A Robust Support Vector Algorithm for Nonparametric Spectral Analysis 💬

José Luis Rojo-Álvarez, *Member, IEEE*, Manel Martínez-Ramón, *Member, IEEE*,
Aníbal R. Figueiras-Vidal, *Senior Member, IEEE*, Ana García-Armada, *Member, IEEE*, and
Antonio Artés-Rodríguez, *Senior Member, IEEE*

Abstract—**A new approach to the nonparametric spectral estimation on the basis of the support vector method (SVM) framework is presented. A reweighted least squared error formulation avoids the computational limitations of quadratic programming. The application to a synthetic example and to a digital communication problem shows the robustness of the SVM spectral analysis algorithm.**

Index Terms—**Spectral analysis, support vector method, weighted least squares, Welch periodogram.**

I. INTRODUCTION

NONPARAMETRIC spectral analysis of time series is a widely scrutinized framework. The most relevant of the classical spectral estimators, the Welch periodogram and the Blackman–Tukey correlogram, are based on Fourier transform representations, either for the observed time series and for its estimated autocorrelation function [1], so their main advantages are the low computational burden required and their simplicity. On the other hand, their spectral resolution is limited due to the windowing effect. Also, the periodogram shows a high sensitivity to outliers, and it is strongly affected by impulsive noise, which is frequently present in many communication systems.

An alternative approach to the classical nonparametric spectral analysis can be drawn from the support vector method (SVM), which was first suggested to obtain maximum margin by separating hyperplanes in classification problems [2], and it has been extended to the general learning theory [3], [4]. In [5], the standard SVM regression algorithm is modified to provide an adequate approach to nonparametric spectral analysis problems, which is called the SVM-Spect formulation. SVM-Spect algorithms are solved via quadrating programming (QP), whose time demand grows exponentially with the length of the time series, making them useless for most of the practical applications.

We present here a iterative reweighted least squares (IRWLS) formulation [6] that overcomes this limitation. In Section II, the new SVM-Spect algorithm is derived from a IRWLS formula-

tion of the problem. In Section III, a simulation on synthetic data, with an asymmetric digital subscriber line (ADSL) communication system application, is presented to support the potential of this approach.

II. SVM-SPECT FROM A WLS FORMULATION

The model for the harmonic decomposition of a discrete time real-valued sequence $\{y_{t_k}\}$, where N consecutive samples are observed at instant times t_k, can be expressed in terms of the Fourier series. For uniform sampling, $t_k = k - 1$ is just the sample lag. The sinusoidal approximation is given by

$$y_{t_k} = \sum_{i=1}^{N_\omega} A_i \cos(\omega_i t_k - \phi_i) + e_{t_k} \qquad (1)$$

where the unknown parameters are amplitudes A_i, phases ϕ_i and frequencies ω_i for a number N_ω of sinusoidal components, and e_{t_k} is the model error for the kth sample. This is a non-linear relationship, except when frequencies are known. In this case, (1) can be linearly expressed by using Cartesian coordinates $c_i = A_i \cos(\phi_i)$ and $d_i = A_i \sin(\phi_i)$

$$y_{t_k} = \sum_{i=1}^{N_\omega} [c_i \cos(\omega_i t_k) + d_i \sin(\omega_i t_n)] + e_{t_k}. \qquad (2)$$

A vast number of methods have been suggested to find the model coefficients, such as least square analysis or the discrete-time Fourier transform [1], [7]. However, the performance of these methods can degrade in situations like presence of atypical samples (outliers), low SNR or low number of available observations. On the other hand, several robust cost functions have been used in SVM regression, like Vapnik's loss function [2], Huber's robust cost [8], or the ridge regression approach [9]. We propose here a more general cost function, which contains the above as particular cases. Fig. 1 depicts the relationship between model approximation error e and its corresponding cost, denoted by $L_P(e)$. In the SVM approach, the L_2 norm of the model coefficients and $L_P(e)$ are simultaneously minimized. We propose here a more general cost function, which contains the above as particular cases. We minimize

