Support Vector Machines for Robust Channel Estimation in OFDM

M. Julia Fernández-Getino García, *Member, IEEE*, José Luis Rojo-Álvarez, *Member, IEEE*,
Felipe Alonso-Atienza, *Student Member, IEEE*, and Manel Martínez-Ramón, *Senior Member, IEEE*

Abstract—A new support vector machine (SVM) algorithm for coherent robust demodulation in orthogonal frequency-division multiplexing (OFDM) systems is proposed. We present a complex regression SVM formulation specificall adapted to a pilots-based OFDM signal. This novel proposal provides a simpler scheme than an SVM classificatio method. The feasibility of our approach is substantiated by computer simulation results obtained for IEEE 802.16 broadband fixe wireless channel models. These experiments allow to scrutinize the performance of the OFDM-SVM system and the suitability of the ε-Huber cost function, in the presence of non-Gaussian impulse noise interfering with OFDM pilot symbols.

Index Terms—Channel estimation, complex, orthogonal frequency-division multiplexing (OFDM), robust estimation, support vector machine (SVM).

I. INTRODUCTION

ORTHOGONAL frequency-division multiplexing (OFDM) is a very attractive technique for high bit-rate transmission in wireless environments [1]. Data symbols are frequency multiplexed with orthogonal subcarriers to minimize the effects of multipath spread. Thus, a frequency-selective channel is transformed into a set of parallel flat-fading Gaussian subchannels, which is what makes equalization a simple task. Moreover, this transmission technique can be efficiently implemented via IDFT/DFT operations. Channel estimation is usually carried out based on pilot symbols using least squares (LS) [2]. However, in a practical environment where impulse noise can be present, this method may not be effective. To our knowledge, nonlinear channels with non-Gaussian noise have not been studied yet in OFDM systems.

The use of support vector machines (SVM) has already been proposed to solve a variety of digital communications problems. The decision feedback equalizer [3], [4] and adaptive multiuser detector for direct sequence code division multiple access (CDMA) signals in multipath channels [5] are addressed by means of binary SVM nonlinear classifiers. In [6], signal equalization and detection for a multicarrier (MC)-CDMA system are based on an SVM linear classification algorithm. Nonlinear channel estimation based on SVM multiregression

for multiple-input multiple-output systems has also been scrutinized [7]. In all these applications, SVM techniques outperform classical methods.

We propose a SVM robust version for channel estimation that is specifically adapted to OFDM data structure. There are two main novelties in our proposal. First, the use of complex regression in SVM formulation provides us with a simpler scheme than describing OFDM signal with either multilevel or nested binary SVM classification algorithms. Second, the adequacy of free parameters in ε-Huber robust cost function [8] is investigated in the presence of impulse noise. Although robustness of some digital communication receivers against impulse noise has been examined by using M-estimates [9], [10], there is no previous work about the performance of SVM algorithms in digital communications under this condition. For simplicity, a linear dispersive channel with non-Gaussian noise is analyzed here. The extension of the proposed linear OFDM-SVM scheme to nonlinear scenarios can be easily introduced by using Mercer's kernels in a similar way as proposed for other communication schemes [4].

In Section II, the OFDM system and impulse noise model are described. Coherent demodulation of OFDM signals with SVM is addressed in Section III. In Section IV, simulation results are presented, and conclusions are drawn in Section V.

II. OFDM MODEL

The discrete-time received OFDM signal for a system with N subcarriers is

$$r_n = \sum_{k=0}^{N-1} S_k H_k e^{j\frac{2\pi}{N}kn} + w_n + b_n g_n \tag{1}$$

where r_n $(n = 0, \ldots, N-1)$ are time-domain samples before DFT transformation, H_k is the channel's frequency response at the kth frequency, S_k is the complex symbol transmitted at the kth subcarrier, and w_n is the complex white Gaussian noise process $N(0, \sigma_w^2)$. The impulse noise is modeled as a Bernoulli–Gaussian process, i.e., the product of a real Bernoulli process b_n with $Pr(b_n = 1) = p$ and a complex Gaussian process $g_n \sim N(0, \sigma_i^2)$ [10]. Then, residual noise at the receiver side is given by the sum of both terms $z_n = w_n + b_n g_n$.

