Optimally Regularized Channel Tracking Techniques for Sequence Estimation Based on Cross-Validated Subspace Signal Processing

Massimiliano (Max) Martone, *Member, IEEE*

***Abstract*—New methods for maximum-likelihood sequence estimation based on the Viterbi algorithm (VA) are presented. In the proposed scheme, the channel estimator and the Viterbi processor operate concurrently. At any given time-step, the sequence provided to the channel estimator comes from the survivor with the best metric value. These already known modifications of the traditional decision-directed VA cause large variance in the estimated channel coefficients. In fact, sequences with a high error rate may be used to perform estimation, and also the adjustment term of the channel tracking algorithm may exhibit abrupt changes caused by a "survivor swap," (that is by the event in which a different survivor has the best metric at step n with respect to the step $n-1$). The proposed regularization procedure forces the channel vector to lie in the appropriate *a priori* known subspace: while the variance decreases, a certain amount of bias is introduced. The variance-bias tradeoff is then automatically adjusted by means of a cross-validation "shrinkage" estimator, which is at the same time optimal in a "small sample" predictive sum of squares sense and asymptically model mean-squared-error optimal. The method is shown by means of hardware experiments on a wide-band base station extremely more effective than per survivor processing, minimum survivor processing, and traditional decision-directed approaches.**

***Index Terms*—Adaptive maximum-likelihood sequence estimation, band-limited channel, equalization, frequency-selective time-varying channels, intersymbol interference.**

I. Introduction

THE BASEBAND segment of next-generation software radios requires significant performance improvement of the demodulator section of the transceiver. Typical propagation scenarios in cellular communications are characterized by possibly rapid frequency-selective fading. This makes necessary the use of advanced baseband signal processing techniques to mitigate distortion induced by intersymbol interference (ISI). While linear techniques to mitigate time-varying ISI are attractive due to their conceptual simplicity, it is well known and generally accepted that a nonlinear processor is definitely more effective. Within the family of nonlinear methods, the decision-feedback equalizer and the maximum-likelihood sequence estimation (MLSE) [17], [19], [20] are extremely popular and established approaches. MLSE is the optimum strategy for sequence estimation in the presence of a perfectly known channel and (white) Gaussian noise.

The fundamental three building blocks of an MLSE-based demodulator are: 1) front-end preprocessing; 2) channel estimator; and 3) tree search processor. There are several approaches to front-end preprocessing for MLSE (see [1]) at different levels of optimality and complexity, where one always tries to simplify the discrete-time model so that the symbol sequence is simply convolved with a finite-impulse response filter with additive white Gaussian noise. The tree-search processing method of choice is usually the Viterbi algorithm (VA), a breadth-first algorithm. Since the channel is *a priori* unknown, a symbol decision-directed method can be used as in [21] to perform estimation. However, this scheme implicitly assumes reliable symbol decisions which may not be available at low signal-to-noise ratio (SNR) and requires that the estimate is performed with a certain delay (the decision delay) with respect to the current state. An alternative solution is to accomplish independent branch coefficients computation (with no delay) for each survivor path [per survivor processing, (PSP)] [8] within the VA. However, in this case, the channel estimation algorithm looses optimality because there is a clear model mismatch (the assumed sequence for all but one survivor is not the transmitted one). In addition, the computational effort in the channel estimation of the PSP technique is practically unfeasible if a large support ISI channel is induced by significant delay spread of the frequency-selective fading channel. The minimum survivor processing (MSP) technique of [3] uses at every step only one channel estimate, but this estimate is derived using the sequence relative to the survivor with the best metric value. However, the channel coefficients are affected by large variance because of the following facts.

- If the survivor with the best metric at step $n+1$ is different from the survivor with the best metric at step n, the symbol sequence fed back to the channel estimator at step $n+1$ may differ significantly from the symbol sequence used in the previous step. We define this event a "survivor swap."
- The survivor with the best metric value may contain a number of errors in the sequence even if a survivor swap did not occur.

In [3], it was suggested that the introduction of a small delay in the channel estimation procedure mitigates the two problems. The use of this delayed channel estimate in the metric is not

Paper approved by B. L. Hughes, the Editor for Theory and Systems of the IEEE Communications Society. Manuscript received September 23, 1998; revised June 1, 1999. This paper was presented in part at the Ninth Virginia Tech/MPRG Symposium on Wireless Personal Communications, Blacksburg, VA, June 2–4, 1999.

The author is with the Telecommunications Group, Watkins–Johnson Company, Gaithersburg, MD 20878-1794 USA (e-mail: max.martone@wj.com).

Publisher Item Identifier S 0090-6778(00)00489-X.

technically correct, and again, significant performance degradation occurs. We present an MSP method that allows tracking the discrete-time channel response with no delay with respect to the current state of the VA. A regularization procedure forces the channel vector to lie in the appropriate *a priori* known subspace. The well-known variance-bias tradeoff [14] is then automatically adjusted by means of a cross-validation "shrinkage" estimator, which is at the same time optimal in a "small-sample" predictive sum of squares (PRESS) sense and asymptotically (model) mean-squared-error (MSE) optimal. The method is shown to be more effective than the PSP algorithm, the MSP approach, and the traditional decision-directed approach. The paper is organized as follows. In Section II and Section III, we quickly review the system model and existing MLSE techniques, respectively. In Section IV, we describe how to regularize the estimates of the coefficients and use them in the VA. In Section V, we describe an adaptive implementation, and in Section VI, we show results of hardware experiments in a realistic cellular system.

II. System Model

A complex baseband modulated signal can be represented as $\tilde{s}(t)=\sum_n a_n p_g(t-nT)$, where a_n are the complex symbols defining the signal constellation used for the particular digital modulation scheme.[1] The filter $p_g(t)$ is a shaping filter (usually a square root raised-cosine filter), and T is the signaling interval. The baseband continuous time representation of this signal at the receiver is

$$\tilde{r}(t)=\sum_n a_n \tilde{h}(t,\, t-nT)+\tilde{n}(t) \tag{1}$$

where $\tilde{h}(t,\tau)$ is the convolution of $\tilde{f}(t,\tau)$, the impulse response of a frequency-selective fading multipath channel, and $\tilde{g}(\tau)=\int_{-\infty}^{+\infty} p_g(\beta)p_g(\tau-\beta)d\beta$, the overall shaping filter (in [15] and [16] is a raised-cosine function), and $\tilde{n}(t)$ is additive white Gaussian noise. The received signal (1) is the input of an ideal low-pass filter (see Fig. 1) with bandwidth W Hz, such that the bandwidth of $\tilde{r}(t)$ W_r satisfies $W > W_r$. Sampling the output of the ideal low-pass filter at $R/T=2W$ rate (R is an integer), we have the discrete-time system whose output samples form a set of sufficient statistics for MLSE

$$y_n=\sum_{l=0}^{NR-1} h(n,\, n-l)\tilde{a}_l+\eta_n \tag{2}$$

where $\tilde{a}_n=\sum_l \delta(n-lR)a_l$ is the sequence of N symbols interleaved with $R-1$ zeroes and

$$\delta(n)=\begin{cases}1, & n=0\\ 0, & n\neq 0\end{cases} \qquad h(n,\,k)=\tilde{h}\left(\frac{nT}{R},\,\frac{kT}{R}\right)$$

and η_n are samples of a white Gaussian discrete-time process with $E\{\eta_n\eta_n^*\}=RN_0/T$ if N_0 is the two-sided power spectral density of the channel noise. Assuming that the effective support of $\tilde{g}(t)$ is $[-KT,\, KT]$ and that JT is the maximum duration of the fading channel delay profile, the effective time span of $\tilde{h}(t,\tau)$ is $(D+1)T/R=(J+2K+1)T$ which basically allows us to consider finite ISI over $D+1$ fractionally sampled samples

$$y_n=\boldsymbol{a}_n^T\boldsymbol{h}_n+\eta_n, \qquad n=0,\,1,\cdots,RN-1 \tag{3}$$

with $\boldsymbol{a}_n=[\tilde{a}_n,\cdots,\tilde{a}_{n-D}]^T$ and $\boldsymbol{h}_n=[h(n,\,0),\cdots,h(n,\,D)]^T$.