The authors are with the Department of Signal Theory and Communications, University Carlos III of Madrid, 28911 Leganes, Spain (e-mail: jlrojo@tsc.uc3m.es; manel@tsc.uc3m.es; arfv@tsc.uc3m.es; agarcia@tsc.uc3m.es; antonio@tsc.uc3m.es).
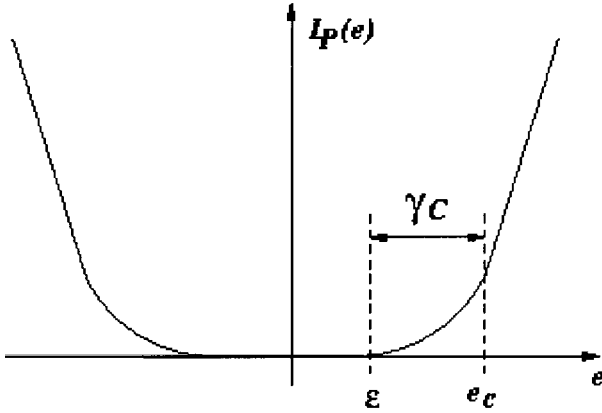
$$\frac{1}{2}\sum_{j=1}^{N_\omega}\left(c_j^2 + d_j^2\right) + \frac{1}{2\gamma}\sum_{k \in I_1}\left(\xi_k^2 + \xi_k^{*2}\right) + C\sum_{k \in I_2}\left(\xi_k + \xi_k^*\right) \quad (3)$$

Fig. 1. Relationship between the error and loss function $L_P(e)$.

with respect to $\{c_j\}$, $\{d_j\}$, $\{xi_k^{(*)}\}$, constrained to

$$y_{t_k} - \sum_{i=1}^{N_\omega} c_i \cos(\omega_i t_k) - \sum_{i=1}^{N_\omega} d_i \sin(\omega_i t_k) \le \varepsilon + \xi_k \quad (4)$$

$$-y_{t_k} + \sum_{i=1}^{N_\omega} c_i \cos(\omega_i t_k) - \sum_{i=1}^{N_\omega} d_i \sin(\omega_i t_k) \le \varepsilon + \xi_k^* \quad (5)$$

$$\xi_k, \xi_k^* \ge 0 \quad (6)$$

for $k = 1, \ldots, N$, where $\{\xi_k^{(*)}\}$[1] are the losses, ε is the insensitivity parameter, γ and C control the trade-off between the L_2 norm regularization of the coefficients and the losses, and I_1, I_2 are the sets of samples whose residuals are in the quadratic or in the linear zone of the cost function, respectively. Note that when γ is small enough this represents the regularized Vapnik's ε-insensitive cost, whereas for $\varepsilon = 0$ it represents Huber's robust cost. It can be easily shown that $e_C = \gamma C$.

The derivation of the dual problem shows that the Cartesian components can be expressed in the solution as

$$c_l^L = \sum_{k=1}^{N} (\alpha_k - \alpha_k^*) \cos(\omega_l t_k) \quad (7)$$

$$d_l^L = \sum_{k=1}^{N} (\alpha_k - \alpha_k^*) \sin(\omega_l t_k) \quad (8)$$

where $\{\alpha_k^{(*)}\}$ are the Lagrange multipliers for the constraints in (4) and (5). Therefore, the lth coefficients correspond to the cross-correlation of the Lagrange multipliers and the sinusoid with frequency ω_l. In [5], these conditions are introduced into the Lagrange functional in order to remove the primal variables, leading to a dual problem that is solved with QP procedures. The computational burden of this approach grows exponentially with the number of data, and the QP problem cannot support an adaptive version.

Further computational advantage can be obtained from an IRWLS-based alternative formulation. Let us denote

$$\mathbf{c} = [c_1, \ldots, c_{N_\omega}]^T \quad (9)$$

$$\mathbf{d} = [d_1, \ldots, d_{N_\omega}]^T \quad (10)$$

[1]In the following, $\{\xi_k^{(*)}\}$ will denote both $\{\xi_k\}$ and $\{\xi_k^*\}$, and the same notation is followed for $\{\alpha_k^{(*)}\}$ and $\{\beta_k^{(*)}\}$.