In coherent OFDM systems, pilot symbols are usually inserted for channel estimation purposes. Then, the channel's frequency response can be first estimated over a subset $\mathcal{K}_p$ of subcarriers, with cardinality $N_p = |\mathcal{K}_p|$, and then interpolated over the remaining subcarriers $(N - N_p)$ by using, for example,

DFT-based techniques with zero-padding in the time domain [2]. Now, the OFDM system can be expressed as

$$r_n = \sum_{k \in \{\mathcal{K}_p\}} P_k H_k e^{j\frac{2\pi}{N}kn} + \sum_{k \notin \{\mathcal{K}_p\}} X_k H_k e^{j\frac{2\pi}{N}kn} + z_n \quad (2)$$

where P_k and X_k are, respectively, the complex pilot or data symbol transmitted at the kth subcarrier. It is well known that if the channel impulse response has a maximum of L resolvable paths (and hence of degrees of freedom), then N_p must be at least equal to L [11].

III. OFDM-SVM COHERENT DEMODULATOR

The proposed signal model for OFDM-SVM is as follows:

$$r_n = \sum_{k \in \{\mathcal{K}_p\}} P_k H_k e^{j\frac{2\pi}{N}kn} + e_n \quad (3)$$

where $e_n = \sum_{k \notin \{\mathcal{K}_p\}} X_k H_k e^{j(2\pi/N)kn} + z_n$ contains the residual noise plus the term due to data symbols. Here, these unknown symbols carrying information will be considered as noise during the training phases. Channel estimation via LS cost function is no longer the maximum likelihood (ML) criterion when dealing with this sort of noise [12]. In order to improve the performance of the estimation algorithm, a robust cost function must be introduced.

SVM algorithms minimize a regularized cost function of the residuals. In [8], the ε-Huber cost function is used, given by

$$L(e_n) = \begin{cases} 0, & |e_n| \leq \varepsilon \\ \frac{1}{2\delta}(|e_n| - \varepsilon)^2, & \varepsilon \leq |e_n| \leq e_C \\ C(|e_n| - \varepsilon) - \frac{1}{2}\delta C^2, & |e_n| \geq e_C \end{cases} \quad (4)$$

where $e_C = \varepsilon + \delta C$, ε is the insensitive parameter, and δ and C control the trade-off between the regularization and the losses. Three different regions allow to deal with different kinds of noise: ε-insensitive zone ignores errors lower than ε; quadratic cost zone uses the L_2-norm of errors, which is appropriate for Gaussian noise; and linear cost zone limits the effect of sub-Gaussian noise. Note that (4) represents Vapnik ε-insensitive cost function when δ is small enough and Huber cost function when $\varepsilon = 0$. Here, for complex e_n, we define $L(e_n) = L(\Re\{e_n\}) + L(\Im\{e_n\})$, where $\Re\{\cdot\}$ and $\Im\{\cdot\}$ denote real and imaginary parts, respectively. The primal problem can be stated as minimizing

$$\frac{1}{2}\sum_{k \in \{\mathcal{K}_p\}} |H_k|^2 + \frac{1}{2\delta}\sum_{n \in I_1}\left(\xi_n + \xi_n^+\right)^2 + C\sum_{n \in I_2}\left(\xi_n + \xi_n^+\right)$$
$$+\frac{1}{2\delta}\sum_{n \in I_3}\left(\zeta_n + \zeta_n^+\right)^2 + C\sum_{n \in I_4}\left(\zeta_n + \zeta_n^+\right) - \frac{1}{2}\sum_{n \in I_2, I_4}\delta C^2 \quad (5)$$

constrained to

$$\Re(r_n) - \sum_{k \in \{\mathcal{K}_p\}}\Re\left(P_k H_k e^{j\frac{2\pi}{N}kn}\right) \leq \varepsilon + \xi_n \quad (6)$$

$$\Im(r_n) - \sum_{k \in \{\mathcal{K}_p\}}\Im\left(P_k H_k e^{j\frac{2\pi}{N}kn}\right) \leq \varepsilon + \zeta_n \quad (7)$$

$$-\Re(r_n) + \sum_{k \in \{\mathcal{K}_p\}}\Re\left(P_k H_k e^{j\frac{2\pi}{N}kn}\right) \leq \varepsilon + \xi_n^+ \quad (8)$$

$$-\Im(r_n) + \sum_{k \in \{\mathcal{K}_p\}}\Im\left(P_k H_k e^{j\frac{2\pi}{N}kn}\right) \leq \varepsilon + \zeta_n^+ \quad (9)$$

$$\xi_n^{(+)}, \zeta_n^{(+)} \geq 0 \quad (10)$$

for $n = 0, \ldots, N - 1$, where pairs of slack variables are introduced for both real $(\xi_n^{(+)})$ and imaginary $(\zeta_n^{(+)})$ residuals; superscript $^+$ and no superscript stand for positive and negative components of residuals, respectively; and $I_1 - I_2$ $(I_3 - I_4)$ are the set of samples for which real (imaginary) parts of the residuals are in the quadratic-linear cost zone.