A. Notation

In the following derivations, vectors and matrices are bold. $\boldsymbol{M}^T$, $\boldsymbol{v}^T$, $\boldsymbol{M}^H$, $\boldsymbol{v}^H$ designate transposition and Hermitian for matrix $\boldsymbol{M}$ and vector $\boldsymbol{v}$, respectively. Complex conjugation for scalars, matrices, and vectors is indicated as u^*, $\boldsymbol{M}^*$, and $\boldsymbol{v}^*$, respectively, while notations $[\boldsymbol{M}]_{l,\,m}$ and $[\boldsymbol{v}]_k$ stand for the $l,\,m$ element of matrix $\boldsymbol{M}$ and the kth element of vector $\boldsymbol{v}$, respectively. We use the notation $\|\boldsymbol{v}\|=\sqrt{\sum_{i=1}^M |v_i|^2}$ for the 2-norm of the complex M-vector $\boldsymbol{v}=[v_1,\cdots,v_M]^T$, and $\boldsymbol{I}_M$ for the $M\times M$ identity matrix.

III. MLSE Techniques

According to the maximum-likelihood principle, the optimal metric corresponding to the hypothesis that the sequence $[\alpha_1,\cdots,\alpha_{RN-1}]^T$ was transmitted is

$$L_N=\sum_{n=0}^{NR-1} l(\boldsymbol{\alpha}_n,\,\boldsymbol{h}_n)=\sum_{n=0}^{NR-1}\left|y_n-\boldsymbol{\alpha}_n^T\boldsymbol{h}_n\right|^2 \tag{4}$$

where $\boldsymbol{\alpha}_n=[\alpha_n,\cdots,\alpha_{n-D}]^T$, and $l(\boldsymbol{\alpha}_n,\,\boldsymbol{h}_n)$ is the branch metric. The VA is widely used to efficiently perform the exhaustive search in the trellis of the hypotheses. In practice, the time-varying coefficients $\boldsymbol{h}_n$ are unknown and must be identified by means of an input/output system identification procedure. If the true transmitted sequence could be provided at any step by the VA, channel identification could be accomplished by minimization of the MSE $E\{|\epsilon_n(\boldsymbol{a}_n,\,\hat{\boldsymbol{h}}_n)|^2\}$ with respect to $\hat{\boldsymbol{h}}_n$, where $\epsilon_n(\boldsymbol{a}_n,\,\hat{\boldsymbol{h}}_n)=y_n-\boldsymbol{a}_n^T\hat{\boldsymbol{h}}_n$. Typical approaches to achieve real-time channel estimation are gradient-based methods and recursive least squares (RLS)-based methods. The typical recursion in both cases is

$$\hat{\boldsymbol{h}}_{n+1}=\hat{\boldsymbol{h}}_n+\boldsymbol{p}_n(\boldsymbol{a}_n,\,\hat{\boldsymbol{h}}_n) \tag{5}$$

where $\boldsymbol{p}_n(\boldsymbol{a}_n,\,\hat{\boldsymbol{h}}_n)=\mu\epsilon_n(\boldsymbol{a}_n,\,\hat{\boldsymbol{h}}_n)\boldsymbol{a}_n^*$, if the algorithm of choice is the least mean squares (LMS) and μ is a step-size parameter that controls the rate of adjustment.[2] However, since the true transmitted sequence at $\boldsymbol{a}_n$ is not available in practice, one is forced to use sequences formed by past decisions on symbols at the output of the Viterbi processor: estimation of the channel is delayed by Δ where Δ is the decision delay of the VA.

An alternative solution is to accomplish independent branch coefficients computation for each survivor path (PSP) [8] within the VA. Denoting $\boldsymbol{\alpha}_n$ as the sequence relative to a generic survivor path, one updates the survivor channel as

$$\hat{\boldsymbol{h}}_{n+1}=\hat{\boldsymbol{h}}_n+\boldsymbol{p}_n(\boldsymbol{\alpha}_n,\,\hat{\boldsymbol{h}}_n). \tag{6}$$

[1]It is $\pi/4$ differential quaternary phase-shift keying (DQPSK) in the US cellular time-division multiple-access (TDMA) standard [15], [16].

[2]Observe that in the case of an RLS approach, $\boldsymbol{p}_n(\boldsymbol{a}_n,\,\hat{\boldsymbol{h}}_n)=\lambda\epsilon_n(\boldsymbol{a}_n,\,\hat{\boldsymbol{h}}_n)\boldsymbol{K}_n^*$, where $\boldsymbol{K}_n=(\boldsymbol{P}_{n-1}\boldsymbol{a}_n/\lambda+\boldsymbol{a}_n^H\boldsymbol{P}_{n-1}\boldsymbol{a}_n)$ and $\boldsymbol{P}_n=\lambda^{-1}(\boldsymbol{P}_{n-1}-\boldsymbol{a}_n^H\boldsymbol{P}_{n-1}\boldsymbol{K}_n)$ with λ as the forgetting factor.

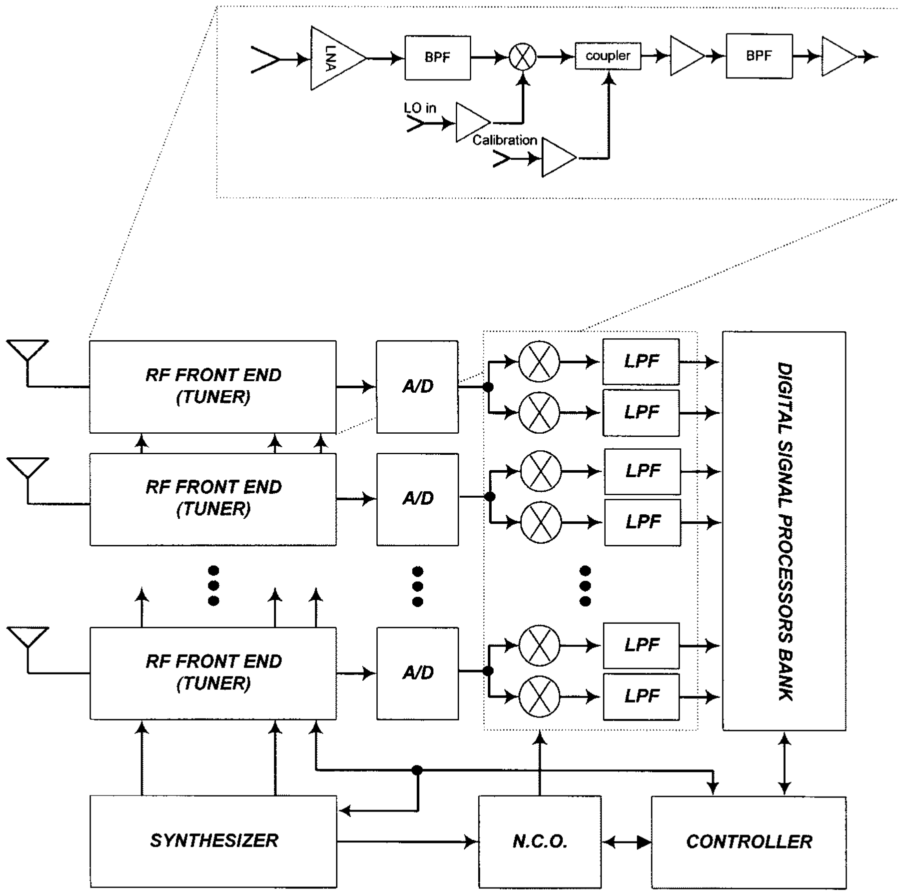

Fig. 1. A block diagram of the transceiver (only the receiver section is shown).

In essence, this approach is an extension of the VA because it implements basically the generalized likelihood detector [18]. An important criticism to the method is the fact that the channel estimation algorithm (gradient based or least squares) used in this particular context looses its optimality. It is in fact evident that the model assumed by the recursive identification algorithm for the generic survivor is

$$y_n = \boldsymbol{\alpha}_n^T \boldsymbol{h}_n + w_n \tag{7}$$

where the disturbance term

$$w_n = \eta_n + (\boldsymbol{a}_n - \boldsymbol{\alpha}_n)^T \boldsymbol{h}_n \tag{8}$$

is evidently non-Gaussian [3]. As a consequence, the adjustment term $\boldsymbol{p}_n(\boldsymbol{\alpha}_n, \hat{\boldsymbol{h}}_n)$ cannot be determined correctly to minimize the MSE. This loss of optimality is crucial because in the known channel VA, the paths diverge in the steps immediately preceding step n and quickly converge going back to the past. Errors in $\boldsymbol{\alpha}_n$ cause the estimated coefficients to be significantly different from $\boldsymbol{h}_n$. This in turn increases the probability of having path metrics in the vicinity of step n that are not distinguishable from the correct and maximum-likelihood one. In addition, the computational effort in the channel estimation of the PSP technique is practically unfeasible if a large support ISI channel is induced by significant delay spread of the frequency-selective fading channel. The MSP technique of [3] uses at every step only one channel estimate, but this estimate is derived using the sequence $\boldsymbol{\alpha}_n$ relative to the survivor with the best metric value. However, $\hat{\boldsymbol{h}}_n$ is a large variance estimate because of the following facts.

