Research Article

Relevance Vector Machines for Enhanced BER Probability in DMT-Based Systems

Ashraf A. Tahat¹ and Nikolaos P. Galatsanos²

¹ School of Electrical Engineering, Princess Sumaya University for Technology, Amman 11941, Jordan

² Electrical and Computer Engineering Department, University of Patras, Rio 26500, Greece

Correspondence should be addressed to Ashraf A. Tahat, tahat@psut.edu.jo

Received 15 December 2009; Revised 15 March 2010; Accepted 27 April 2010

Academic Editor: Cyril Leung

Copyright © 2010 A. A. Tahat and N. P. Galatsanos. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A new channel estimation method for discrete multitone (DMT) communication system based on sparse Bayesian learning relevance vector machine (RVM) method is presented. The Bayesian frame work is used to obtain sparse solutions for regression tasks with linear models. By exploiting a probabilistic Bayesian learning framework, sparse Bayesian learning provides accurate models for estimation and consequently equalization. We consider frequency domain equalization (FEQ) using the proposed channel estimate at both the transmitter (*preequalization*) and receiver (*postequalization*) and compare the resulting bit error rate (BER) performance curves for both approaches and various channel estimation techniques. Simulation results show that the proposed RVM-based method is superior to the traditional least squares technique.

1. Introduction

One of the research goals during the last decades was to provide broadband communication capabilities to and from the customer premises [1, 2]. To cope with the time dispersive transmission character of wireline and wireless communications, multicarrier modulation (MC) offers a viable solution. Since the early seventies a widespread interest was created due to an all digital implementation based on the cost-effective fast Fourier transform (FFT) algorithm [3]. Nowadays, MC systems are included in digital audio or video broadcasting (DAB/DVB) [3, 4] in wireless local area networks such as IEEE802.11a/g [5] and HIPERLAN2 [6], wireless wide area networks like WiMAX and LTE systems [7], and in wireline communication over twisted pairs such as in ADSL, ADSL2+, VDSL [8, 9] and also in power-line local area networks [10].

At the present time, the transmission format as presented in [3] is more commonly known as Discrete Multi-tone (DMT) modulation or Orthogonal Frequency Division Multiplexing (OFDM). An inherent feature of DMT over traditional OFDM is that DMT systems make use of the channel state information (CSI) at the transmitter [11, 12].

DMT divides the available bandwidth into several parallel frequency bins or tones, where each tone is quadrature amplitude modulated (QAM). Modulation and demodulation are efficiently carried out by means of an inverse Fourier transform (IFFT) and a fast Fourier transform (FFT), respectively. Contrary to analog MC systems, FFT-based implementation permits considerable overlap between sub-channels and therefore provides a high bandwidth efficiency.

As mentioned, the various flavors of Digital Subscriber Line technologies (xDSL) represent a good example of DMT practical applications over traditional twisted pair phone lines to connect between customer premises and the Central Office (CO). We note that the bit rates highly depend on the length of the twisted pair: higher bit rates are only offered on relatively short loops due to signal attenuation caused by the telephone line. Since line attenuation increases with length and frequency, short loops are able to exploit a broader frequency spectrum and hence enable higher data rates. Tones with higher Signal-to-Noise Ratio (SNR) are able to carry more bits for a predetermined Bit Error Rate (BER) than those with low SNR. The measurement of SNR is usually performed during the initialization phase of the modem during which the frequency response of the channel is also estimated.

FIGURE 1: (a) DMT system block diagram with FEQ at receiver. (b) DMT system block diagram with FEQ at transmitter.

FIGURE 2: Block diagram of the baseband DMT modulator.

To deliver high bit rates towards the customer, DMT modems depend on advanced digital signal processing to mitigate several loop (channel) impairments such as time dispersion, noise, echo, and radio frequency interference (RFI) [13, 14]. In this paper, we formulate a Relevance Vector Machine (RVM) method [15] for a DMT-baseband system in a regression mode to improve channel estimation when the SNR is low to moderate, which in turn, will improve the probability of error performance when compared with the traditional Least Squares (LS) approach [16].

In Section 2 of this paper, we describe the models used for analysis and design. In Section 3 we present the RVM channel estimation methodology. In Section 4 we present the channel equalization approaches used. In Section 5 we present simulation results, and we conclude in Section 6.