The derivation of a similar dual problem can be found in [8], and only the new steps that are specific to our proposal are pointed out next. In brief, the primal-dual functional is obtained by introducing the constraints into the primal functional by means of Lagrange multipliers $\{\alpha_{R,n}\}$, $\{\alpha_{R,n}^+\}$, $\{\alpha_{I,n}\}$, $\{\alpha_{I,n}^+\}$, for the real (subscript R) and imaginary (subscript I) parts of the residuals. By making zero the primal-dual functional gradient with respect to H_k, we have the following expression for channel estimated values at pilot positions:

$$\hat{H}_k = \sum_{n=0}^{N-1} \psi_n P_k \quad (11)$$

where $\psi_n = (\alpha_{R,n} - \alpha_{R,n}^+) + j(\alpha_{I,n} - \alpha_{I,n}^+)$. For notation, we define the following column vector:

$$\mathbf{v}_n(k) = \left[P_k e^{j\frac{2\pi}{N}kn}\right], \quad k \in \{\mathcal{K}_p\} \quad (12)$$

and the following Gram matrix as $\mathbf{R}(n,m) = \mathbf{v}_n^H \mathbf{v}_m$. Now, by placing optimal solution (11) into the primal-dual functional and grouping terms, a compact form of the functional problem can be stated in vector form, which consists of maximizing

$$-\frac{1}{2}\boldsymbol{\psi}^H(\mathbf{R} + \delta\mathbf{I})\boldsymbol{\psi} + \Re(\boldsymbol{\psi}^H \mathbf{r}) - \left(\boldsymbol{\alpha}_R + \boldsymbol{\alpha}_R^+ + \boldsymbol{\alpha}_I + \boldsymbol{\alpha}_I^+\right)\mathbf{1}\varepsilon \quad (13)$$

constrained to $0 \leq \{\alpha_{R,n}\}, \{\alpha_{R,n}^+\}, \{\alpha_{I,n}\}, \{\alpha_{I,n}^+\} \leq C$, where $\boldsymbol{\psi} = [\psi_0, \ldots, \psi_{N-1}]^T$; $\mathbf{I}$, $\mathbf{1}$ are the identity matrix and the all-ones column vector, respectively; $\boldsymbol{\alpha}_R$ is the vector containing the corresponding Lagrange multipliers, with the other subsets being similarly represented; and $\mathbf{r} = [r_0, \ldots, r_{N-1}]^T$. Note that (13) is a quadratic form and, thus, real-valued, and it represents a natural extension of the dual functional in SVM real regression for complex-valued problems. The channel values at pilot positions (11) can be obtained by optimizing (13) with respect to $\{\alpha_{R,n}\}, \{\alpha_{R,n}^+\}, \{\alpha_{I,n}\}, \{\alpha_{I,n}^+\}$ and then substituting into (11).

IV. SIMULATION RESULTS

In order to test the performance of the OFDM-SVM scheme, a scenario for IEEE 802.16 fixed Broadband Wireless Access Standard has been considered [13]. In second-generation systems for this type of application, non-line-of-sight (NLOS) conditions are present. To simulate this environment, we use the Modified Stanford University Interim SUI-3 channel model for

|	Tap 1	Tap 2	Tap 3	Units
Delay	0	0.5	1	μs
Power	0	-5	-10	dB
K factor (90 %)	1	0	0	linear
Doppler	0.4	0.4	0.4	Hz

omnidirectional antennas, with $L = 3$ taps, a maximum delay spread of $\tau_{\max} = 1\,\mu$s, and maximum Doppler frequency $f_m = 0.4$ Hz [13]. The main parameters of this channel are summarized in Table I. It can be observed that the channel exhibits an rms delay spread of $\tau_{\rm rms} = 0.305\,\mu$s. Also, it must be noticed that K-factors are given in linear values and not in dB values; values shown in Table I mean that 90% of the cell locations have K-factors greater than or equal to the K-factor specified. Finally, the specified Doppler is the maximum frequency parameter f_m of the round-shaped spectrum. Additionally, the SUI-3 channel models specify a normalization factor equal to -1.5113 dB, which must be added to each tap power to get 0 dB as the total mean power.