- If the survivor with the best metric at step $n+1$ is different from the survivor with the best metric at step n, the sequence $\boldsymbol{\alpha}_{n+1}$ fed back to the channel estimator at step $n+1$ may differ significantly from $\boldsymbol{\alpha}_n$, the symbol sequence used in the previous step. We define this event as "survivor swap."
- The survivor with the best metric value may contain a number of errors in the sequence $\boldsymbol{\alpha}_n$, even if a survivor swap did not occur.

In [3], it was suggested that the introduction of a delay $0 \leq d \leq \Delta$ in the channel estimation procedure mitigates the two problems. The use of $\hat{\boldsymbol{h}}_{n-d}$ in the metric function $l(\boldsymbol{a}_n, \hat{\boldsymbol{h}}_{n-d})$ is however not technically correct, and again, significant performance degradation occurs.

Prompted by the these observations, we look for a low-variance channel estimate as follows.

1) By forcing the channel estimates to lie in a known subspace; it is well known that by doing so, we trade variance for bias
2) By using cross validation as a regularization procedure to automatically (and optimally) adjust the tradeoff between variance and bias in the channel estimate.

IV. Optimal Channel Estimation Based on Cross Validation

The recursive channel estimation can be seen as

$$\hat{\boldsymbol{h}}_{n+1} = \sum_{i=1}^{n} \boldsymbol{p}_i(\boldsymbol{\alpha}_i, \hat{\boldsymbol{h}}_i). \tag{9}$$

Now observe that since in practice, $\hat{\boldsymbol{h}}_1$ is estimated using a known[3] preamble, each term in summation (9) depends on the initial condition $\hat{\boldsymbol{h}}_1$. We can rewrite (9) as

$$\hat{\boldsymbol{h}}_{n+1} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{h}_n(i) \tag{10}$$

with $\boldsymbol{p}_i(\boldsymbol{\alpha}_i, \hat{\boldsymbol{h}}_i) = (1/n)\boldsymbol{h}_n(i)$ and model the channel bias in each term as

$$\boldsymbol{h}_n(i) = \boldsymbol{h}_{n+1} + \boldsymbol{\xi}_n(i) \tag{11}$$

where $\boldsymbol{\xi}_n(i)$ is a random disturbance with *unknown* distribution. We will show that it is possible to design an estimator which is optimal for any "small" n in a PRESS, while being asymptotically optimal in the (model) mean squared sense.

A. Orthonormal Basis for $\boldsymbol{h}_n$

The first step is to design an orthonormal basis for $\boldsymbol{h}_n$ exploiting the "structure" of the channel impulse response $\tilde{h}(t, \tau)$. Observe that

$$\tilde{h}(t, \tau) = \int_0^{NT} \tilde{f}(t, \beta)\tilde{g}(\tau - \beta)\, d\beta \tag{12}$$

can be discretized, as shown in [4], by

$$\tilde{h}(t, \tau) \simeq T_I \sum_{l=0}^{M} \tilde{f}(t, lT_I)\tilde{g}(\tau - lT_I). \tag{13}$$

The accuracy of the approximation is arbitrary[4] (see [4]). The (fractionally) sampled channel then becomes

$$h(n, k) \simeq T_I \sum_{l=0}^{M} f_l(n) g_l(k) \tag{14}$$

where $g_l(k) = \tilde{g}(k(T/R) - lT_I)$ and $f_l(n) = \tilde{f}(n(T/R), lT_I)$. Neglecting the approximation error (which can be made arbitrarily small), it is possible to write (14) in vector form as

$$\boldsymbol{h}_n = \boldsymbol{G}\boldsymbol{f}_n \tag{15}$$

where $[\boldsymbol{G}]_{k+1, i+1} = T_I g_i(k)$ and $[\boldsymbol{f}_n]_{i+1} = f_i(n)$. Basically (15) says that $\boldsymbol{h}_n$ is in the subspace spanned by the columns of $\boldsymbol{G}$, which, being known, allows the selection of an appropriate orthonormal basis for the channel vector. Selection of a basis is one way of incorporating prior signal information into any estimation procedure. The value of estimating the signal as represented with respect to the basis is that the estimator is insensitive to disturbance components outside the space spanned by the basis, that is, the known subspace.

Consider the singular value decomposition (SVD) [7]) of $\boldsymbol{G} = \overline{\boldsymbol{U}}\boldsymbol{\Sigma}\boldsymbol{V}^H$, where $\overline{\boldsymbol{U}}$ is $D+1 \times D+1$, $\boldsymbol{\Sigma}$ is $D+1 \times M+1$, and $\boldsymbol{V}$ is $M+1 \times M+1$, and write

$$\boldsymbol{h}_n = \overline{\boldsymbol{U}}\,\overline{\boldsymbol{\theta}}_n \tag{16}$$

where $\overline{\boldsymbol{\theta}}_n = \boldsymbol{\Sigma}\boldsymbol{V}^H\boldsymbol{f}_n$. Consider $\boldsymbol{U} = [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_Q]^T$, the $D+1 \times Q$ matrix formed by the first Q columns (singular vectors) of $\overline{\boldsymbol{U}}$ corresponding to the significant singular values in $\boldsymbol{\Sigma}$. Applying the transformation

$$\boldsymbol{\theta}_n = \boldsymbol{U}^H\boldsymbol{h}_n \tag{17}$$

we define a "subspace parameter" $\boldsymbol{\theta}_n$. The Q orthonormal columns of $\boldsymbol{U}$ span a Q-dimensional subspace defined $\mathcal{U}_Q$, such that $\mathcal{U}_Q \subseteq \mathcal{C}^{D+1}$. Observe the following if the transformation $\boldsymbol{U}$ is selected with $Q < D+1$.

- The computational complexity is reduced because less parameters have to be estimated in the subspace parameter.
- More importantly, the estimate of $\boldsymbol{\theta}_n$ is insensitive to the disturbance components of $[\boldsymbol{\xi}_n(i)]_m$, for $m = Q+1, Q+2, \cdots, D+1$, and the resulting estimator has a smaller variance than the full space estimator of $\boldsymbol{h}_n$.

However, these two advantages are obtained at the expenses of estimation bias because in practice

$$\boldsymbol{h}_n = \boldsymbol{U}\boldsymbol{\theta}_n = \boldsymbol{P}_U\boldsymbol{h}_n \tag{18}$$

with $\boldsymbol{P}_U = \boldsymbol{U}\boldsymbol{U}^H$ holds only approximately true and while the variance increases with dimensions of the subspace the bias decreases [14]. Evidently, the goal of selecting a transformation $\boldsymbol{U}$ is to optimize the tradeoff between bias, variance, and computational load. The PRESS method can be seen as a method to automatically adjust the amount of bias and variance introduced by a given subspace estimation procedure [5]. In Fig. 2, we show the evident variance-bias paradigm for a 20-taps channel fading at 8 km/h estimated in a perfect training situation (that is with completely known data), using RLS, but using transformations projecting the parameters in subspaces with reduced dimensions. The figure shows the variance and the squared bias of the estimated $\hat{\boldsymbol{h}}_{n+1}$ using different orders of the subspace. Observe that the variance increases with the subspace order, while the bias exhibits the opposite behavior. In this particular case, the MSE of the coefficients exhibits a minimum, which allows advantages with respect to the full-rank estimator. In some other situations, one may desire less variance, while a slight increase of MSE is still tolerable.