$$\mathbf{w} = [\mathbf{c}^T, \mathbf{d}^T]^T \quad (11)$$

$$\mathbf{x}_i = \begin{bmatrix} \cos(\omega_1 t_i) \\ \vdots \\ \cos(\omega_{N_\omega} t_i) \\ \sin(\omega_1 t_i) \\ \vdots \\ \sin(\omega_{N_\omega} t_i) \end{bmatrix} \quad (12)$$

$$\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]. \quad (13)$$

Then, the Lagrange functional [2] for the minimization of (3) constrained to (4)–(6) is given by

$$L_{PD} = \frac{1}{2}\|\mathbf{w}\| + \sum_{k=1}^{N} \alpha_k \left(y_{t_k} - \mathbf{w}^T\mathbf{x}_k - \varepsilon\right)$$
$$+ \sum_{k=1}^{N} \alpha_k^* \left(-y_{t_k} - \mathbf{w}^T\mathbf{x}_k - \varepsilon\right)$$
$$+ \frac{1}{2\gamma} \sum_{k\in I_1} \left(\xi_k^2 + \xi_k^{*2}\right) + \sum_{k\in I_2} C\left(\xi_k + \xi_k^*\right)$$
$$- \sum_{k=1}^{N} [(\beta_k\xi_k + \beta_k^*\xi_k^*) + \alpha_k\xi_k + \alpha_k^*\xi_k^*] \quad (14)$$

and it has to be minimized with respect to $\mathbf{w}$, $\{\xi_k^{(*)}\}$, and maximized with respect to $\{\alpha_k^{(*)}\}$, $\{\beta_k^{(*)}\}$, constrained to $\{\alpha_k^{(*)}\}$, $\{\beta_k^{(*)}\}$, $\{\xi_k^{(*)}\} \ge 0$. From the Karush–Khun–Tucker (KKT) conditions, we have $(\partial L_{PD})/(\partial \xi_k^{(*)}) = 0$, so that

$$C - \beta_k^{(*)} - \alpha_k^{(*)} = 0, \quad i \in I_1 \quad (15)$$

$$\frac{1}{\gamma}\xi_k^{(*)} - \beta_k^{(*)} - \alpha_k^{(*)} = 0, \quad i \in I_2 \quad (16)$$

and the terms of L_{PD} on $\xi_k^{(*)}$ can be removed because (15) and (16) must hold at the solution. If we denote $e_k = y_{t_k} - \mathbf{w}^T\mathbf{x}_k$ (i.e., e_k is the residual of the kth observation), and by denoting

$$\lambda_k = \frac{2\alpha_k}{y_{t_k} - \mathbf{w}^T\mathbf{x}_k - \varepsilon} = \frac{2\alpha_k}{e_k - \varepsilon} \quad (17)$$

$$\lambda_k^* = \frac{2\alpha_k^*}{-y_{t_k} + \mathbf{w}^T\mathbf{x}_k - \varepsilon} = \frac{2\alpha_k^*}{-e_k - \varepsilon} \quad (18)$$

the functional can be written down with the standard IRWLS formulation

$$L_{PD} = \frac{1}{2}\|\mathbf{w}\|^2 + \sum_{k=1}^{N} \left(\lambda_k(e_k - \varepsilon)^2 + \lambda_k^*(-e_k - \varepsilon)^2\right). \quad (19)$$

A straightforward relationship between the residuals and the Lagrange multipliers can be derived from the KKT conditions. In brief, for $e_k \in [0, \varepsilon]$, $\xi_k = 0$ and hence $\alpha_k = 0$. For $e_k \in [\varepsilon, e_C]$, (15) holds and $\alpha_k = \xi_k/\gamma = (e_k - \varepsilon)/\gamma$. For $e_k > e_C$, (16) holds and $\alpha_k = C$. This relationship and its corresponding for α_k^* are summarized in (21) and (22).

By making zero gradient $\nabla_\mathbf{w} L_p$, the following matrix-form expressed equation is obtained

$$\mathbf{w} = [\mathbf{I} + \mathbf{X}^T\mathbf{D}_{\lambda+\lambda^*}\mathbf{X}]^{-1}\mathbf{X}^T[\mathbf{D}_{\lambda+\lambda^*}\mathbf{y} - \mathbf{D}_{\lambda-\lambda^*}\varepsilon] \quad (20)$$

where $\mathbf{D}_\lambda$ denotes the diagonal matrix whose main diagonal kth element is λ_k. The samples for which $\lambda_k = 0$ can be removed from the calculations, hence reducing the computational burden. In the above step, $\lambda_k^{(*)}$ are treated as constants, as usual in IWRLS algorithms.