2. System Model

Consider the system block diagram shown in Figure 1, which represents a baseband Discrete Multi-Tone (DMT) link that includes a DMT modulator, the channel model, DMT demodulator, and frequency domain equalizer (FEQ).

Figure 1(a) depicts the DMT system with FEQ implemented at the receiver (*postequalization*), while in Figure 1(b) we show the configuration of FEQ performed at the transmitter (*preequalization*) [11].

2.1. The DMT Modulator. DMT transmission [3] splits a high-rate data stream, R , into N lower rate streams transmitted simultaneously over subcarriers. The symbol rate for any of the N complex data streams is R/N symbols/second, which are sent to a $2N$ -point inverse fast Fourier transform (IFFT) block after zero padding, forming a complex conjugate symmetry around the center of the IFFT and converting the frequency-domain data symbols into $2N$ time-domain real-valued samples as shown in Figure 2.

A cyclic prefix (CP) of length ν is prepended to the $2N$ -point time-domain samples to form the cyclically extended DMT symbol.

This is similar to the configuration used for DMT-based Asymmetric Digital Subscriber Line (ADSL), but with fixed modulation on each subchannel (i.e., QPSK). The cyclic prefix length, ν , is chosen to encompass the maximum *delay spread* of the channel to prevent intersymbol interference (ISI) and making the DMT symbol appear periodic over the time span of interest.

2.2. The Channel Model. When the channel impulse response (CIR) is smaller than or equal in length to the CP length, ν , the $2N$ -DMT-sampled time points (Baud) and the channel are circularly convolved. This enables straight forward frequency-domain channel equalization at

the receiver. However, when the order of the CIR is larger than the cyclic prefix, more complex equalization techniques are required [17]. In this paper, the channel that we will consider has a CIR that does not exceed the CP. Hence, inserting of the CP at the transmitter and then discarding the first ν samples at the receiver give the received real-valued sampled-time domain samples over one DMT data block (baud) as follows:

$$\mathbf{r} = \mathbf{C}\mathbf{x} + \mathbf{n}, \quad (1)$$

where the vector $\mathbf{r} = [r_0, r_1, r_2, \dots, r_{2N-1}]^T$ contains the received time-domain samples, the vector $\mathbf{x} = [x_0, x_1, x_2, \dots, x_{2N-1}]^T$ contains the transmitted time-domain samples, and the vector $\mathbf{n} = [n_0, n_1, n_2, \dots, n_{2N-1}]^T$ contains the samples of the band-limited additive noise with average power (variance) σ_n^2 . The matrix multiplication operator in (1) represents circular convolution without intersymbol interference (ISI). The circulant matrix $\mathbf{C} \in \mathbb{R}^{2N \times 2N}$

$$\mathbf{C} = \begin{bmatrix} h_0 & 0 & \dots & 0 & h_L & \dots & h_1 \\ h_1 & h_0 & & & & & \vdots \\ \vdots & & \ddots & & & & h_L \\ h_L & & & & 0 & & \vdots \\ 0 & h_L & & & & & \vdots \\ \vdots & \ddots & \ddots & & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & h_L & \dots & h_1 & h_0 \end{bmatrix}, \quad (2)$$

can be diagonalized by pre- and post-multiplication with the $2N$ -point discrete Fourier transform (DFT) matrix, $\mathbf{W}_{2N}$, and inverse DFT (IDFT) matrix $(\mathbf{W}_{2N})^{-1}$ [18] that is,

$$\mathbf{W}_{2N}\mathbf{C}\mathbf{W}_{2N}^{-1} = \text{diag}\{[H_0 \dots H_{2N-1}]\}, \quad (3)$$

where $[H_0, H_1, H_2, \dots, H_{2N-1}]^T$ is the frequency response of the channel.