Remark: Observe that if the discretization interval T_I is selected such that $M+1 < D+1$, then the subspace $\mathcal{U}_Q$ with $Q = M+1$ gives a close to zero bias. However, this subspace selection is not always possible in general. □

[3]In the absence of a preamble, using a blind method, the initialization of $\hat{\boldsymbol{h}}_1$ is relatively arbitrary.

[4]It must hold $(M+1)T_I = NT$. Accuracy is achieved by choosing a discretization interval T_I sufficiently small.

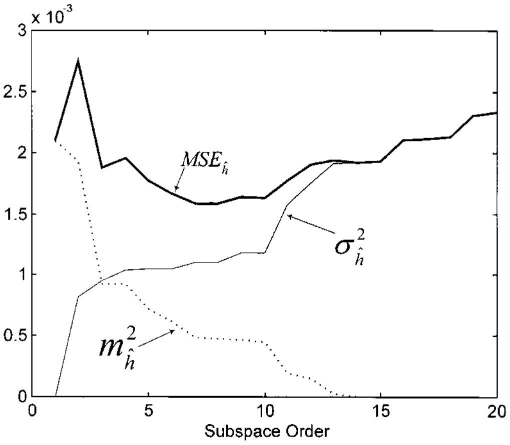

Fig. 2. The evident tradeoff between bias and variance in channel estimation. A 20-taps Rayleigh fading channel is estimated using RLS in reduced-order subspaces.

B. PRESS Method

Since we have selected a proper basis for $\hat{\boldsymbol{h}}_n$, it is immediate to see that

$$\hat{\boldsymbol{h}}_{n+1} = \boldsymbol{U}\hat{\boldsymbol{\theta}}_{n+1} \Rightarrow \hat{\boldsymbol{\theta}}_{n+1} = \boldsymbol{U}^H\hat{\boldsymbol{h}}_{n+1} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{U}^H\hat{\boldsymbol{h}}_n(i). \tag{19}$$

The estimator $\hat{\boldsymbol{\theta}}_{n+1}$ can also be expressed as the solution of a least squares problem

$$\hat{\boldsymbol{\theta}}_{n+1} = \arg\min_{\boldsymbol{\theta}_{n+1}} \left\| \hat{\boldsymbol{h}}_{n+1} - \boldsymbol{U}\boldsymbol{\theta}_{n+1} \right\|^2. \tag{20}$$

In an attempt to regularize the least squares solution, we formulate a "penalized" least squares (PLS) problem

$$\tilde{\boldsymbol{\theta}}_{n+1} = \arg\min_{\boldsymbol{\theta}_{n+1}} \left\{ \left\| \hat{\boldsymbol{h}}_{n+1} - \boldsymbol{U}\boldsymbol{\theta}_{n+1} \right\|^2 + \boldsymbol{\theta}_{n+1}^H \boldsymbol{\Lambda}_n \boldsymbol{\theta}_{n+1} \right\} \tag{21}$$

where $\boldsymbol{\Lambda}_n$ is a diagonal matrix and $[\boldsymbol{\Lambda}_n]_{i,i} = \lambda_{n,i} \geq 0$ are the regularization parameters that control the amount of penalization in each component of the least squares estimate. The term $\boldsymbol{\theta}_{n+1}^H \boldsymbol{\Lambda}_n \boldsymbol{\theta}_{n+1}$ is the "penalizing functional." Increasing $\lambda_{n,i}$ increases the cost associated with channel estimates having a large component in the corresponding subspace of the transform space spanned by the columns of $\boldsymbol{U}$. The unique solution to (21) is[5] well known [5]

$$\tilde{\boldsymbol{\theta}}_{n+1}(\boldsymbol{\Gamma}_n) = (\boldsymbol{I} + \boldsymbol{\Lambda})^{-1}\hat{\boldsymbol{\theta}}_{n+1} = \hat{\boldsymbol{\theta}}_{n+1} - \boldsymbol{\Gamma}_n\hat{\boldsymbol{\theta}}_{n+1} \tag{22}$$

where $\boldsymbol{\Gamma}_n$ is a diagonal matrix with $[\boldsymbol{\Gamma}_n]_{i,i} = (\lambda_{n,i}/1+\lambda_{n,i}) = \gamma_{n,i}$, and since $\lambda_{n,i}$ is a real nonnegative number, we have $\gamma_{n,i} \in [0, 1]$.

Remark: As noted by a reviewer, since the channel estimates are already forced to lie in the subspace spanned by $\boldsymbol{U}$, it would appear that there is no need for penalization. However, $\boldsymbol{h}_n = \boldsymbol{U}\boldsymbol{\theta}_n$ in our approach is only an efficient orthonormal representation for the parameter, not *the main mechanism to reduce variance*. Ideally, we would like further reduction of variance by eliminating more columns from $\boldsymbol{U}$ and using constrained solutions of the form

$$\hat{\boldsymbol{\theta}}'_{n+1} = \arg\min_{\boldsymbol{\theta}_{n+1}\in\Theta_{\tilde{Q}}} \left\| \hat{\boldsymbol{h}}_{n+1} - \boldsymbol{U}\boldsymbol{\theta}_{n+1} \right\|^2 \tag{23}$$

with $\Theta_{\tilde{Q}} = \{\boldsymbol{\theta}_{n+1}: \theta_{n+1,i} = 0, \tilde{Q} \leq i \leq Q\}$, for some $\tilde{Q} < Q$. However, further reduction of the subspace size increases bias. The penalized least squares solution makes an optimal tradeoff between these constrained estimators and the full subspace solution (20), in fact, we would obtain the estimator (23) from (22) if we selected $\gamma_{n,i} = 0$, for $i = 0, 1, \cdots, \tilde{Q}-1$, and $\gamma_{n,i} = 1$, for $i = \tilde{Q}, \cdots, Q$. □

If we define

$$\tilde{\boldsymbol{\theta}}_{n+1}(\boldsymbol{\Gamma}_n)_l = (\boldsymbol{I} - \boldsymbol{\Gamma}_n)\sum_{k=1,\,k\neq l}^{n}\boldsymbol{U}^H\boldsymbol{h}_n(k) \tag{24}$$

we can design a validation function to be minimized with respect to $0 \leq \gamma_{n,i} \leq 1$ as

$$\mathcal{V}(\boldsymbol{\Gamma}_n) = \sum_{l=1}^{n}\left\| \tilde{\boldsymbol{\theta}}_{n+1}(\boldsymbol{\Gamma}_n)_l - \boldsymbol{U}^H\boldsymbol{h}_n(l) \right\|^2 \tag{25}$$

which is a PRESS. Basically, $\tilde{\boldsymbol{\theta}}_{n+1}(\boldsymbol{\Gamma}_n)_l$ is a PLS estimate of $\boldsymbol{\theta}_{n+1}$ with the lth vector omitted. Then, this estimate is used to predict $\boldsymbol{U}^H\boldsymbol{h}_n(l)$. Summing the squares of the errors for any l, one obtains the PRESS function $\mathcal{V}(\boldsymbol{\Gamma}_n)$ which is a small-sample optimality criterion. The objective is to find

$$\boldsymbol{\Gamma}_n^{(\mathrm{PRESS})} = \arg\min_{\boldsymbol{\Gamma}_n}\mathcal{V}(\boldsymbol{\Gamma}_n). \tag{26}$$

A closed-form solution to the minimization problem can be found as in [5]. Define

$$\hat{\boldsymbol{\theta}}_{n+1} = \left[\hat{\theta}_{n+1,1} \quad \hat{\theta}_{n+1,2}\cdots\hat{\theta}_{n+1,Q}\right]^T$$
$$\tilde{\boldsymbol{\theta}}_{n+1} = \left[\tilde{\theta}_{n+1,1} \quad \tilde{\theta}_{n+1,2}\cdots\tilde{\theta}_{n+1,Q}\right]^T$$
$$\hat{\theta}_{n+1,i} = \frac{1}{n}\sum_{l=1}^{n}\boldsymbol{u}_i^H\boldsymbol{h}_n(l)$$

and

$$\sigma_{n,i}^2 = \frac{1}{n-1}\sum_{l=1}^{n}\left|\boldsymbol{u}_i^H\boldsymbol{h}_n(l) - \hat{\theta}_{n+1,i}\right|^2.$$

Setting the derivative of $\mathcal{V}(\boldsymbol{\Gamma}_n)$ with respect to $\gamma_{n,i}$ equal to zero, it is easy to obtain that

$$\gamma_{n,i}^{(\min)} = \frac{\hat{\sigma}_{n,i}^2}{\dfrac{\hat{\sigma}_{n,i}^2}{n} + (n-1)\left|\hat{\theta}_{n+1,i}\right|^2}.$$

It is evident that $\gamma_{n,i}^{(\min)} \geq 0$. If $0 \leq \gamma_{n,i}^{(\min)} \leq 1$, $\gamma_{n,i}^{(\min)}$ is the required minimizer of $\mathcal{V}(\boldsymbol{\Gamma}_n)$. If instead $\gamma_{n,i}^{(\min)} > 1$, we have that the required minimizer is equal to 1 because $\mathcal{V}(\boldsymbol{\Gamma}_n)$ is strictly

[5]Observe that (22) is a "shrinkage" estimator because it shrinks the estimate close to the origin.

