematic description about the M-ary SVM for 16-QAM detection is given.

2.1. SVM as Classifier

Based on statistical learning theory, SVM is developed from linear classifier which aims at finding the optimal separating hyperplane that maximizes the margin (i.e., maximizes the smallest distance between the hyperplane and any of the samples) [16]. Meanwhile, the kernel method is also adopted by the SVM to solve the nonlinear separable situation [17]. For the two-class linearly separable problem in an n-dimensional feature space, the optimal separating hyperplane $f(\boldsymbol{v})$ is derived through training L vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_L$ where $\boldsymbol{v}_k \in \boldsymbol{R}^n$ and each vector corresponds to a category label $y_k \in \{-1, +1\}$. The form of $f(\boldsymbol{v})$ is given by

$$f(\boldsymbol{v}) = \omega^T \cdot \boldsymbol{v} + b = 0 \tag{1}$$

where, the vector $\omega = (\omega_1, \omega_2, \ldots \omega_n)^T$ and the scalar b are the parameters of the hyperplane. Assuming that all the training data are separated by (1) correctly, we have $y_k \cdot (\omega^T \cdot \boldsymbol{v}_k + b) > 0$ (this is because if $y_k = +1$, $f(\boldsymbol{v}_k) > 0$; and if $y_k = -1$, $f(\boldsymbol{v}_k) < 0$) for every training instance-label pair $(\boldsymbol{v}_k, y_k)$. So the Euclidean distance between vector $\boldsymbol{v}_k$ and the hyperplane is

$$\gamma_k = \frac{|\omega^T \cdot \boldsymbol{v}_k + b|}{\|\omega\|} = \frac{y_k \cdot f(\boldsymbol{v}_k)}{\|\omega\|}. \tag{2}$$

It is noted that the γ_k remains unchanged if we make the rescaling $\omega \to \lambda\omega$ and $b \to \lambda b$. Thus, we can set $y_k \cdot f(\boldsymbol{v}_k) = 1$ for the closest vector to the hyperplane. Now, the problem of obtaining the hyperplane with the maximum margin can be converted into the following equivalent quadratic programming problem [16]:

$$\underset{\omega,b}{\text{argmin}} \frac{1}{2}\|\omega\|^2 \tag{3a}$$

$$\text{subject to } y_k(\omega^T \boldsymbol{v}_k + b) \geq 1, \quad k = 1, 2, \ldots, L. \tag{3b}$$

With the help of the Lagrange multipliers $\alpha = \{\alpha_1, \alpha_1, \ldots, \alpha_L\}$ where $\alpha_k \geq 0$ $k = 1, 2, \ldots, L$ and the Karush-Kuhn-Tucker (KKT) conditions [18], optimization problem (3a) with the inequality constraints (3b) can be solved in its dual form

$$\underset{\alpha}{\text{argmax}} : \sum_{i=1}^{L} \alpha_i - \frac{1}{2} \sum_{i=1,j=1}^{L} y_i \cdot y_j \cdot \alpha_i \cdot \alpha_j \cdot \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle \tag{4a}$$

$$\text{s.t. } \alpha_i \geq 0, \quad i = 1, 2, \ldots, L \tag{4b}$$

$$\sum_{i=1}^{L} \alpha_i y_i = 0 \tag{4c}$$

where, $\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle$ is the inner products of vectors $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$. The derivation of step (3) to (4) and the detail of the KKT conditions can be found in Appendix A. After the Lagrange multipliers $\{\alpha_k\}$ are obtained, the ω and b in (1) can be readily computed [16]. Then the unknown coming data can be detected by the trained model through judging the sign of $f(\boldsymbol{v})$

$$\hat{c} = \text{sign}[f(\boldsymbol{v})] = \text{sign}\left\{ \sum_{k=1}^{L} \alpha_k \cdot y_k \cdot \langle \boldsymbol{v}, \boldsymbol{v}_k \rangle + b \right\}. \tag{5}$$

According to the KKT conditions' three properties, i.e., (a) $\alpha_k \geq 0$; (b) $1 - y_k(\omega^T \cdot \boldsymbol{v}_k + b) \leq 0$; (c) $\alpha_k[1 - y_k(\omega^T \cdot \boldsymbol{v}_k + b)] = 0$, we can find that most α_k in (5) will be zero except those whose corresponding $\boldsymbol{v}_k$ locate at the margin boundaries (they are called the support vectors). Thus, only a small amount of inner products in (5) need to be calculated when detecting new data.

If the data cannot be separated linearly by (1) in n-dimensional feature space, but can be separated in a higher dimensional space through a mapping function $\phi(\boldsymbol{v}) : \boldsymbol{R}^n \to \boldsymbol{R}^m$, then the inner products in (4a) and (5) would be substituted by $\langle \phi(\boldsymbol{v}_i), \phi(\boldsymbol{v}_j) \rangle$ and $\langle \phi(\boldsymbol{v}), \phi(\boldsymbol{v}_k) \rangle$, respectively. However, the dimensionality of mapped space is generally too high to be operated. In this case, we can search for the corresponding kernel function $\kappa(\boldsymbol{v}_i, \boldsymbol{v}_j)$, satisfying the conditions of Mercer's theorem [14], such that $\kappa(\boldsymbol{v}_i, \boldsymbol{v}_j) = \langle \phi(\boldsymbol{v}_i), \phi(\boldsymbol{v}_j) \rangle$. The inner products of higher dimensional space can be calculated by $\kappa(\boldsymbol{v}_i, \boldsymbol{v}_j)$ in the original feature space equivalently without finding the $\phi(\boldsymbol{v})$. The Gaussian radial basis function

$$\kappa(\boldsymbol{v}_i, \boldsymbol{v}_j) = \exp\left(-\frac{\|\boldsymbol{v}_i - \boldsymbol{v}_j\|^2}{2\sigma^2} \right) \tag{6}$$

is a widely used kernel and will be adopted in the simulation section. The detail idea of the kernel function and Mercer's theorem are presented in Appendix B.