When the transmitted symbols are generated in the frequency domain (i.e., $\mathbf{x} = \mathbf{W}_{2N}^{-1}[X_0, X_1, X_2, \dots, X_{2N-1}]^T$), and the received samples are converted to the frequency domain (i.e., $[R_0, R_1, R_2, \dots, R_{2N-1}]^T = \mathbf{W}_{2N}\mathbf{r}$), we obtain the following very simple input-output relationship:

$$\begin{bmatrix} R_0 \\ \vdots \\ R_{2N-1} \end{bmatrix} = \begin{bmatrix} H_0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & H_{2N-1} \end{bmatrix} \begin{bmatrix} X_0 \\ \vdots \\ X_{2N-1} \end{bmatrix} + \begin{bmatrix} N_0 \\ \vdots \\ N_{2N-1} \end{bmatrix}, \quad (4)$$

or, alternatively as follows:

$$\begin{bmatrix} R_0 \\ \vdots \\ R_{2N-1} \end{bmatrix} = \begin{bmatrix} X_0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & X_{2N-1} \end{bmatrix} \begin{bmatrix} H_0 \\ \vdots \\ H_{2N-1} \end{bmatrix} + \begin{bmatrix} N_0 \\ \vdots \\ N_{2N-1} \end{bmatrix}, \quad (5)$$

where $[N_0, N_1, N_2, \dots, N_{2N-1}]^T = \mathbf{W}_{2N}\mathbf{n}$. From (4) and (5), we can see that each received frequency-domain symbol on each of the subchannels is simply a scaled version of the transmitted frequency-domain symbol plus white Gaussian noise. Moreover, every sub-channel can be processed

independently of the other subchannels. In other words, block transmission with CP has converted a time dispersive channel into $2N$ parallel, narrowband flat sub-channels, or tones each having a channel gain H_k and additive Gaussian noise, hence the term discrete multi-tone transmission (DMT).

3. Channel Estimation

An important aspect of DMT is the fact that it uses an optimized frequency division allocation of energy and bits to maximize the achievable data rates that can be transmitted over band-limited channels. During startup of the DMT modem, the receiver measures the quality of the signals received (SNR) on each tone and reports this information to the transmitter [9], this is repeated periodically for dynamic channels. To estimate the channel during initialization, a pseudorandom frequency-domain training (pilot) sequence, $X_b(k)$, is employed. Where $\mathbf{X}_b$ is some constant modulus frequency-domain training sequence (i.e., $|X_b(k)| = \text{constant}$ for $k = 0, 1, \dots, 2N-1$). Here, without loss of generality, we normalize $\mathbf{X}_b$ so that $|X_b(k)| = 1$. Now, the received pilot symbols can be expressed in vector form as follows:

$$\mathbf{R}_b = \mathbf{X}_b\mathbf{H} + \mathbf{N}. \quad (6)$$

In (6), $\mathbf{R}_b$ is the $2N \times 1$ received signal vector, the diagonal matrix $\mathbf{X}_b$ is the transmitted (pilot) signal, $\mathbf{N}$ is a vector containing the complex noise of the $2N$ subcarriers, and $\mathbf{H}$ is a vector that contains the overall complex channel gains between transmitter and receiver.

The problem at hand is to estimate the frequency-domain channel vector, H_k , ($k = 0, 1, 2, 3, \dots, N-1$) as the FFT of L unknown sample-spaced time-domain tap gains, where L is chosen to encompass the maximum expected delay spread and does not exceed the CP length, ν . The frequency-domain model of the channel for each of the $2N$ tones is given by

$$H_k = \sum_{n=0}^{L-1} h_n e^{-j2\pi kn/2N}, \quad (7)$$

where h_n is channel tap at discrete time n .

To perform channel estimation from (6), we start by multiplying each of the frequency-domain symbols by the conjugate of the training symbols to produce the vector $\mathbf{T}_b$ as follows:

$$\mathbf{T}_b = \mathbf{X}_b^H \mathbf{R}_b = \mathbf{X}_b^H \mathbf{X}_b \mathbf{H} + \mathbf{X}_b^H \mathbf{N}, \quad (8)$$

where the superscript H denotes conjugate-transpose. Because of the property of unitary magnitude of the training symbols, (8) can be expressed as

$$\mathbf{T}_b = \mathbf{X}_b^H \mathbf{R}_b = \mathbf{H} + \mathbf{N}'. \quad (9)$$

In the noise-free case, $\mathbf{T}_b$ in (9) will contain the perfect frequency-domain channel gains. Due to channel noise, $\mathbf{T}_b$ contains the channel estimates that are corrupted by the additive noise, $\mathbf{N}'$. This is equivalent to least squares (LS) estimation [16] in which no assumption about the channel

impulse response length has been incorporated (i.e., h_n can have values larger than zero for $n = 0, 1, \dots, 2N-1$).