TABLE I
GRADIENT PRESS-REGULARIZED ALGORITHM

INPUT	
--	$y_n, \boldsymbol{\alpha}_n, \tilde{\boldsymbol{\theta}}_n, \mathbf{U}, \mu, \beta, q, Q,$
SUBSPACE PROJECTION	
Step 1	$\mathbf{z}(\boldsymbol{\alpha}_n) = \mathbf{U}^T \boldsymbol{\alpha}_n$
Step 2	$\epsilon(\boldsymbol{\alpha}_n, \tilde{\boldsymbol{\theta}}_n) = y_n - \mathbf{z}(\boldsymbol{\alpha}_n)^T \tilde{\boldsymbol{\theta}}_n$
ADAPTIVE FILTERING	
Step 1	$\hat{\boldsymbol{\theta}}_{n+1} = \tilde{\boldsymbol{\theta}}_n + \mu\epsilon(\boldsymbol{\alpha}_n, \tilde{\boldsymbol{\theta}}_n)\mathbf{z}^*(\boldsymbol{\alpha}_n)$
REGULARIZATION	
Step 1 $i = 1, 2, ..., Q$	$\hat{\sigma}^2_{n,i} = \frac{1}{q-1}\sum_{l=1}^{q}\left\|\mu \mathbf{z}^*(\boldsymbol{\alpha}_{n-l+1})\epsilon(\boldsymbol{\alpha}_{n-l+1}, \tilde{\boldsymbol{\theta}}_{n-l+1}) - \hat{\boldsymbol{\theta}}_{n+1}\right\|^2,$
Step 2 $i = 1, 2, ..., Q$	$[\boldsymbol{\Gamma}_n]_{i,i} = \beta[\boldsymbol{\Gamma}_{n-1}]_{i,i} + \mathcal{T}\left(\frac{\hat{\sigma}^2_{n,i}}{\hat{\sigma}^2_{n,i}q^{-1} - (q-1)\|\hat{\theta}_{n+1,i}\|^2}\right)$
Step 3	$\tilde{\boldsymbol{\theta}}_{n+1} = (\mathbf{I} - \boldsymbol{\Gamma}_n)\hat{\boldsymbol{\theta}}_{n+1}.$

decreasing in $0 \le \gamma_{n,i} \le 1$ whenever $\hat{\sigma}^2_{n,i} > (\hat{\sigma}^2_{n,i}/n) + (n-1)|\hat{\theta}_{n+1,i}|^2$. The optimal choice for $\gamma_{n,i}$ is then

$$\gamma^{(\text{PRESS})}_{n,i} = \mathcal{T}\left(\frac{\hat{\sigma}^2_{n,i}}{\dfrac{\hat{\sigma}^2_{n,i}}{n} + (n-1)\left|\hat{\theta}_{n+1,i}\right|^2}\right) \tag{27}$$

where

$$\mathcal{T}(x) = \begin{cases} x, & x < 1 \\ 1, & x \ge 1. \end{cases}$$

Observe that the PRESS estimator has an interesting statistical interpretation: it is asymptotically equivalent to the estimator that minimizes $E\{|\theta_{n+1,i} - \tilde{\theta}_{n+1,i}|^2\}$, the model MSE. In fact, modeling $\boldsymbol{h}_{n+1}$ as deterministic and $\boldsymbol{\xi}_n(i)$ as a zero-mean random disturbance with unknown distribution and $E\{\boldsymbol{\xi}_n(i)^H \boldsymbol{\xi}_n(i)\} < \infty$, one obtains that the value of $\gamma_{n,i}$ minimizing $E\{|\theta_{n+1,i} - \tilde{\theta}_{n+1,i}|^2\}$ is

$$\gamma^{(\text{MSE})}_{n,i} = \frac{E\left\{\left|\hat{\theta}_{n+1,i}\right|^2\right\} - E\left\{\theta^*_{n+1,i}\hat{\theta}_{n+1,i}\right\}}{E\left\{\left|\hat{\theta}_{n+1,i}\right|^2\right\}} = \frac{\sigma^2_{n,i}}{\sigma^2_{n,i} + n|\theta_{n+1,i}|^2} \tag{28}$$

where $\sigma^2_{n,i} = E\{|\boldsymbol{u}_i^H(\boldsymbol{h}_n(i) - \boldsymbol{h}_{n+1})|^2\}$. It is proved in [6] that asymptotically $\gamma^{(\text{PRESS})}_{n,i} \Rightarrow \gamma^{(\text{MSE})}_{n,i}$.

V. Adaptive Implementation

The transformation $\boldsymbol{U}$ allows the modification of the original data model into

$$y_n = \boldsymbol{z}(\boldsymbol{a}_n)^T \boldsymbol{\theta}_n + \eta_n \tag{29}$$

where $\boldsymbol{z}(\boldsymbol{a}_n) = \boldsymbol{U}^T \boldsymbol{a}_n$ is the "transformed sequence." For the recursive procedure, the following occurs at every step.

- It obtains the estimate $\hat{\boldsymbol{\theta}}_{n+1}$, adjusting the regularized estimate of the previous step.
- It computes the regularization parameters.
- It applies the regularization to $\hat{\boldsymbol{\theta}}_{n+1}$.

The adaptation then becomes

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{n+1} &= \tilde{\boldsymbol{\theta}}_n + \boldsymbol{U}^H \boldsymbol{p}_n(\boldsymbol{\alpha}_n, \tilde{\boldsymbol{\theta}}_n) \\ &= \tilde{\boldsymbol{\theta}}_n + \mu\epsilon\left(\boldsymbol{\alpha}_n, \tilde{\boldsymbol{\theta}}_n\right)\boldsymbol{z}^*(\boldsymbol{\alpha}_n)\end{aligned} \tag{30}$$

where $\epsilon(\boldsymbol{\alpha}_n, \tilde{\boldsymbol{\theta}}_n) = y_n - \boldsymbol{z}(\boldsymbol{\alpha}_n)^T \tilde{\boldsymbol{\theta}}_n$. To obtain $\boldsymbol{\Gamma}_n$, we use a windowed estimate of $\hat{\sigma}^2_{n,i}$

$$\hat{\sigma}^2_{n,i} = \frac{1}{q-1}\sum_{l=1}^{q}\left|\mu \boldsymbol{z}^*(\boldsymbol{\alpha}_{n-l+1})\epsilon\left(\boldsymbol{\alpha}_{n-l+1}, \tilde{\boldsymbol{\theta}}_{n-l+1}\right) - \hat{\boldsymbol{\theta}}_{n+1}\right|^2$$

for $i = 1, 2, \cdots, Q$, where q is the size of the sliding window. The algorithm is summarized in Table I.