However, the data may overlap due to the noise, and exact separation of the training data can lead to poor generalization [16]. In this case, soft margin method which allows some of the training data to be misclassified is efficient to control the trade-off between minimizing training errors and

model complexity. This method is achieved by introducing the slack variables ξ_k for each training vector and a penalty factor C. If the training vector $\boldsymbol{v}_k$ is on or within the correct margin boundary, then $\xi_k = 0$, otherwise $\xi_k = |y_n - f(\boldsymbol{v}_k)|$. Furthermore, if $\boldsymbol{v}_k$ lie in the region between hyperplane and correct margin, then $0 < \xi_k \leq 1$; and if v_k is misclassified to the opposite side of the hyperplane, then $\xi_k > 1$. Hence, (3a) and (3b) can be expressed as follows [16]:

$$\underset{\omega,b}{\text{argmin}} \frac{1}{2}\|\omega\|^2 + C\sum_{k=1}^{L}\xi_k \tag{7a}$$

$$\text{s.t.} \quad y_k(\omega^T \boldsymbol{v}_k + b) \geq 1 - \xi_k \tag{7b}$$

$$\xi_k \geq 0, \qquad k = 1, 2, \ldots, L \tag{7c}$$

where the parameter $C > 0$ controls the trade-off between the slack variable penalty and the margin. As a result, the dual form now becomes [16]

$$\underset{\alpha}{\text{argmax}} : \sum_{i=1}^{L}\alpha_i - \frac{1}{2}\sum_{i=1,j=1}^{L} y_i \cdot y_j \cdot \alpha_i \cdot \alpha_j \cdot \kappa(\boldsymbol{v}_i, \boldsymbol{v}_j) \tag{8a}$$

$$\text{s.t.} \quad 0 \leq \alpha_i \leq C, \qquad i = 1, 2, \ldots, L \tag{8b}$$

$$\sum_{i=1}^{L}\alpha_i y_i = 0. \tag{8c}$$

In this case, the decisions of new data are also performed by (5), and the support vectors with $\alpha_k > 0$ consist of the data which satisfy $y_k(\omega^T \boldsymbol{v}_k + b) = 1 - \xi_k$ (this again follows the KKT conditions).

2.2. M-ary SVM for 16-QAM Detection

The support vector machine is fundamentally a two-class classifier. When the signals have to be divided into multiple classes (i.e., the number of categories $K > 2$, for example, the 16-QAM detection in which each constellation cluster should be considered as one class), it is necessary to build a multiclass classifier by combining multiple two-class SVMs. One commonly used multiple classifier method is the "one-against-all" approach [16]. In this scheme, the number of SVMs is the same as the number of the classes. Each SVM is trained using the data belong to a exact class c_k as the positive instances, as well as the data belong to the rest of $K - 1$ classes as the negative instances. Another useful scheme for multiclass problem is called "one-against-one" [19], in which $K(K-1)/2$ different two-classes models are trained on all possible pairs of classes, and then to detect the data according to which class that has the highest number of "vote". However, the above approaches usually require a large number of classifiers and an additional processing step to resolve ambiguities for the overlapping region.

The M-ary SVM used for multiple hypothesis test problem proposed in [15] only needs $\lceil \log_2 K \rceil$ SVMs for K-classes problem. In this scheme, each signal's class is labeled in binary format with each bit modeled by a conventional SVM. In this paper, we focus on this method and use it as a detector in the coherent optical 16-QAM system, in which the signals are aggravated by fiber nonlinearities and ASE noise when they propagate through the fiber link. Fig. 1 illustrates the processing structure. The core unit of the M-ary SVM detector is the binary support vector machines array. Each SVM conducts binary classification with soft margin under a given classification strategy for the received signal. The whole procedure consisted of the training stage and the testing stage. Before training or testing stage, the in-phase and quadrature of each QAM symbol are used to construct the feature vector. The main purpose of the training stage is to establish the separating hyperplane of each binary SVM. First, every symbol in the training set is assigned one of the 16 classes. Furthermore, we assign the class for each constellation point

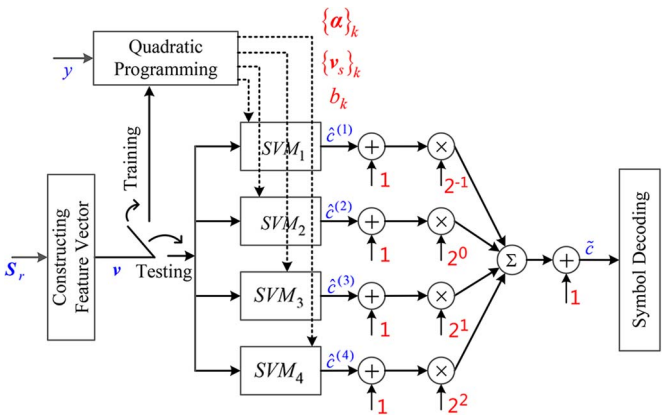


Fig. 1. Processing structure of M-ary support vector machine used for 16-QAM coherent optical signal detection with fiber nonlinear phase noise.