The LS estimator for the impulse response, $\mathbf{H}$, minimizes the criterion $(\mathbf{R}_b - \mathbf{X}_b \mathbf{H})^H (\mathbf{R}_b - \mathbf{X}_b \mathbf{H})$ and is given by

$$\mathbf{H}_{LS} = (\mathbf{X}_b^H \mathbf{X}_b)^{-1} \mathbf{X}_b^H \mathbf{R}_b, \quad (10)$$

where due to the unitary magnitude of the training symbols we have

$$\mathbf{H}_{LS} = \mathbf{X}_b^H \mathbf{R}_b = \mathbf{T}_b. \quad (11)$$

When the values of the channel impulse response are forced to be zero for sample numbers larger than the cyclic prefix duration, ν , then this will be equivalent to “modified LS” estimation [16].

Our objective is to improve the performance of the channel estimation at low to moderate SNR after truncation of the channel impulse response to ν [19]. Specifically, we will use a regression model similar to the one used for the relevance vector machine method (RVM) [15, 20]. We apply this model to baseband-DMT channel estimation by first taking the inverse Fourier transform of (9) to fit a regression model as follows:

$$\mathbf{t}_b = \mathbf{h} + \mathbf{n}', \quad (12)$$

where $\mathbf{t}_b = [t_0, t_1, t_2, \dots, t_{2N-1}]^T$ is the target vector time domain samples, $\mathbf{h} = [h_0, h_1, h_2, \dots, h_\nu, 0, \dots, 0]^T$ is the $2N \times 1$ channel vector, and the vector $\mathbf{n}' = [n_0, n_1, n_2, \dots, n_{2N-1}]^T$ contains samples of the band-limited additive Gaussian noise with variance σ'^2 .

In order to estimate $\mathbf{h}$ from the observations of $\mathbf{t}_b$ in (12) and ameliorate the effects of the noise, we will use sparse Bayesian regression. This approach has the property of setting the appropriate regression weights to zero automatically, avoiding the fitting of noise in the signal $\mathbf{t}_b$.

Assume that the channel can be approximated using the function $y(n)$ which is the linear combination of kernel functions as follows:

$$y(n) = \sum_{i=0}^{\nu} w_i \phi(n-i). \quad (13)$$

This is a convolution that can be written in matrix vector as

$$\mathbf{y} = \Phi \mathbf{w}. \quad (14)$$

where $\mathbf{w} = [w_0, w_1, w_2, \dots, w_i, \dots, w_\nu]^T$ are the model weights, and Φ is the $\nu \times (\nu+1)$ convolution (design) matrix that is created from the kernel we select. One commonly used kernel is the Gaussian-shaped given by

$$\phi(n-i) = \exp\left(\frac{-\|n-i\|^2}{2\sigma^2}\right). \quad (15)$$

However, there are many other choices (e.g., spline, polynomial, etc.) [20]. Now, our model becomes

$$\mathbf{t}_b = \mathbf{y} + \mathbf{e} = \Phi \mathbf{w} + \mathbf{e}, \quad (16)$$

where $\mathbf{y} = [y_0, y_1, y_2, \dots, y_\nu]^T$ is an approximation function and $\mathbf{e} = [e_0, e_1, e_2, \dots, e_\nu]^T$ is the vector with the regression model error.

The Bayesian model used assumes that the errors are modeled as independent identically distributed zero-mean Gaussian random variables, with variance σ^2 , so

$$p(\mathbf{e}) = \prod_{i=1}^{\nu} \mathcal{N}(e_i | 0, \sigma^2). \quad (17)$$

The parameter σ^2 can be set in advance if known, but can be also estimated from the data. This error model implies a multivariate Gaussian likelihood for the target vector $\mathbf{t}_b$

$$p(\mathbf{t}_b | \mathbf{w}, \sigma^2) = (2\pi\sigma^2)^{-2N/2} \exp\left(\frac{-\|\mathbf{t}_b - \Phi \mathbf{w}\|^2}{2\sigma^2}\right). \quad (18)$$

Maximum-likelihood estimation of $\mathbf{w}$ from (18) is equivalent to the Least Squares solution given by

$$\mathbf{w}_{LS} = (\Phi^T \Phi)^{-1} \Phi^T \mathbf{t}_b, \quad (19)$$

and leads to severe overfitting of the data [20].