The fixed-point algorithm that is used for solving this IRWLS problem can be summarized as follows.

1) Start with an arbitrary $\mathbf{w}_0$. Set $r = 0$.
2) Calculate errors $e_k = y_{t_k} - \mathbf{w}_r^T \mathbf{x}_k$.
3) Calculate the Lagrange multipliers by

$$\alpha_k = \begin{cases} C, & e_k \geq e_C \\ \frac{1}{\gamma}(e_k - \varepsilon), & \varepsilon \leq e_k \leq e_C \\ 0, & \text{elsewhere} \end{cases} \quad (21)$$

$$\alpha_k^* = \begin{cases} C, & e_k \leq -e_C \\ \frac{1}{\gamma}(-e_k - \varepsilon), & -\varepsilon \geq e_k \geq -e_C \\ 0, & \text{elsewhere} \end{cases} \quad (22)$$

where $e_C = \varepsilon + \gamma C$.

4) Calculate λ_k, λ_k^* as given in (17) and (18).
5) Solve (20) to obtain $\mathbf{w}_{r+1}$.
6) Set $r = r + 1$. If $r \leq R$ (maxiter) then go to Step 2).

Note that the solution coefficients are obtained by the empirical cross correlation between the sinusoidal functions and the Lagrange multipliers [as given in (7) and (8)]. As the last ones are a nonlinear transformation of the residuals, given by (21) and (22), controlling the value of C allows to reduce the impact of an outlier in the solution.

III. Application Examples

A. Insensitivity to Outliers

A simple synthetic data example is first presented to show the capacity of SVM-Spect to deal with outliers. A discrete time process is given by

$$x_k = \sin(2\pi f k) + v_k + j_k \quad (23)$$

where $f = 0.3$ Hz; v_n is a white, Gaussian noise sequence with zero mean and variance $\sigma^2 = 0.1$; and j_n is an impulsive noise process, generated as a sparse sequence for which 30% of the samples, randomly placed, are high-amplitude values given by $\pm 10 + U(-0.5, 0.5)$ (where $U()$ denotes the uniform distribution in the given interval), and the remaining are null samples. The number of observed samples is $N = 128$, and we set $N_\omega = N/2 = 64$ [Fig. 2(a)].

In order to avoid too sparse solutions, $\varepsilon = 0$ is used. A low value of γ leads to a major emphasis on minimizing the losses, so that overfitting to the observations occurs in this case. We select a moderately high $\gamma = 10$.