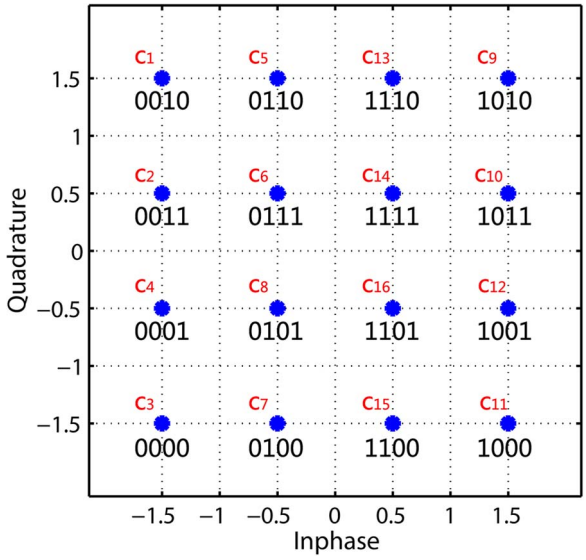


Fig. 2. Coding and class assignment strategy for each constellation point in 16-QAM systems.

according to Fig. 2 to make the separating hyperplane as simple as possible, and the Gray code is also employed to improve the reliability of the system. Although the received signals are distorted by NLPN, the category information of each training data can also be identified through the class label stored in the receiver. Then, the binary classification strategy of the i th SVM for the training set is given by [15]

$$C_i^+ = \left\{ n \in S | \bmod \left(\left\lfloor (n-1) \cdot 2^{-(i-1)} \right\rfloor, 2 \right) = 0 \right\} \tag{9a}$$

$$C_i^- = \left\{ n \in S | \bmod \left(\left\lfloor (n-1) \cdot 2^{-(i-1)} \right\rfloor, 2 \right) \neq 0 \right\} \tag{9b}$$

where $S = \{1, 2, \ldots, 16\}$ is the set of classes assigned to modulated symbols, n is the class label of the training data, C_i^+ and C_i^- are the positive and negative class sets of the i th SVM, and satisfy the conditions $C_i^+ \cap C_i^- =$, $C_i^+ \cup C_i^- = S$. Finally, the separating hyperplane of each SVM is obtained by quadratic programming (QP) [14] after relabeling the training data belong to C_i^+ and C_i^- as $+1$ and -1,

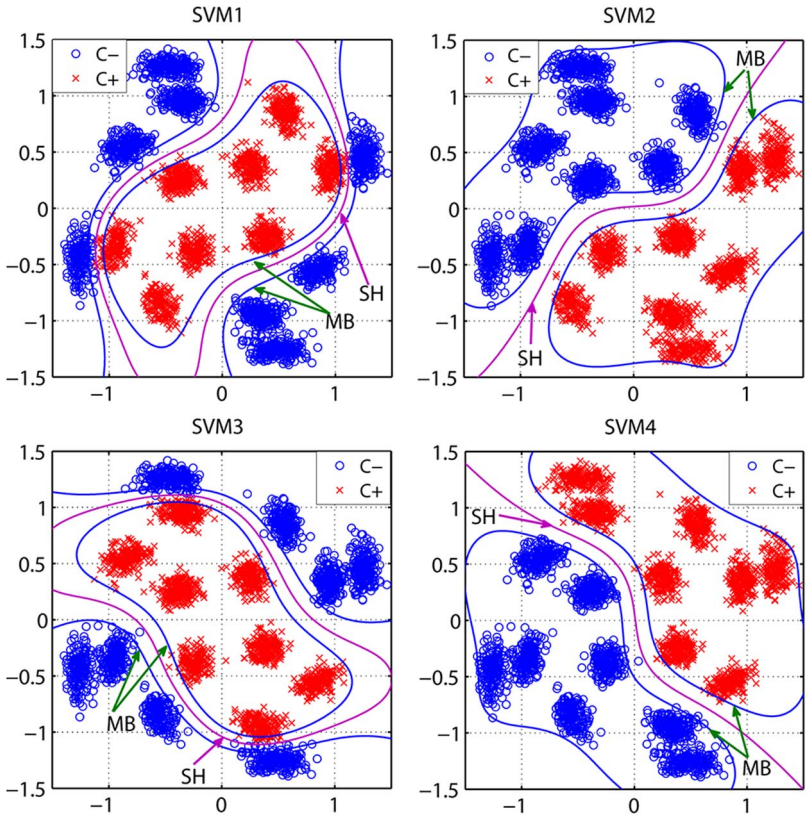


Fig. 3. M-ary SVM detection example for 16-QAM signal dominated by nonlinear phase noise (NLPN) with 1600 km transmission link and 0 dBm launch power. (a) classification result by SVM1; (b) classification result by SVM2; (c) classification result by SVM3; (d) classification result by SVM4.