In order to overcome this problem, a flexible Gaussian prior over the weights $\mathbf{w}$ and Bayesian inference will be used [20]. More specifically,

$$p(\mathbf{w}, \boldsymbol{\alpha}) = \prod_{i=1}^{\nu} \mathcal{N}(w_i | 0, \alpha_i^{-1}), \quad (20)$$

with $\boldsymbol{\alpha}$ is a vector of ν hyperparameters [15, 20]. The flexibility of this prior is based on the fact that it uses a *separate hyperparameter* for every weight w_i .

The weights $\mathbf{w}$ are marginalized in according to

$$p(\mathbf{t}_b; \mathbf{a}, \sigma^2) = \int p\left(\frac{\mathbf{t}_b}{\mathbf{w}}; \sigma^2\right) p(\mathbf{w}; \mathbf{a}) d\mathbf{w}. \quad (21)$$

This integral can be found in closed form and is given by

$$p(\mathbf{t}_b | \boldsymbol{\alpha}, \sigma^2) = N(\mathbf{t}_b | \mathbf{0}, \mathbf{B}^{-1} + \Phi \mathbf{A}^{-1} \Phi^T), \quad (22)$$

where we have defined $\mathbf{A} = \text{diag}(\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_\nu)$, $\mathbf{B} = \sigma^2 \mathbf{I}_\nu$, and $\mathbf{I}_\nu$ is the $\nu \times \nu$ identity matrix. The pdf $p(\mathbf{t}_b | \boldsymbol{\alpha}, \sigma^2)$ is called the *marginal likelihood* (or *evidence*) for the hyperparameters $\boldsymbol{\alpha}$ and σ^2 [21]. Because the involved pdfs are Gaussian, calculation of the *posterior* in closed form is also possible and is given by

$$p(\mathbf{w} | \mathbf{t}_b, \boldsymbol{\alpha}, \sigma^2) = N(\mathbf{w} | \boldsymbol{\mu}_{w/t}, \boldsymbol{\Sigma}_{w/t}), \quad (23)$$

with

$$\boldsymbol{\Sigma}_{w/t} = (\Phi^T \mathbf{B} \Phi + \mathbf{A})^{-1}, \quad \boldsymbol{\mu}_{w/t} = \boldsymbol{\Sigma}_{w/t} \Phi^T \mathbf{B} \mathbf{t}_b. \quad (24)$$

The regression estimates obtained by Bayesian Inference are given by $\Phi \boldsymbol{\mu}_{w/t}$ where the hyperparameters $\boldsymbol{\alpha}$ and σ^2 used in $\boldsymbol{\mu}_{w/t}$ are the *Maximum Likelihood* (ML) estimates given the observations $\mathbf{t}_b$ and are found by maximizing $p(\boldsymbol{\alpha}, \sigma^2 | \mathbf{t}_b)$ [15]. However, maximization of (22) with respect to $\boldsymbol{\alpha}$ and

σ^2 , which is termed as the *Type II ML* [20, 21], is equivalent to finding the maximum of $p(\alpha, \sigma^2 | \mathbf{t}_b)$ assuming a uniform hyperprior [15, 20].

The values of α and σ^2 which maximize (22) cannot be obtained in closed form. However, they can be found by stationary point analysis of (22). More specifically, they are found by the following iterative update equations [22]:

$$\alpha_i^{\text{new}} = \frac{\gamma_i}{\mu_i^2}, \quad (\sigma^2)^{\text{new}} = \frac{\|\mathbf{t}_b - \Phi \mu_{w/t}\|^2}{\nu - \sum_i \gamma_i}, \quad (25)$$

where μ_i is the i th component of the posterior mean $\mu_{w/t}$ from (24) and the quantities γ_i are defined by $\gamma_i = 1 - \alpha_i \Sigma_{ii}$ with Σ_{ii} being the i th diagonal element of the posterior weight covariance and $\Sigma_{w/t}$ from (24) computed with the current α and σ^2 values.