The appropriate *a priori* choice of the free parameter C can be addressed by considering that, according to (7) and (8), the solution is a function (in fact, proportional to the empirical cross correlation) of the multipliers and the data. Also, (21) and (22) reveal that a high-amplitude residual, corresponding to an outlier, will produce a high-amplitude multiplier, which will distort the solution. But if the maximum value that the multiplier
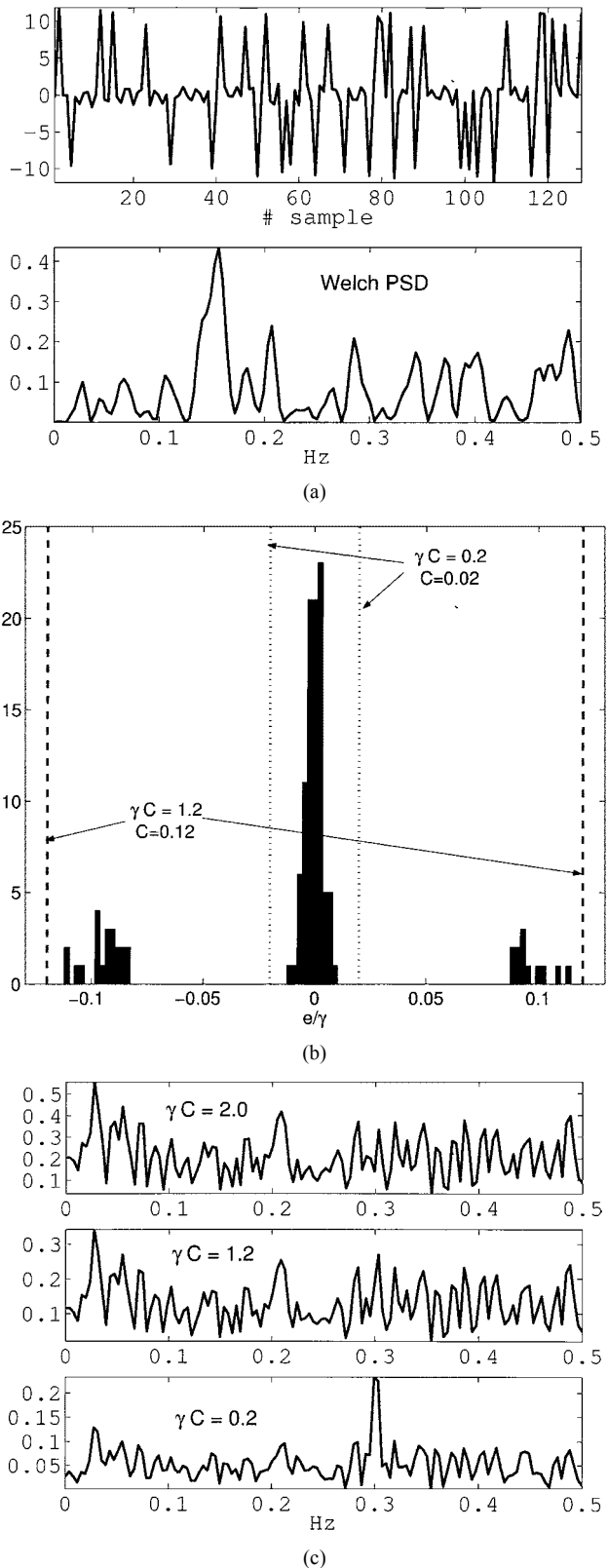

Fig. 2. Insensitivity of SVM-Spect to outliers. (a) Sinusoid whithin impulsive noise (up) and its Welch periodogram (down). (b) Histogram of the residuals (scaled to $\gamma = 10$) and control of the outlier impact onto the solution with C. (c) SVM-Spect spectral estimators for different values of insensitivity, which is controlled by the product γC.

can take is properly limited by C, the impact of the outlier on the solution is weakened. Fig. 2(b) shows that C should be low

enough to exclude the residual amplitudes greater than the base level. This level can be obtained from previous methods of estimation of the prediction error, from *a priori* knowledge of the problem, or from training data. Fig. 2(c) shows the results for $\gamma C = 2$, 1.2, and 0.2. Other experiments, not included here, allow to show that similar solutions are obtained for a range of values of γC being low enough.

B. ADSL Channel Estimation

A potential application where robustness of SVM-Spect can be exploited is ADSL channel estimation. Discrete multitone (DMT) modulation is used in ADSL because with the aid of cyclic prefix [10], the equalization of the channel is done simply by estimating the channel transfer function at the subcarrier frequency and dividing the received signal in each subcarrier by this estimation. Some of the DMT subcarriers are used to transmit known symbols (pilots) for channel estimation. In brief, a simplified model of an ADSL system using DMT parameters, as specified in ANSI T1-413 [11],[2] in a two-tap frequency-selective channel with additive white Gaussian noise and impulse noise, is simulated. A complete characterization of impulse noise can be found in [12] where probability density functions of voltages, interarrival times and impulse length are described. Since the simulation length involved in channel estimation is shorter than interarrival times, a simplified model is used with only one impulse following a Gaussian amplitude and fixed duration much shorter than the DMT symbol duration. The channel is estimated at different SNR and impulse noise conditions (measured by the standard deviation of impulse noise voltages σ_I) following two methods: the standard fast Fourier transform (FFT)-based averaging over two and four DMT symbols, and the SVM-Spect averaging over two symbols. For the last one, an optimum set of the free parameters (ε, C, and γ) are previously determined by cross-validation, and these parameters remain fixed. The training process will allow in this case to adapt to slow changes in the channel, and hence it is more convenient that fixing and freezing the free parameters *a priori*. The relative error is calculated at each realization as

$$\text{RE} = \int 20 \log \left| \frac{|H_e(f)| - |H(f)|}{|H(f)|} \right| df \qquad (24)$$

where $H_e(f)$ represents either the FFT or the SVM-Spect channel estimator, for a number of 500 realizations at each SNR explored value.