RLS methods can also be used to improve tracking performance. A particular modification of the QR-RLS (recursive least squares based on QR decomposition) algorithm is derived in the Appendix and summarized in Table II. The regularization is obtained by just replacing (30) with

$$\hat{\boldsymbol{\theta}}_{n+1} = \tilde{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{g}}(n+1)^* r_{n+1} \tag{31}$$

where $\tilde{\boldsymbol{g}}(n+1)$, r_{n+1} are obtained as described in the Appendix. To obtain $\boldsymbol{\Gamma}_n$, we use the estimate of $\hat{\sigma}^2_{n,i}$

$$\hat{\sigma}^2_{n,i} = \frac{1}{q-1}\sum_{l=1}^{q}\left|-\tilde{\boldsymbol{g}}(n-l+2)^* r_{n-l+2} - \hat{\boldsymbol{\theta}}_{n+1}\right|^2$$

for $i = 1, 2, \cdots, Q$. In both methods, the size q of the sliding window has to be designed in such a way to accomodate the time-varying characteristics of the channel, moreover the regularization parameters are computed as

$$\gamma^{(\text{PRESS})}_{n,i} = \mathcal{T}\left(\frac{\hat{\sigma}^2_{n,i}}{\dfrac{\hat{\sigma}^2_{n,i}}{q} - (q-1)\left|\hat{\theta}_{n+1,i}\right|^2}\right). \tag{32}$$

In rapidly fading environments, the size of the window q has to be only a few samples, which causes the estimates of $\gamma^{(\text{PRESS})}_{n,i}$ to be extremely noisy. A practical approach is to filter the estimated regularization parameters with a one-pole filter having

TABLE II
QR-RLS PRESS-REGULARIZED ALGORITHM

INPUT	
--	$y_n,\ \boldsymbol{\alpha}_n,\ \widetilde{\boldsymbol{\theta}}_n,\ \mathbf{U},\ \mathbf{V}^{(n)},\ \lambda,\ \beta,\ q,\ Q,$
SUBSPACE PROJECTION	
Step 1	$\mathbf{z}(\boldsymbol{\alpha}_n) = \mathbf{U}^T\boldsymbol{\alpha}_n$
Step 2	$\epsilon_n = y_n - \mathbf{z}(\boldsymbol{\alpha}_n)^T\widetilde{\boldsymbol{\theta}}_n$
ADAPTIVE FILTERING	
Step 1	$\bar{\mathbf{V}}^{(n)} = \begin{bmatrix} \lambda\mathbf{V}^{(n)} & 0 & \mathbf{V}^{(n)^{-H}}\lambda^{-1} \\ \mathbf{z}(\boldsymbol{\alpha}_n)^T & \epsilon_n & \mathbf{0}^T \end{bmatrix}$
Step 2	$\bar{\mathbf{V}}_1^{(n)} = \bar{\mathbf{V}}^{(n)}$
Step 3: Rotations	
for $l = 1, 2, \ldots, Q$ do	$a = \left[\bar{\mathbf{V}}_l^{(n)}\right]_{l,l},\quad b = \left[\bar{\mathbf{V}}_l^{(n)}\right]_{Q+1,l}$ $c_l = \frac{a}{\sqrt{\lvert a\rvert^2+\lvert b\rvert^2}},\quad s_l = \frac{b^*}{\sqrt{\lvert a\rvert^2+\lvert b\rvert^2}}$ $\tilde{\mathbf{Q}}^{(l)} = \mathbf{I}_{Q+1}$ $\left[\tilde{\mathbf{Q}}^{(l)}\right]_{Q+1,Q+1} = c_l,\ \left[\tilde{\mathbf{Q}}^{(l)}\right]_{l,Q+1} = s_l,$ $\left[\tilde{\mathbf{Q}}^{(l)}\right]_{Q+1,l} = -s_l^*,\ \left[\tilde{\mathbf{Q}}^{(l)}\right]_{l,l} = c_l,$ $\bar{\mathbf{V}}_{l+1}^{(n)} = \tilde{\mathbf{Q}}^{(l)}\bar{\mathbf{V}}_l^{(n)}$
end	
Step 4: Partition $\bar{\mathbf{V}}_{Q+1}^{(n)}$	$\bar{\mathbf{V}}_{Q+1}^{(n)} \Rightarrow \begin{bmatrix} \mathbf{V}^{(n+1)} & \overline{\mathbf{Y}}^{(n+1)} & \mathbf{V}^{(n+1)^{-H}} \\ \mathbf{0}^T & r_{n+1} & \tilde{\mathbf{g}}(n+1)^T \end{bmatrix}.$
Step 5	$\widehat{\boldsymbol{\theta}}_{n+1} = \widetilde{\boldsymbol{\theta}}_n - \tilde{\mathbf{g}}(n+1)^* r_{n+1}$
REGULARIZATION	
Step 1 $i = 1, 2, \ldots, Q$	$\widehat{\sigma}^2_{n,i} = \frac{1}{q-1}\sum_{l=1}^{q}\left\lvert -\tilde{\mathbf{g}}(n-l+2)^* r_{n-l+2} - \widehat{\boldsymbol{\theta}}_{n+1}\right\rvert^2,$
Step 2 $i = 1, 2, \ldots, Q$	$[\boldsymbol{\Gamma}_n]_{i,i} = \beta[\boldsymbol{\Gamma}_{n-1}]_{i,i} + \mathcal{T}\left(\frac{\widehat{\sigma}^2_{n,i}}{\widehat{\sigma}^2_{n,i}q^{-1} - (q-1)\lvert\hat{\theta}_{n+1,i}\rvert^2}\right)$
Step 3	$\widetilde{\boldsymbol{\theta}}_{n+1} = (\mathbf{I} - \boldsymbol{\Gamma}_n)\widehat{\boldsymbol{\theta}}_{n+1}.$

response $H(z) = 1/(1-\beta z^{-1})$. This gives the following update:

$$\gamma_{n,i}^{(\mathrm{PRESS})} = \beta\gamma_{n-1,i}^{(\mathrm{PRESS})} + \mathcal{T}\left(\frac{\hat{\sigma}^2_{n,i}}{\frac{\hat{\sigma}^2_{n,i}}{q} - (q-1)\left\lvert\hat{\theta}_{n+1,i}\right\rvert^2}\right). \tag{33}$$

Observe that the "transformed" branch metric is

$$l\left(\boldsymbol{U}^T\boldsymbol{\alpha}_n, \tilde{\boldsymbol{\theta}}_n\right) = \left\lvert y_n - \boldsymbol{z}(\boldsymbol{\alpha}_n)^T\tilde{\boldsymbol{\theta}}_n\right\rvert^2 \tag{34}$$

which can be used directly into the VA.

VI. PERFORMANCE ANALYSIS RESULTS

The cellular TDMA system under analysis is IS-136 [15], [16]. Hardware experiments are performed using the Watkins–Johnson wide-band dual-mode (AMPS and IS-136) base station system $Base_2$. A block diagram of the receiver section of the base station is shown in Fig. 1. The slots are 162 symbols long and the symbol period is 41.2 ms. Observe that

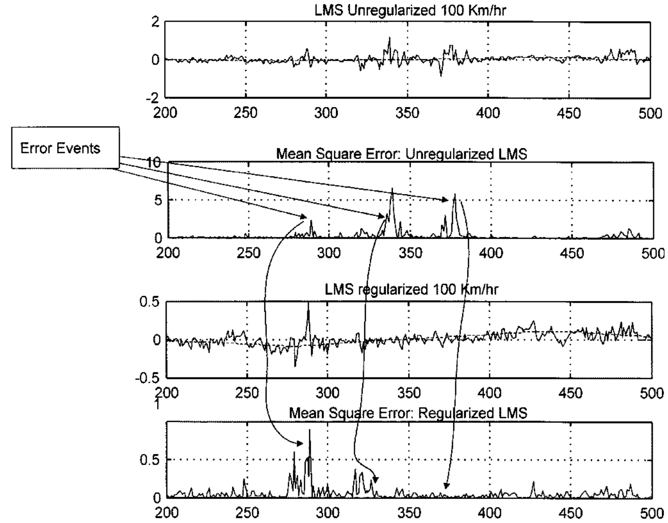

Fig. 3. Tracking a two-ray Rayleigh faded channel at 100 km/h with no delay between the VA and the channel estimator using the MSP technique.

there are two antennas. The metric has to be modified for the two independent fading branches as in the following:

$$L_N = \sum_{n=0}^{NR-1}\sum_{i=1}^{2}\left\lvert y_n^{(i)} - \boldsymbol{\alpha}_n^T\boldsymbol{h}_n^{(i)}\right\rvert^2 \tag{35}$$

where $y_n^{(i)}$, $\boldsymbol{h}_n^{(i)}$ are signal samples and discrete-time channel from the ith antenna $i = 1, 2$. The dispersive channel is created using a hardware fading emulator which models a two-ray Rayleigh fading channel [15], [16]. The delay interval is the difference in time of arrival between the two rays at each antenna. The Doppler frequency is related through wavelength $\tilde{\lambda}$ to the ith mobile transmitter velocity V_i expressed in km/h. We compare the following approaches.