The learning algorithm thus proceeds by iterating between (24) and (25) until some convergence criteria have been satisfied [15, 22]. It can be shown that this iteration is equivalent to an expectation-maximization (EM) algorithm [20]. The E-step is (24), and the M-step is (25). This relation guarantees convergence of the proposed algorithm [20].

An important property of this model is that the optimal value of most hyperparameters α_i is typically *infinite* [15, 22]. From (24) this implies that the corresponding μ_i is zero. Thus, the regressor $y(n) = \sum_{i=0}^{\nu} \mu_i \varphi(n-i)$ is *sparse* since many of its weights μ_i are zero. In matrix form, the RVM-channel estimate is expressed as

$$h_{\text{RVM}} = \Phi \mu_{w/t}. \quad (26)$$

4. Frequency-Domain Channel Equalization

If the channel is known for the noise-free case, we can perform the equalization by inverting (1) assuming $\mathbf{C}$ is non-singular. This is the zero-forcing (ZF) equalizer [23]. The eigenvalues of the $2N \times 2N$ circulant matrix are equal to the DFT coefficients of the first column [24]. Since the first column has the channel coefficients in ascending order, from (6) these eigenvalues are

$$H_k = \sum_{n=0}^{L-1} h_n e^{-j2\pi kn/2N} = \sum_{n=0}^{L-1} h_n W^{-kn}, \quad (27)$$

where $W = e^{j2\pi/2N}$, and H_k represents the channel frequency response. Note that $h_n = 0$ for $L < n < 2N$.

The circulant matrix $\mathbf{C}$ can be diagonalized with the DFT matrix [24] as mentioned in (3) as follows:

$$\mathbf{W}_{2N} \mathbf{C} \mathbf{W}_{2N}^{-1} = \text{diag}\{[H_0 \cdots H_{2N-1}]\} = \Lambda_C. \quad (28)$$

Because the DFT (and thus its matrix $\mathbf{W}_{2N}$) is invertible, we deduce that the circulant matrix, $\mathbf{C}$, is invertible if and only if Λ_C is invertible. And, the diagonal matrix will be invertible if it has no zero on its diagonal (i.e., the channel transfer function has no zero on the DFT grid). The diagonal elements of $[\Lambda_C]^{-1}$ are $1/H_k$, which can be regarded as a set of DFT-domain equalizers. The symbols are equalized at

the receiver as in Figure 1(a) in the frequency domain before detection as follows:

$$\hat{\mathbf{X}} = \Lambda_C^{-1} \mathbf{R}, \quad (29)$$

where $\hat{\mathbf{X}}$ is a $2N \times 1$ vector, and $\mathbf{R}$ is a $2N \times 1$ received signal vector. Of course, in practice channel noise can be amplified by $[\Lambda_C]^{-1}$ if the coefficients, $1/H_k$, are large. This is the result of an optimization based on the ZF criterion which aims at equalizing the symbols regardless of noise level. Where on each subcarrier, the received symbol is multiplied by a complex multiplier to off set the effect of the channel, which is manifested in the form of attenuation and phase rotation. We can express the equalized received symbol on each subcarrier in equation form as

$$\hat{X}_k = X_k + \frac{1}{H_k} N_k. \quad (30)$$

In this version (*postequalization*), the receiver has all the complexity.

If the channel is known at the transmitter (e.g., transmitter has access to CSI through a feedback channel [25]), as it is the case in DMT [11, 12]. We can move $[\Lambda_C]^{-1}$ to the transmitter. Equalization in this configuration is implemented at the transmitter (*preequalization*) as shown in Figure 1(b). This results in superior performance [11] as will be demonstrated in the simulation section. The symbols are equalized in the frequency domain before modulation and transmission as follows:

$$\check{\mathbf{X}} = \Lambda_C^{-1} \mathbf{X}, \quad (31)$$

where $\check{\mathbf{X}}$ is a $2N \times 1$ symbols vector. And the received symbol on each subcarrier after demodulation is expressed as

$$\hat{X}_k = X_k + N_k. \quad (32)$$

We can see that channel noise will not be amplified for this second version. In addition, this yields a useful configuration for the cases where the receiver is simpler.