Fig. 3 shows that the SVM-Spect method outperforms the FFT-based channel estimation for low SNR values and impulses of high amplitude. In the upper figure, it can be seen that the percentage of subcarriers reserved for channel estimation (i.e., overhead) can be reduced from 20% to 10% while maintaining or even improving (for low SNR) the performance obtained with FFT-based estimation.
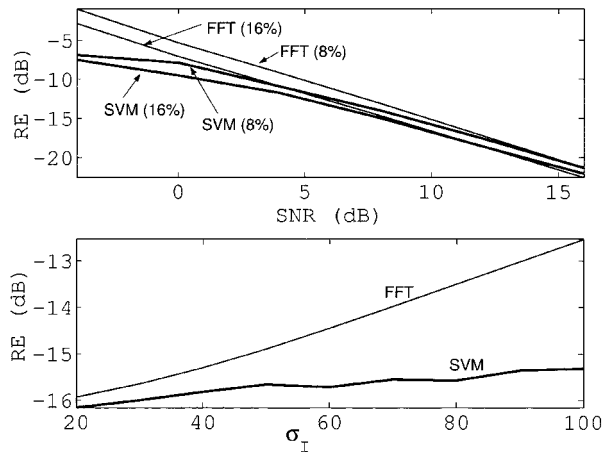
Fig. 3. ADSL channel estimation. (Up) Averaged relative error with SNR. (Down) Averaged relative error for SNR = 20 dB and impulsive noise in the ADSL channel.

IV. CONCLUSION

The application of the SVM to classical, nonparametric spectral analysis is a promising framework, given that insensitivity to outliers can be easily handled by controlling its free parameters on an easy way, and also, its WLS-based implementation allows to use it on real-time problems. The SVM spectral analysis is a potential robust approach to improve the performance of communication systems in impulsive noise environments.

REFERENCES

[1] S. Marple, *Digital Spectral Analysis With Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[2] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.

[3] B. Schölkopf, K. K. Sung, C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik, "Comparing support vector machines with Gaussian Kernels to radial basis function classifiers," *IEEE Trans. Signal Processing*, vol. 45, pp. 2758–2765, Nov. 1997.

[4] M. Pontil and A. Verri, "Support vector machines for 3D object recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, pp. 637–646, June 1998.

[5] J. Rojo-Álvarez, A. García-Alberola, M. Martínez-Ramón, M. Valdés, A. Figueiras-Vidal, and A. Artés-Rodríguez, "Support vector robust algorithms for nonparametric spectral analysis," in *Proc. ICANN*, Madrid, Spain, 2002.

[6] F. Pérez-Cruz, A. Navia-Vázquez, P. L. Alarcón-Diana, and A. Artés-Rodríguez, "An IRWLS procedure for SVR," in *Proc. EUSIPCO 2000*, Tampere, Finland, Sept. 2000.

[7] P. Laguna, G. Moody, and R. Mark, "Power spectral density of unevenly sampled data by least-square analysis: Performance and application to heart rate signals," *IEEE Trans. Biomed. Eng.*, vol. 45, pp. 698–715, June 1998.

[8] K. Müller, A. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen, and V. Vapnik, "Predicting time series with support vector machines," in *Advances in Kernel Methods—Support Vector Learning*, C. B. B. Schölkopf and A. Smola, Eds. Cambridge, MA: MIT Press, 1999, pp. 243–254.

[9] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[10] J. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Mag.*, vol. 28, pp. 5–14, May 1990.

[11] *Network and customer installation interfaces—Asymmetric digital subscriber line (ADSL) metallic interface*, ANSI T1-413, 1998.

[12] W. Henkel, T. Kessler, and H. Chung, "Coded 64-cap ADSL in an impulse-noise environment—modeling of impulse noise and first simulation results," *IEEE J. Select. Areas Commun.*, vol. 13, pp. 1611–1621, Dec. 1995.