- CLMS: Channel estimation with a single LMS channel estimator, which is delayed $d = \Delta = 6$ symbols. This is the typical approach of [2].
- CQRLS: Channel estimation with a single RLS based on QR decomposition (QR-RLS) channel estimator, which is delayed $d = \Delta = 6$ symbols.
- PSPLMS: Channel estimation with one LMS channel estimator per survivor using the PSP technique with $d = 0$ [8].
- PSPQRLS: Channel estimation with one QR-RLS channel estimator per survivor using the PSP technique with $d = 0$.
- MSPLMS: Channel estimation with a single LMS channel estimator using the MSP technique with $d = 2$. This is the approach of [3].
- MSPQRLS: Channel estimation with a single QR-RLS channel estimator using the MSP technique with $d = 2$.
- CVMSPLMS(Q): Channel estimation with a single regularized LMS channel estimator using the MSP technique with $d = 0$ and order of the subspace equal to Q.
- CVMSPQRLS(Q): Channel estimation with a single regularized QR-RLS channel estimator using the MSP technique with $d = 0$ and order of the subspace equal to Q.

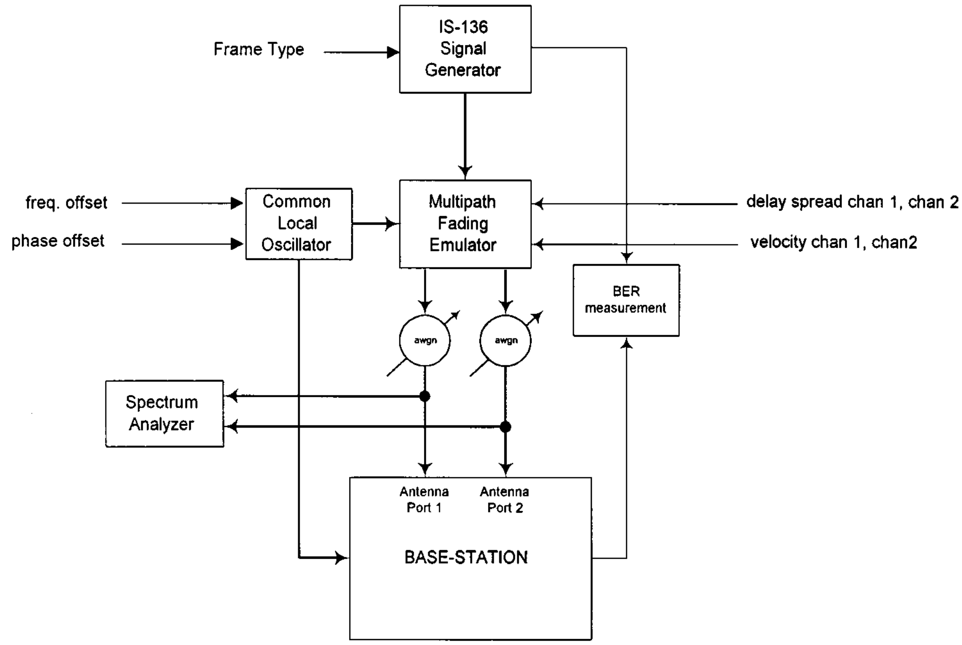

Fig. 4. Hardware test setup for lab experiments.

The value of Q order of the "representation subspace" for the channel can be typically selected as the number of singular values in the SVD of $\boldsymbol{G}$ with magnitude larger than a small number ranging between 0.01 and 0.001. We report the performance of the method with different values of Q. The oversampling factor is $R = 4$. Fig. 3 shows experiments where we try to investigate the tracking capability of the method in a computer simulation experiment. The plots show the trajectory of the real part of the maximum energy fading coefficient compared to the estimated one. The speed of the mobile is 100 km/h and the delay interval is 20.6 μs. We compare CVMSPLMS(4) with MSPLMS. The SNR per bit is equal to 20 dB. The size of the sliding window for computation of the regularization parameters is $q = 5$. Observe the relative resistance to error events exhibited by the cross-validated subspace approach. The following figures describe only hardware implementation results. The hardware test setup is depicted in Fig. 4. The IS-136 signal generator simulates transmission of digital traffic channel (DTC) frames coming from three different mobiles. Additive Gaussian noise is injected on both diversity channels. Observe that the DSP modem receives a sampling rate of 80 kHz after RF conversion and digital quadrature downconversion. The wordlength used is 16, and the algorithm has been implemented using analog devices fixed-point DSP processors. A polyphase raised-cosine filter (rolloff factor equal to 0.35) concatenated with an adjacent interference rejection filter transforms the rate to $4/T = 97.2$ kHz. Automatic gain control is operated on a slot-by-slot basis. An open-loop synchronizer chooses the optimum positioning of the metric computer with a resolution of $4/T$ rate. In addition, as required by the specification [15], [16], there is carrier frequency offset between the local oscillator and the transmitted carrier frequency of about 213 Hz. Frequency offset is estimated and removed using a second-order phase-locked loop. A detailed description of the synchronizer and the frequency offset compensator is omitted because it is beyond the scope of this paper. The trellis has four states because the span of the ISI channel is two symbol periods, which means that it is assumed $D + 1 = 8$. The complexity of each algorithm is reported in Table III in MIPS (million of instructions per second) for a two-antenna receiver. The complexity is relevant to the processing of one single reverse digital traffic channel. The above algorithms have been implemented in analog devices ADSP-2181 processors (40-MHz clock rate, 16 bits fixed point arithmetic) with the exclusion of the PSP algorithms that have been implemented in simulated fixed-point C-language and executed on baseband samples collected from the same hardware platform. Bit-error rate (BER) analysis results are shown in Figs. 5–7. Fig. 5 shows results at 100 km/h versus delay interval, while the SNR per bit is 25 dB. Fig. 6 is for delay interval equal to 20.6 μs, varying the mobile speed and SNR per bit equal to 22 dB. Fig. 7 depicts results for the LMS case with the MSP technique for both regularized and not. It is evident that there is a significant advantage using the cross-validated method.

TABLE III
COMPUTATIONAL COMPLEXITY WITH TWO ANTENNAS, FOUR STATES, VA OVERSAMPLING FACTOR 4, ISI SPAN OF TWO SYMBOLS

Algorithm	MIPS (Millions of Instructions per Second)
CLMS	20
CQRLS	34
PSPLMS	37
PSPQRLS	43
MSPLMS	26
MSPQRLS	32
CVMSPLMS(6)	27
CVMSPQRLS(6)	33

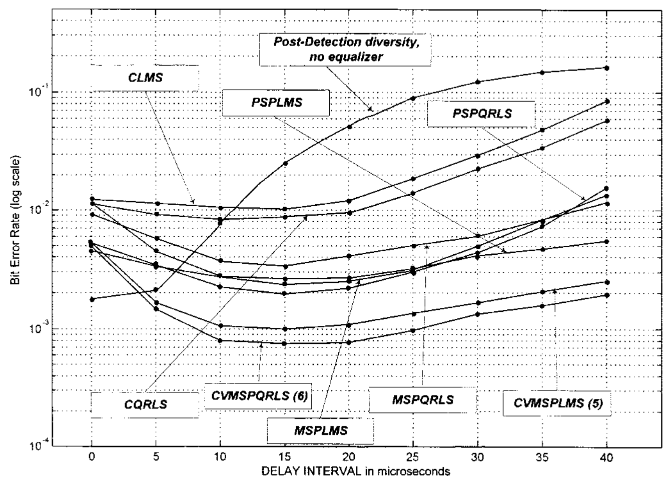

Fig. 5. BER at SNR per bit equal to 25 dB and 100 km/h versus delay interval.

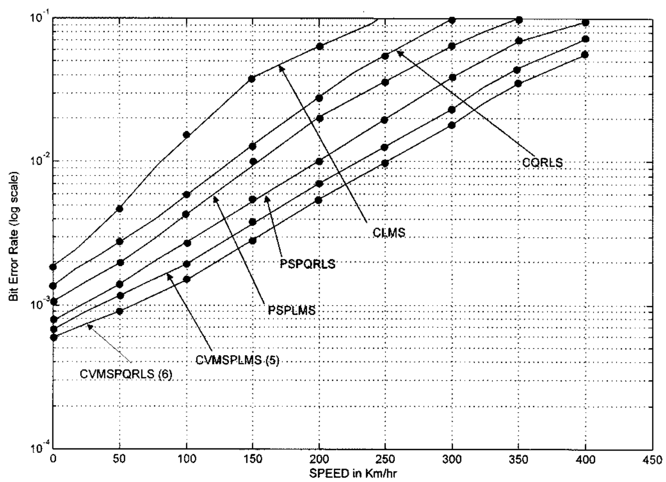

Fig. 6. BER at SNR per bit equal to 22 dB and 20.6 μs delay interval and versus mobile speed.