5. Simulation Results

To evaluate the performance of the DMT/RVM system, it was simulated using MATLAB. The RVM-channel estimation scheme was evaluated and compared to other methods in addition to the ideal perfect case. The following are the common simulation parameters: $2N = 512$ is the DMT FFT size; $\nu = 32$ is the Cyclic Prefix length. The additive channel noise is band-limited white Gaussian with power σ_n^2 . The actual channel length is $L = 32$. The used subchannels consisted of tones 6 to 255. Also, each of the subcarriers uses a QPSK modulation scheme.

In Figure 3, we present an example target vector time domain samples of (12), which in the noise-free case should be the actual impulse response of the channel and are superimposed on the same plot.

FIGURE 3: An example target vector.

FIGURE 4: Channel impulse response and its estimate obtained using the RVM method.

The channels were assumed to remain constant after each sounding using the pilot sequence. Figure 4 depicts the actual sampled-time domain channel in the cyclic prefix range where it is equal to zero for values of time larger than the cyclic prefix. The target vector is also plotted in the same time range, which is equivalent to the Least Squares (LS) estimate of the channel. The proposed Bayesian sparse regression algorithm is used to process the target vector samples to filter out the noise at an SNR of 5 dB. The locations of the non-zero regression weights are identified on the plot with circles. Notice that we use *only 7 nonzero weights* from a total of 32. The remaining 25 are set to zero by the sparse model.

FIGURE 5: Frequency response of actual and estimated channel.

FIGURE 6: BER curves versus SNR with equalization at Transmitter.

To demonstrate the effects of estimating the channel impulse response on the frequency attenuation of each subcarrier (and hence equalization), in Figure 5 we plotted the frequency response of the estimated channel, the actual channel, and the original noise corrupted channel.

From Figure 5 we observe that in the frequency domain the estimated channel by the sparse model has smoother variation between adjacent tones and is closer to the desired frequency response.

When FEQ is implemented at the transmitter using the estimated channel at the receiver, BER performance curves (labeled TX-) demonstrate in Figure 6 the superior

FIGURE 7: Comparisons of BER curves with FEQ at receiver to that at transmitter.

performance attained as a result of the more accurate sparse regression channel (RVM) estimate over the LS one. Also, the BER performance curve based on the actual channel (perfect estimate) is superimposed on the same plot of Figure 6 as a performance bench mark.

We also compare in Figure 7 the accuracy in channel estimation when FEQ is implemented at the receiver as compared to that at the transmitter. We can make two observations. First, even though it is intuitive that FEQ at the transmitter (*preequalization*) will achieve better performance, we demonstrate that it is far superior to that at the receiver (*postequalization*) for the same channel estimate. Second, the improved accuracy in channel estimation has limited effect on BER performance curves when FEQ is implemented at the receiver (labeled RX-), and BER curves are virtually identical for the actual channel, sparse regression (RVM) channel estimate, and the LS channel estimate. Low SNR on attenuated subcarriers renders the improved channel accuracy not important in this case, since equalization will also enhance noise simultaneously.

6. Conclusions

Proper equalization is commonly used to improve utilization of available bandwidth in a DMT-data transmission system. In this paper, an improved channel estimation method in a DMT communication system based on sparse Bayesian learning and the RVM method was presented. Deploying the RVM method, more accurate channel estimates in a low to medium SNR environments were attained. Although, frequency-domain equalization at the transmitter (*preequalization*) will intuitively yield better BER performance, we demonstrated the performance enhancement achieved when using the improved channel estimate via the RVM

method. We also show through simulations that preequalization is more sensitive to channel estimation accuracy than frequency equalization at the receiver side (*postequalization*). Moreover, when compared to the traditional least squares technique, the proposed enhanced channel estimation scheme had superior performance and exhibited promising characteristics, particularly when applied to an ADSL system.