respectively. When the separating hyperplane of each SVM is established, the *i*th SVM classifies the unknown symbols in the testing stage according to the discriminant function

$$\hat{c}^{(i)} = \text{sign}\left\{\sum_{k\in V} \alpha_k^{(i)} \cdot y_k^{(i)} \cdot \langle \boldsymbol{v}, \boldsymbol{v}_k^{(i)} \rangle + b_i\right\} \tag{10}$$

where V is the set include the support vectors of i th SVM. Before decoding the symbols, the specific category for the symbol is calculated by

$$\tilde{c} = 1 + \sum_{i=1}^{4} \left(\hat{c}^{(i)} + 1\right) \cdot 2^{i-2}. \tag{11}$$

The operation is as shown in Fig. 1 and the result $\tilde{c}$ belongs to the set S. A detection example using the M-ary SVM for 16-QAM coherent optical system dominated by nonlinear phase noise with 1600 km transmission link and 0 dBm launch power is showed in Fig. 3. In each subplot, crosses represent the data whose corresponding class label is $+1$, and the label -1 is marked by circles. Moreover, the separating hyperplane (SH) and margin boundaries (MB) of each binary classifier in Fig. 3, are obtained through the training sequence according to the mechanism of SVM. For example, if c_{10} is classify correctly by all four SVMs, the $\hat{c}^{(i)}$ obtained by (10) are $\{\hat{c}^1 = +1, \hat{c}^2 = -1, \hat{c}^3 = -1, \hat{c}^4 = +1\}$. Consequently, the final detection results $\tilde{c} = 10 = c_{10}$. Finally, the symbol is decoded with the coding mapping in Fig. 2.

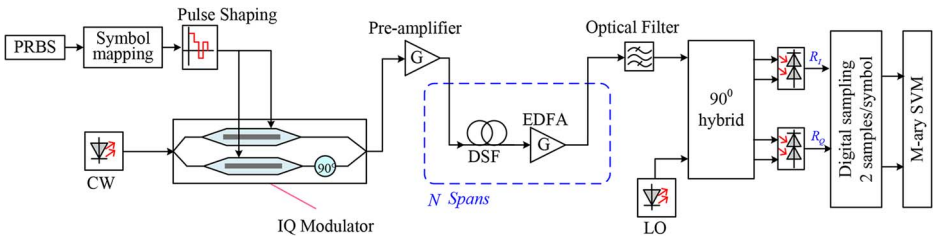


Fig. 4. Block diagram of 112-Gb/s 16-QAM single-polarization coherent optical simulation system. CW: Continuous Wave; LO: Local Oscillator; DSP: digital signal processor.

According to the above analysis, the M-ary SVM converts the nonlinear mitigation problem into a detection one without considering the nonlinear characteristics of the channel. Therefore, this method is applicable to different transmission links. The complexity of each SVM just depends on fewer support vectors, and the number of SVMs required is smaller than other multiple classifiers.

3. Simulation System Description

In order to investigate the performance of the M-ary SVM for fiber nonlinear phase noise, numerical simulation for a 112-Gb/s 16-QAM single-polarization system has been set up (Fig. 4 shows the schematic) by VPItransmissionMaker 8.6 and MATLAB. The transmitter laser is a continuous wave with its wavelength fixed at 1550 nm. The IQ modulator is composed of two Mach–Zehnder modulators (MZM) and an additional 90° phase shift that is attached to one branch. The modulating signals are generated by electrical multilevel pulses shaping filter whose raising time equals 1/8 of the symbol duration. The pulse shaping filter converts the in-phase and quadrature information of 16 QAM symbols into electrical signals with different voltage values. In our simulation, the pseudo-random binary sequence (PRBS) with length of 2^{16} is adopted, which is then mapped to 16-QAM symbols according to the strategy illustrated in Fig. 2. The modulated optical signals are coupled into the fiber after they are amplified by the pre-amplifier.

Following the model of [8]–[11], in our simulation, the transmission link consists of $N \times 80$ km dispersion-shifted fiber (DSF) spans. The fiber loss coefficient is 0.2 dB/km, corresponding to a 16 dB loss of each span. The fiber chromatic dispersion is neglected. The fiber nonlinear coefficient is $\gamma = 1.3\ W^{-1}km^{-1}$. An erbium-doped fiber amplifier (EDFA) is placed in the end of each span to compensate the fiber's loss, and the noise figure of the EDFA is set to be 6 dB.

At the receiving end, a Gaussian optical filter with 30 GHz bandwidth is firstly used to reduce the ASE noise. Then the signal is mixed with the local oscillator via the optical 90° hybrid. The local oscillator, as well as transmitted laser, is assumed to have zero linewidth. The in-phase and quadrature components of the electric signals are down sampled to one sample per symbol and detected by the M-ary SVM. For each simulation, 1000 symbols are used for training data to get the separating model of each SVM (i.e., the optimal hyperplane of each binary SVM) before the M-ary SVM acts as the detector for nonlinear signals. In the testing stage, the BER is counted to evaluate the systems performance. For each symbol, the in-phase and quadrature tributaries are used as the elements of the feature vectors. In our simulation, the LibSVM which is a widely used SVM library [20] is adopted to construct the model of SVMs in the M-ary SVM. Throughout the simulations, we focus on the impacts of ASE noise and fiber nonlinearities, while other effects such as the laser frequency drifting and the effect of polarization mode dispersion (PMD) are ignored.