Subsequent distortion as impulse noise is modeled with a Bernoulli–Gaussian process $(\rm p = 0.05)$. Note that, due to the short length in samples of the preamble in our OFDM system, there is a low probability of a spike falling into it.

This OFDM system consists of $N = 64$ subcarriers conveying QPSK symbols. We consider a packet-based transmission, where each packet consists of a header at the beginning of the packet with a known training sequence or preamble to carry out channel estimation, followed by a certain number $L_x = 20$ of OFDM data symbols. At the preamble, there are two OFDM symbols with $N_p = 16$ pilot subcarriers with randomly generated symbols. Each OFDM symbol is appended a cyclic prefix to overcome the delay spread $\tau_{\max}$ of the channel. For transmission, we have chosen a channel bandwidth of $W = 2$ MHz. Since we sample at Nyquist rate, this yields a sampling interval $T_s = 1/f_s = 0.5\,\mu$s; this means a length for the cyclic prefix L_{CP} of two samples. The total length of each OFDM symbol becomes $(N + L_{CP})$ samples. This cyclic prefix is removed at the receiver side, before performing any demodulation or detection operation. After the estimation (with either SVM or LS) of channel coefficients at pilot positions $\hat{H}_k$, we use them to compute the interpolation of the channel. Next, we perform zero forcing (ZF) equalization using the interpolated channel. Detection is carried out with a hard-decision slicer over the equalized data.

Channel estimation for the coherent demodulation of this OFDM system is performed with SVM algorithm. For comparison purposes, LS channel estimates in the frequency domain (after DFT demodulation) are simultaneously obtained in all cases. In both algorithms, a DFT-based technique with zero-padding in the time domain is used to interpolate the channel's frequency response for data subcarrier positions.

The signal-to-impulse ratio (SIR) is defined as SIR $= (E(r_n - z_n))/E(g_n)$, and it ranged from -21 to 21 dB. The bit-error rate (BER) and the mean-squared error (MSE)
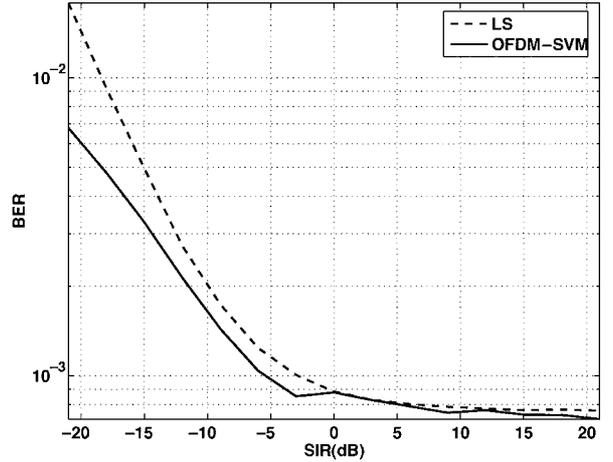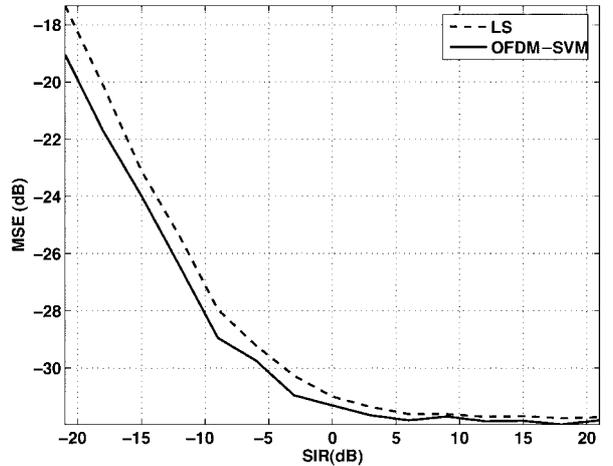
Fig. 1. Performance of OFDM-SVM versus LS. (a) BER performance as a function of SIR. (b) MSE performance as a function of SIR.

improvements that can be attained with SVM are represented in Fig. 1. For high SIR values, a performance similar to LS can be obtained, whereas for low SIR values, SVM outperforms LS by properly choosing C, δC, and ε.