References

- [1] Broadband Forum, "Next generation broadband access white paper," Marketing Report 185, August 2009.
- [2] E. Dahlman, S. Parkvall, and P. Beming, *3G Evolution: HSPA and LTE for Mobile Broadband*, Academic Press, Oxford, UK, 2007.
- [3] S. B. Weinstein and P. M. Ebert, "Data transmission by frequency-division multiplexing using the discrete Fourier transform," *IEEE Transactions on Communications*, vol. 19, no. 5, part 1, pp. 628–634, 1971.
- [4] The European Telecommunications Standards Institute, "Digital video broadcasting (DVB); framing structure, channel coding and modulation for terrestrial digital terrestrial television," Tech. Rep. EN 300 744 V1.4.1, ETSI, 2001.
- [5] The Institute of Electrical and Electronics Engineers, "Wireless LAN medium access control (MAC) and physical layer (PHY) specification," IEEE standard 802.11a, 1999.
- [6] The European Telecommunications Institute, "Broadband radio access networks (BRAN); high performance radio local area networks (HIPERLAN) type 2; system overview," Tech. Rep. 101 683 V1.1.1, ETSI, 2000.
- [7] S. Glisic, *Advanced Wireless Communications: 4G Technologies*, John Wiley & Sons, Chichester, England, 2004.
- [8] P. J. Kyees, R. C. McConnell, and K. Sistanizadeh, "ADSL. A new twisted-pair access to the information highway," *IEEE Communications Magazine*, vol. 33, no. 4, pp. 52–59, 1995.
- [9] H. Dedieu, *Fundamentals of DSL Technology*, Auerbach, Boca Raton, Fla, USA, 2006.
- [10] S. Baig and N. D. Gohar, "A discrete multitone transceiver at the heart of the PHY layer of an in-home power line communication local area network," *IEEE Communications Magazine*, vol. 41, no. 4, pp. 48–53, 2003.
- [11] Z. Wang and G. B. Giannakis, "Wireless multicarrier communications: where Fourier meets Shannon," *IEEE Signal Processing Magazine*, vol. 17, no. 3, pp. 29–48, 2000.
- [12] F. J. Simois and J. I. Acha, "Linear IIR precoding for non-redundant transmission over minimum-phase channels in wireline multitone systems," *IEEE Transactions on Communications*, vol. 57, no. 5, pp. 1496–1504, 2009.
- [13] L. de Clercq, M. Peeters, S. Schelstraete, and T. Pollet, "Mitigation of radio interference in xDSL transmission," *IEEE Communications Magazine*, vol. 38, no. 3, pp. 168–173, 2000.
- [14] C.-C. Li and Y.-P. Lin, "Receiver window designs for radio frequency interference suppression in DMT systems," *IET Signal Processing*, vol. 3, no. 1, pp. 33–39, 2009.
- [15] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, no. 3, pp. 211–244, 2001.
- [16] J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Borjesson, "On channel estimation in OFDM systems," in *Proceedings of the IEEE 45th Vehicular Technology Conference*, pp. 815–819, July 1995.

- [17] T. Pollet, M. Peeters, M. Moonen, and L. Vandendorpe, "Equalization for DMT-based broadband modems," *IEEE Communications Magazine*, vol. 38, no. 5, pp. 106–113, 2000.
- [18] G. Golub and C. Van Loan, *Matrix Computations*, The John Hopkins University Press, Baltimore, Md, USA, 3rd edition, 1996.
- [19] A. A. Tahat and N. P. Galatsanos, "Relevance vector machines for DMT based systems," in *Proceedings of the 20th International Conference on Electronics Communications and Computers (CONIELECOMP '10)*, pp. 31–36, February 2010.
- [20] C. Bishop, *Pattern Recognition and Machine Learning*, Springer, Berlin, Germany, 2006.
- [21] J. Berger, *Statistical Decision Theory and Bayesian Analysis*, Springer, New York, NY, USA, 1985.
- [22] A. Faul and M. Tipping, "Analysis of sparse Bayesian learning," in *Proceedings of the Neural Information Processing Systems: Natural and Synthetic (NIPS '01)*, pp. 383–389, December 2001.
- [23] J. Proakis, *Digital Communications*, McGraw-Hill, New York, NY, USA, 3rd edition, 1995.
- [24] A. Papoulis, *Signal Analysis*, McGraw Hill, New York, NY, USA, 1977.
- [25] D. J. Love, R. W. Heath Jr., V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1341–1365, 2008.