4. Results and Discussions

To evaluate the nonlinear mitigation performance of the M-ary SVM for coherent optical 16-QAM systems, the Gaussian radial basis function in (6) is employed as kernel of each SVM. The optimal value of the penalty factor C in (7a) and the kernel parameter σ in (6) can be obtained with the cross validation method [21]. In our simulation, they are $C = 3$ and $\sigma = 2$. Moreover, we compare the M-ary SVM scheme with amplitude-dependent phase rotations at the transmitter and the receiver (we

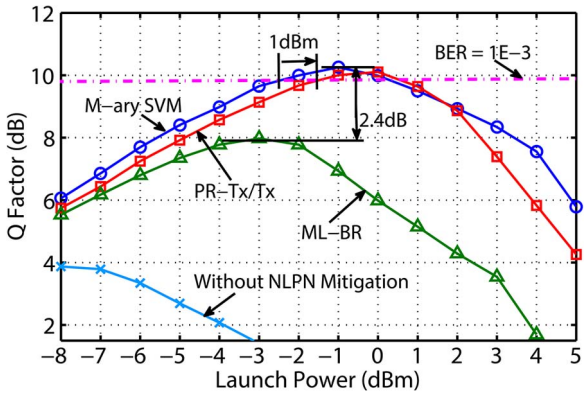


Fig. 5. Q factor versus optical launch power for 16-QAM systems with a transmission length of 20 spans (1600 km).

denote it as PR-Tx/Rx) [9], and the maximum likelihood with back rotation (this method is also introduced as reference scheme by [10], [11], and we denote it as ML-BR). According to the technique of PR-Tx/Rx scheme [9], the transmitted signal phase is firstly prerotated by

$$\bar{\theta}_{NL} - E_p\left[\frac{\gamma L P_{rec}}{2}\right] = \bar{\theta}_{NL} - \frac{\gamma L(P+\sigma^2)}{2} \tag{12}$$

where, $\bar{\theta}_{NL}$ is the mean nonlinear phase shift of 16 signal clusters, $E_p[\cdot]$ denotes the expectation given a transmitted signal power P, γ is the nonlinear coefficient of the fiber, L is the transmission distance and σ^2 is the variance of the ASE noise of the inline amplifiers. Then the received signal with power P_{rec} is rotated by an angle $\phi = -\gamma L P_{rec}/2$ at the receiver of the PR-Tx/Rx scheme. For the ML-BR method, the detector only considers the deterministic channel effects, and the received signal symbols are determined by

$$\hat{\boldsymbol{a}}(\hat{\boldsymbol{r}}_n) = \underset{\boldsymbol{a}\in\Omega^2}{\operatorname{argmin}} \left\| \hat{\boldsymbol{r}}_n - \boldsymbol{a}\cdot\exp\left(i\gamma L_{eff} N_{span}\|\boldsymbol{a}\|^2\right)\right\|^2. \tag{13}$$

In (13), $\boldsymbol{a}$ is the two-dimensional complex data vector corresponds to one of constellation points of the transmitted signal. L_{eff} is the effective length of the fiber span and N_{span} denotes the number of fiber spans. Furthermore, the situation in which the NLPN mitigation is neglected is also considered in our simulation. For this case, the received symbols are directly decided by

$$\hat{\boldsymbol{a}}(\hat{\boldsymbol{r}}_n) = \underset{\boldsymbol{a}\in\Omega^2}{\operatorname{argmin}} \|\hat{\boldsymbol{r}}_n - \boldsymbol{a}\|^2. \tag{14}$$

The received 16-QAM signals with different launch power after 1600 km fiber transmission are processed by the M-ary SVM, PR-Tx/Rx, ML-BR and direct decision, respectively. Then the Q factors are evaluated for each scheme. Fig. 5 shows the Q factor as a function of optical launch power. We can find the performance is poor when the NLPN is not mitigated at the receiver. For all of the other NLPN mitigation methods, with the increase of the launch power, the optical signal-to-noise ratio (OSNR) of the receiver becomes higher, corresponding to a gradually decrease effect of the ASE noise. When the launch power exceeds the optimal values (-3 dBm for the ML-BR, -1 dBm for the M-ary SVM and 0 dBm for the PR-Tx/Rx), the Q factors reduce again, resulting in remarkable nonlinearities. It can also be observed that a much better performance can be achieved by employing the M-ary SVM and PR-Tx/Rx than the ML-BR. The optimal Q factor of M-ary SVM exceeds that of ML-BR by 2.4 dB. The launch power dynamic range (LPDR), which represents the difference

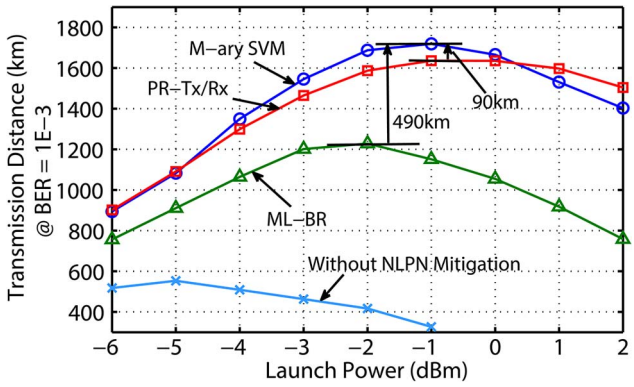


Fig. 6. Transmission distance versus optical launch power at a BER of 1×10^{-3} for 16-QAM systems.