Given that the choice of the free parameters in SVM is often a key issue, we studied the performance variation in the system due to changes in their values. Therefore, they were explored according to $C \in (1, 10^2)$, $\delta C \in (10^{-4}, 0.5)$, $\varepsilon \in (0, 1)$, and at 40 points per interval. Each exploration consisted of 500 realizations. Note that two of the most common robust cost functions (regularized LS and Huber cost) are implicit in this SVM approach.

The difference in BER between SVM and LS has been analyzed, given the definition of $\Delta\log(\rm BER) = \log(\rm BER_{SVM}) - \log(\rm BER_{LS})$, in order to easily detect working zones where SVM works better ($\Delta\log(\rm BER) > 0$), similar ($\Delta\log(\rm BER) = 0$) or worse ($\Delta\log(\rm BER) < 0$) than LS. Fig. 2 shows that the BER performance of the SVM algorithm can be superior to LS for properly chosen values of the free parameters and that there is a wide range of values for it. For instance, good performance remains for values of C ranging from 1 to 100 and for δC within a range from $10^{0.5}$ to 10^3.

Therefore, a reasonable first choice of the free parameters can be done, and the performance of the algorithm will not be highly

Fig. 2. Performance of OFDM-SVM with free parameters, compared to LS in terms of $\Delta \log 10 (\mathrm{BER})$. (a) SIR versus C. (b) SIR versus δC. (c) SIR versus ε.

affected by moderate changes in the conditions. The choice of δC depends on the power of non-Gaussian noise. Taking into account that thermal noise is mainly a zero-mean, low amplitude contribution, and that noise impulses are high-amplitude spikes, δC must be adjusted to produce a quadratic cost around a rough estimate of the thermal noise, and linear otherwise. For almost all noise values, ε must be small and linearly growing with the thermal noise. These results are coherent with [14] and [15], in which two similar rules for the choice of these parameters suggest that a near-optimal value of ε in SVM regressors must be proportional to the power of thermal noise.

V. CONCLUSION

A new SVM algorithm for OFDM coherent demodulation has been proposed. Rather than a classification approach, our formulation is based on a complex regression expression specifically developed for pilot-based OFDM systems. Simulations have confirmed the capabilities of OFDM-SVM in the presence of impulse noise interfering with the pilot symbols. The proposal allows a simple implementation and a straightforward choice of free parameters, and its cost function is robust against several different noise scenarios. This novel scheme turns out to be a framework for the development of new applications of the SVM technique in OFDM systems, and further problem statements of OFDM-SVM can be explored by taking into account the temporal-spectral relationship of the OFDM signal. The well-known Mercer theorem will provide a natural nonlinear extension of this algorithm (for instance, using radial basis functions). This may lead to a significant benefit for OFDM communications in those scenarios in which nonlinear distortion is present.

REFERENCES

[1] H. Sampath, S. Talwar, J. Tellado, V. Erceg, and A. Paulraj, "A fourth-generation MIMO-OFDM broadband wireless system: Design, performance and field trial results," *IEEE Commun. Mag*, vol. 40, no. 9, pp. 143–149, Sep. 2002.

[2] O. Edfors, M. Sandell, J. van de Beek, S. Wilson, and P. O. Börjesson, Analysis of DFT-based channel estimators for OFDM Div. Signal Process., Luleå Univ. Technol., Sweden, Res. Rep. TULEA, 1996, vol. 17.

[3] S. Chen, S. Gunn, and C. Harris, "Decision feedback equaliser design using support vector machines," *Proc. Inst. Elect. Eng., Vis. Image Signal Process.*, vol. 147, no. 3, pp. 213–219, Jun. 2000.

[4] D. Sebald and A. Buclew, "Support vector machine techniques for nonlinear equalization," *IEEE Trans. Signal Process.*, vol. 48, no. 11, pp. 3217–3226, Nov. 2000.

[5] S. Chen, A. K. Samingan, and L. Hanzo, "Support vector machine multiuser receiver for DS-CDMA signals in multipath channels," *IEEE Trans. Neural Netw.*, vol. 12, no. 3, pp. 604–611, May 2001.

[6] S. Rahman, M. Saito, M. Okada, and H. Yamamoto, "An MC-CDMA signal equalization and detection scheme based on support vector machines," in *Proc 1st Int. Symp. Wireless Communication Systems*, 2004, pp. 11–15.