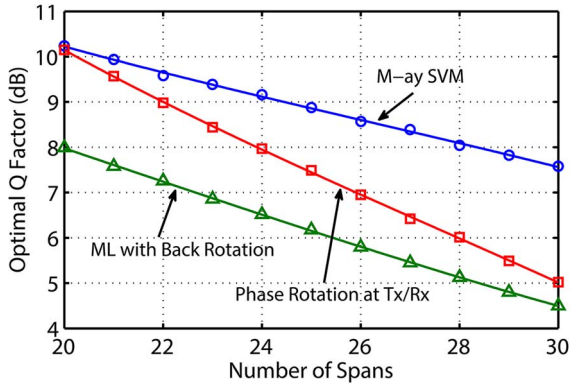


Fig. 7. Optimal Q factor versus the number of fiber spans.

between the two power values corresponding to the points where Q factors is equal to 9.8 dB (i.e., the BER is 1×10^{-3}), is employed to evaluate the performance of the M-ary SVM and the PR-Tx/Rx. The LPDRs for the M-ary SVM and the PR-Tx/Rx are approximately 2.5 dBm and 1.5 dBm, respectively. Hence, the M-ary SVM has 1 dBm more LPDR than that of the PR-Tx/Rx.

The maximum transmission distance corresponding to a BER of 1×10^{-3} is determined for a given launch power, and the results for above three methods are depicted in Fig. 6. When the NLPN mitigation is not considered, the maximum transmission distance can only reach 550 km at -5 dBm launch power. The system can take larger launch power and reach longer distance when either the M-ary SVM or the PR-Tx/Rx is employed to mitigate the NLPN. The maximum transmission distances of the systems for the M-ary SVM and PR-Tx/Rx are achieved at the same launch power, which equals to -1 dBm, while that for the ML-BR scheme is -2 dBm. The achievable maximum transmission distance of the M-ary is 490 km and 90 km longer than those of the ML-BR and PR-Tx/Rx, respectively. For further investigation, the changes of the optimal Q factors with the increasing number of fiber spans for different mitigation approaches are shown in Fig. 7. We can find the gap of the optimal Q factors between the M-ary SVM and the ML-BR is approximately equals to 2 dB, while that for the M-ary SVM and PR-Tx/Rx increases gradually. The M-ary SVM and PR-Tx/Rx almost have the same value of optimal Q factor when the number of spans is 20. The difference of the optimal Q factor for the two methods will increase to 2.6 dB when the transmission distance equals to 30 spans.

In the nonlinear phase detection stage, each SVM in Fig. 1 conducts binary classification for the testing data according to (10). As a result, the computational complexity of the M-ary SVM is relay on the number of the support vectors. In Fig. 8 we illustrate the number of SVs versus the launch

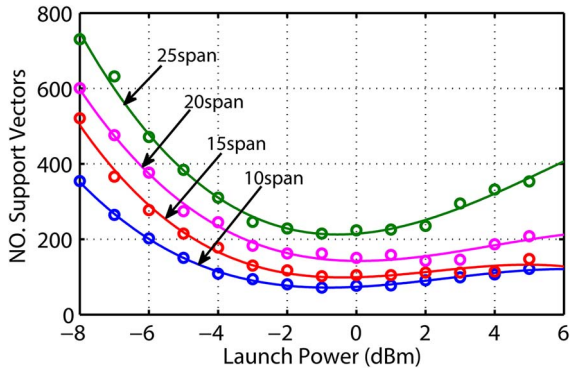


Fig. 8. Number of support vectors of the M-ary SVM versus the launch power for different transmission distance.

power for different transmission distance. At the optimal power which is consistent with Fig. 6, the number of SVs is the least, and the computational complex is the lowest. In the low power area, the number of support vectors (SV) is large resulting from ASE noise. When the launch power exceeds the optimal value, the separating hyperplanes of all SVMs become complicated and this is caused by heavy NLPN. Thus, the number of SVs increases again. On the other hand, the number of SVs will increase with the transmission length.

5. Conclusion

In this paper, the M-ary SVM is employed as a non-parameter detection scheme for 16-QAM coherent optical systems. Through transforming QAM signals distorted by nonlinear phase noise to a hypothesis test problem, the M-ary SVM method can improve the system performance without considering the specific characteristics of the fiber link. To investigate the performance of the M-ary SVM, we conducted numerical simulations for 112-Gb/s single channel 16-QAM systems in which the signals are deteriorated by nonlinearities of the fiber and the ASE noise of the in-line amplifiers. Numerical results show that compared with the method proposed by Lau and Kahn and the maximum likelihood detection with back rotation, the M-ary SVM can perform better while no information of the transmission link is required. This paper provides an idea of mitigating the NLPN by using the M-ary SVM based on numerical simulation. The performance of this scheme for the practical systems, and optical signals impaired with other factors, will be the topics that need further investigations.