[7] M. P. Sánchez-Fernández, M. de Prado-Cumplido, J. Arenas-García, and F. Pérez-Cruz, "SVM multiregression for nonlinear channel estimation in multiple-input multiple-output systems," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2298–2307, Aug. 2004.

[8] J. L. Rojo-Álvarez, G. Camps-Valls, M. Martínez-Ramón, E. Soria-Olivas, A. N. Navia Vázquez, and A. R. Figueiras-Vidal, "Support vector machines framework for linear signal processing," *Signal Process.*, vol. 85, no. 12, pp. 2316–2326, 2005.

[9] W. Bai, C. He, L. G. Jiang, and X. X. Li, "Robust channel estimation in MIMO-OFDM systems," *Electron. Lett.*, vol. 39, no. 2, pp. 242–244, Jan. 2003.

[10] M. Ghosh, "Analysis of the effect of impulse noise on multicarrier and single carrier QAM systems," *IEEE Trans. Commun.*, vol. 44, no. 2, pp. 145–147, Feb. 1996.

[11] M. J. Fernández-Getino García, J. M. Páez-Borrallo, and S. Zazo, "DFT-based channel estimation in 2D-pilot-symbol-aided OFDM wireless systems," in *Proc IEEE Vehicular Technology Conf.*, 2001, vol. 2, pp. 815–819.

[12] A. Papoulis, *Probability Random Variables and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.

[13] V. Erceg *et al.*, Channel models for fixed wireless applications IEEE 802.16 Broadband Wireless Working Group, IEEE802.16a-03/01, Tech. Rep., 2003.

[14] J. T. Kwok and I. W. Tsang, "Linear dependency between ε and the input noise in ε-support vector regression," *IEEE Trans. Neural Netw.*, vol. 14, no. 3, pp. 544–553, May 2003.

[15] V. Cherkassky and Y. Ma, "Practical selection of SVM parameters and noise estimation for SVM regression," *Neural Ne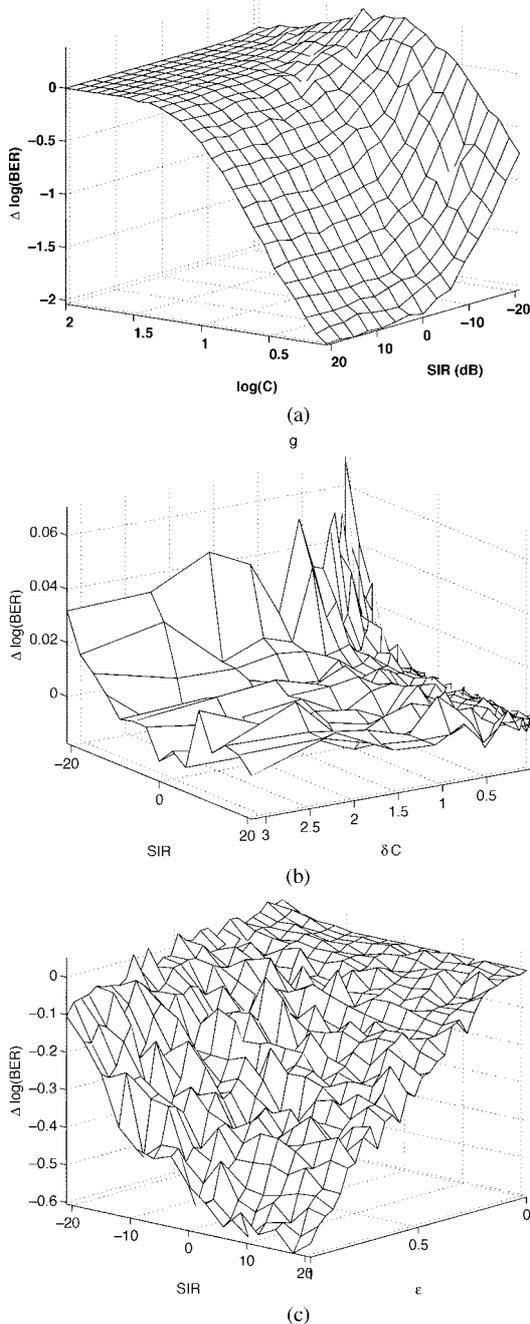tw.*, vol. 17, no. 1, pp. 113–126, 2004.