Appendix I
KKT Conditions and the Derivation of Dual Form

According to [18], if $\boldsymbol{x}^* \in \boldsymbol{R}^n$ is the solution for minimizing $f(\boldsymbol{x})$ under m inequality constrains $c_i(\boldsymbol{x}) \geq 0$, $i \in m$, and the Lagrangian of this optimization problem is

$$L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) - \sum_{i \in m} \lambda_i (c_i(\boldsymbol{x})) \tag{A.1}$$

then, there exist Lagrange multipliers λ^* such that $\boldsymbol{x}^*$, λ^* satisfy the following conditions:

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}, \lambda) = 0 \tag{A.2a}$$
$$c_i(\boldsymbol{x} \geq 0), \ i \in m \tag{A.2b}$$
$$\lambda_i \geq 0, \ i \in m \tag{A.2c}$$
$$\lambda_i \cdot c_i(\boldsymbol{x}) = 0, \forall i. \tag{A.2d}$$

The conditions of (A.2a) to (A.2d) are called the KKT conditions for above optimization problem with inequality constrains. So, The Lagrangian of (3) can be defined as

$$L(\omega, b, \alpha) = \frac{1}{2}\|\omega\|^2 - \sum_{i=1}^{L} \alpha_i \cdot \left(y_i(\omega^T x_i + b) - 1\right). \tag{A.3}$$

We can find $\max L(\omega, b, \alpha) = (1/2)\|\omega\|^2$, if $\alpha_i \geq 0$ and $y_i(\omega^T x_i + b) \geq 1$, otherwise $\max L(\omega, b, \alpha) = \infty$. According to the KKT conditions, firstly, we solve by setting the derivatives with respect to ω and b equal to zero

$$\frac{\partial}{\partial \omega} L(\omega, b, \alpha) = \omega - \sum_{i=1}^{L} \alpha_i y_i x_i = 0 \tag{A.4a}$$

$$\frac{\partial}{\partial \omega} L(\omega, b, \alpha) = \sum_{i=1}^{L} \alpha_i y_i = 0. \tag{A.4b}$$

Then, using $\omega = \sum_{i=1}^{L} \alpha_i y_i x_i$ and $\sum_{i=1}^{L} \alpha_i y_i = 0$ to eliminate ω and b from the Lagrangian (A.3)

$$\begin{aligned} L(\omega, b, \alpha) &= \frac{1}{2}\omega^T \omega - \sum_{i=1}^{L} \alpha_i \cdot y_i \cdot \omega^T \cdot x_i - \sum_{i=1}^{L} \alpha_i \cdot y_i \cdot b + \sum_{i=1}^{L} \alpha_i \\ &= -\frac{1}{2}\left(\sum_{i=1}^{L} \alpha_i \cdot y_i \cdot x_i\right)^T \cdot \sum_{i=1}^{L} \alpha_i \cdot y_i \cdot x_i + \sum_{i=1}^{L} \alpha_i \\ &= \sum_{i=1}^{L} \alpha_i - \frac{1}{2} \sum_{i=1, j=1}^{L} y_i \cdot y_j \cdot \alpha_i \cdot \alpha_j \cdot \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle. \end{aligned} \tag{A.5}$$

If the α^* is obtained by resolving $\max L(\omega, b, \alpha)$, then ω can be calculated by (A.4a), b can be gotten through $\omega^T x_s + b = 0$, where x_s locates at the hyperplane.

Appendix II
The Kernel Method and Mercer's Theorem

The basic concept of the kernel method is explained by following example. We assume the data set belonging to two classes is divided by a circular in the two-dimensional space. We define $\boldsymbol{v} = (x, y)$ as the data vector, and the equation of the circle is

$$a_1 \cdot x + a_2 \cdot x^2 + a_3 \cdot y + a_4 \cdot y^2 + a_5 \cdot x \cdot y + a_6 = 0. \tag{B.1}$$

As we know, above data set cannot be separated linearly in the two-dimensional space. Hence, we can construct a mapping function according to (B.1) as $\phi(\boldsymbol{v}) = (x, x^2, y, y^2, x \cdot y)^T$. Thus (B.1) becomes linear equation after dimension mapping. For two arbitrary data $\boldsymbol{v}_1 = (x_1, y_1)$ and $\boldsymbol{v}_2 = (x_2, y_2)$, the inner product after dimension mapping is

$$\langle \phi(\boldsymbol{v}_1), \phi(\boldsymbol{v}_2) \rangle = x_1 x_2 + x_1^2 x_2^2 + y_1 y_2 + y_1^2 y_2^2 + 2 x_1 x_2 y_1 y_2. \tag{B.2}$$

On the other hand, we can note that

$$(\langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle + 1)^2 = 2 x_1 x_2 + x_1^2 x_2^2 + 2 y_1 y_2 + y_1^2 y_2^2 + 2 x_1 x_2 y_1 y_2 + 1. \tag{B.3}$$

Comparing (B.2) and (B.3), if we rescale the elements of $\phi(\boldsymbol{v})$ and add a constant dimension, namely, let $\phi(\boldsymbol{v}) = (\sqrt{2}x, x^2, \sqrt{2}y, y^2, \sqrt{2}xy, 1)^T$, then $\langle\phi(\boldsymbol{v}_1, \boldsymbol{v}_2)\rangle = (\langle \boldsymbol{v}_1, \boldsymbol{v}_2\rangle + 1)^2$ According to above analysis, the inner product of mapped high-dimension can be calculated by the kernel function $(\langle \boldsymbol{v}_1, \boldsymbol{v}_2\rangle + 1)^2$ in the original space, directly. Hence, the intensive computation in high-dimension space can be avoid.

If a function K is a valid kernel, it should satisfy the Mercer's theorem: Let $K : \boldsymbol{R}^n \longrightarrow \boldsymbol{R}$ be given. Then for K to be a valid (Mercer) kernel, it is necessary and sufficient that for any $\{x^{(1)}, \cdots x^{(m)}\}$, $(m < \infty)$, the corresponding kernel matrix $\boldsymbol{K}(x^i, x^j)$, $i, j \in m$ is symmetric positive semi-define. More information about the kernel method and the Mercer's theorem can be found in [14], [16], [17].

References

[1] P. J. Winzer, A. H. Gnauck, C. R. Doerr, M. Magarini, and L. L. Buhl, "Spectrally efficient long-haul optical networking using 112-Gb/s polarization-multiplexed 16-QAM," *J. Lightwave Technol.*, vol. 28, no. 4, pp. 547–556, Feb. 2010.

[2] S. Makovejs, D. S. Millar, D. Lavery, C. Behrens, R. I. Killey, S. J. Savory, and P. Bayvel, "Characterization of long-haul 112 GB/s PDM-QAM-16 transmission with and without digital nonlinearity compensation," *Opt. Exp.*, vol. 18, no. 12, pp. 12 939–12 947, Jun. 2010.

[3] S. Kumar, "Analysis of nonlinear phase noise in coherent fiber-optic systems based on phase shift keying," *J. Lightwave Technol.*, vol. 27, no. 21, pp. 4722–4733, Nov. 2009.

[4] N. Ekanayake and H. M. V. R. Herath, "Effect of nonlinear phase noise on the performance of *M*-ary PSK signals in optical fiber links," *J. Lightwave Technol.*, vol. 31, no. 3, pp. 447–454, Feb. 2013.

[5] J. P. Gordon and L. F. Mollenauer, "Phase noise in photonic communications systems using linear amplifiers," *Opt. Lett.*, vol. 15, no. 23, pp. 1351–1353, Dec. 1990.

[6] S. Kumar and L. Liu, "Reduction of nonlinear phase noise using optical phase conjugation in quasi-linear optical transmission systems," *Opt. Exp.*, vol. 15, no. 5, pp. 2166–2177, Mar. 2007.

[7] O. Kuzucu, Y. Okawachi, R. Salem, M. A. Foster, A. C. Turner-Foster, M. Lipson, and A. L. Gaeta, "Spectral phase conjugation via temporal imaging," *Opt. Exp.*, vol. 17, no. 22, pp. 20 605–20 614, Oct. 2009.

[8] K.-P. Ho and J. M. Kahn, "Electronic compensation technique to mitigate nonlinear phase noise," *J. Lightwave Technol.*, vol. 22, no. 3, pp. 779–783, Mar. 2004.

[9] A. P. T. Lau and J. M. Kahn, "Signal design and detection in presence of nonlinear phase noise," *J. Lightwave Technol.*, vol. 25, no. 10, pp. 3008–3016, Oct. 2007.

[10] A. S. Tan, H. Wymeersch, P. Johannisson, E. Agrell, P. Andrekson, and M. Karlsson, "An ML-based detector for optical communication in the presence of nonlinear phase noise," in *Proc. IEEE Int. Conf. Commun*, Jun. 2011, pp. 1–5.

[11] N. Jiang, Y. Gong, J. Karout, H. Wymeersch, P. Johannisson, M. Karlsson, E. Agrell, and P. Andrekson, "Stochastic backpropagation for coherent optical communications," presented at the European Conf. Optical Commun., Geneva, Switzerland, Sep. 2011, Paper We.10.P1.

[12] T. S. R. Shen, A. P. T. Lau, and C. Yu, "Simultaneous and independent multi-parameter monitoring with fault localization for DSP-based coherent communication systems," *Opt. Exp.*, vol. 18, no. 23, pp. 23 608–23 619, Nov. 2010.

[13] D. Rafique and A. D. Ellis, "Nonlinear penalties in long-haul optical networks employing dynamic transponders," *Opt. Exp.*, vol. 19, no. 10, pp. 9044–9049, May 2011.

[14] V. N. Vapnik, *The Nature of Statistical Learning Theory*. Berlin, Germany: Springer-Verlag, 1995.

[15] D. J. Sebald and J. A. Bucklew, "Support vector machines and the multiple hypothesis test problem," *IEEE. Trans. Signal Process.*, vol. 49, no. 11, pp. 2865–2872, Nov. 2001.

[16] C. M. Bishop, *Pattern Recognition and Machine Learning*. Berlin, Germany: Springer-Verlag, 2006.

[17] B. Schölkopf and A. J. Smola, *Learning With Kernels*. Cambridge, MA, USA: MIT Press, 2002.

[18] R. Fletcher, *Practical Methods of Optimization*. Hoboken, NJ, USA: Wiley, 1987.

[19] U. H.-G. Kreel, "Pairwise classification and support vector machines," in *Advances in Kernel Methods: Support Vector Learning*, B. Schölkopf, C. J. C. Burges, and A. J. Smola, Eds. Cambridge, MA, USA: MIT Press, 1999, pp. 255–268.

[20] C.-C. Chang and C.-J. Lin, "LIBSVM : A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 2, p. 27, Apr. 2011.

[21] G. Wahba, Y. Lin, and H. Zhang, "Generalized approximate cross validation for support vector machines, or, another way to look at margin-like quantities," Dept. Stat., Univ. Wisconsin-Madison, Madison, WI, USA, Tech. Rep. 1006, Apr. 1999.