VII. Conclusions

We have studied a new practical solution to the adaptive channel estimation problem in MLSE-based receivers. In the proposed scheme, the channel estimator and the trellis-search processor operate concurrently. At any given time-step, the sequence provided to the channel estimator comes from the survivor with the best metric value. These already known modifications of the traditional decision-directed Viterbi-based equalizer cause large variance in the estimated channel coefficients, because sequences with high error rate may be used to perform estimation, and because the adjustment term of the channel tracking algorithm may exhibit abrupt changes caused by a "survivor swap," that is, by the event in which a different survivor has best metric at step n with respect to step $n-1$. By forcing the channel estimates to lie in a known subspace and using cross validation as a regularization procedure, we automatically (and optimally) adjust the amount of variance and bias that contributes to the MSE of the channel estimate. Specializing the results to the current U.S. standard IS-136 [15], [16] and to specific hardware implementations of the base station transceiver, we presented the results of an extensive investigation (in part also presented in [9], [11]–[13]) whose

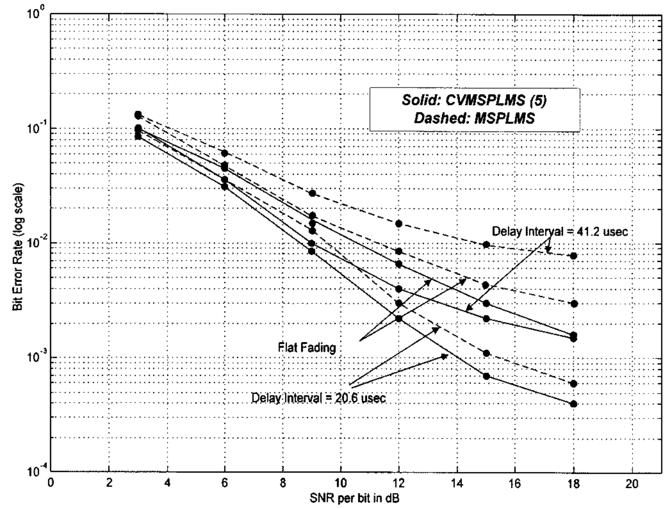

Fig. 7. BER at 0-20.6-41.2 μs delay interval for LMS-based MSP techniques.

goal was to select the most appropriate strategy for equalization in a diversity receiver for field operations. Our results indicate that the MSP technique of [3] with no delay between channel tracker and Viterbi processor, in conjunction with the novel subspace-regularized approach presented, outperforms the traditional decision-directed approach and the PSP method and is superior to these last two even if regularization is applied to them.

Appendix

We have shown in [10] that at each $n+1$ step, the problem

$$\min_{\boldsymbol{\theta}_{n+1}} \left\| \begin{bmatrix} \lambda \tilde{\boldsymbol{Z}}^{(n)} \\ \boldsymbol{z}(\boldsymbol{\alpha}_{n+1})^T \end{bmatrix} \boldsymbol{\theta}_{n+1} - \begin{bmatrix} \lambda \boldsymbol{Y}^{(n)} \\ y_{n+1} \end{bmatrix} \right\|^2 \tag{36}$$

where λ is the *forgetting factor* [7]

$$\tilde{\boldsymbol{Z}}^{(n)^T} = [\boldsymbol{z}(\boldsymbol{\alpha}_1), \cdots, \boldsymbol{z}(\boldsymbol{\alpha}_n)]\boldsymbol{\Lambda}(n)$$

is the data matrix, and $\boldsymbol{Y}^{(n)^T} = [y_1, y_2, \cdots, y_n]\boldsymbol{\Lambda}(n)$ with $\boldsymbol{\Lambda}(n) = \mathrm{diag}[\lambda^{n-1}, \lambda^{n-2}, \cdots, 1]$ is equivalent to solving for $\partial\boldsymbol{\theta}_n = \boldsymbol{\theta}_{n+1} - \boldsymbol{\theta}_n$ the problem

$$\min_{\partial\boldsymbol{\theta}_n} \left\| \begin{bmatrix} \boldsymbol{V}^{(n+1)} \\ \mathbf{0}^T \end{bmatrix} \partial\boldsymbol{\theta}_n - \boldsymbol{Q}(n) \begin{bmatrix} \mathbf{0} \\ \epsilon_{n+1} \end{bmatrix} \right\|^2. \tag{37}$$

The orthogonal matrix $\boldsymbol{Q}(n)$ is such that

$$\boldsymbol{Q}(n)\tilde{\boldsymbol{Z}}^{(n+1)} = \begin{bmatrix} \boldsymbol{V}^{(n+1)} \\ \mathbf{0}^T \end{bmatrix}$$

where $\boldsymbol{V}^{(n+1)}$ is upper triangular. As a consequence, one can append a new data row $\boldsymbol{z}(\boldsymbol{\alpha}_{n+1})^T$ to $\boldsymbol{V}^{(n)}$, sweep the last row of the augmented matrix by means of an orthogonal transformation $\tilde{\boldsymbol{Q}}(n)$, apply the same transformation to the vector $[\boldsymbol{0}^T, \epsilon_{n+1}]^T$, and solve a $Q \times Q$ triangular system by backsubstitution (for details, see [10]). $\tilde{\boldsymbol{Q}}(n)$ can be found as a product of Q Givens rotation matrices [7]. Further, one can avoid the backsubstitution (implicitly a serial process) to enhance parallelism. The matrix factorization lemma [7, p. 590] says that for any two matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ related to $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ by orthogonal transformations, that is

$$\tilde{\boldsymbol{Q}}[\boldsymbol{A}_1 \quad \boldsymbol{A}_2] = [\boldsymbol{B}_1 \quad \boldsymbol{B}_2] \tag{38}$$

where $\tilde{\boldsymbol{Q}}^H\tilde{\boldsymbol{Q}} = \boldsymbol{I}$, it holds the following fundamental relation

$$\boldsymbol{B}_1^H\boldsymbol{B}_2 = \boldsymbol{A}_1^H\boldsymbol{A}_2. \tag{39}$$

Fact 1: The inverse hermitian of $\boldsymbol{V}^{(n)}$ (indicated as $\boldsymbol{V}^{(n)^{-H}}$) can be recursively computed using

$$\tilde{\boldsymbol{Q}}(n)\begin{bmatrix}\boldsymbol{V}^{(n)^{-H}}\lambda^{-1}\\ \boldsymbol{0}^T\end{bmatrix} = \begin{bmatrix}\boldsymbol{V}^{(n+1)^{-H}}\\ \tilde{\boldsymbol{g}}(n+1)^T\end{bmatrix}. \tag{40}$$

Proof: Consider the factorization lemma with

$$\boldsymbol{A}_1 = \begin{bmatrix}\boldsymbol{V}^{(n)^{-H}}\lambda^{-1}\\ \boldsymbol{0}^T\end{bmatrix} \qquad \boldsymbol{A}_2 = \begin{bmatrix}\lambda\boldsymbol{V}^{(n)}\\ \boldsymbol{z}(\boldsymbol{\alpha}_{n+1})^T\end{bmatrix}$$

and apply the orthogonal transformation to $[\boldsymbol{A}_1 \quad \boldsymbol{A}_2]$. It is possible to write

$$\tilde{\boldsymbol{Q}}(n)\begin{bmatrix}\boldsymbol{V}^{(n)^{-H}}\lambda^{-1} & \lambda\boldsymbol{V}^{(n)}\\ \boldsymbol{0}^T & \boldsymbol{z}(\boldsymbol{\alpha}_{n+1})^T\end{bmatrix} = \begin{bmatrix}\boldsymbol{V}^{(n+1)^{-H}} & \boldsymbol{V}^{(n+1)}\\ \tilde{\boldsymbol{g}}(n+1)^T & \mathbf{o}^T\end{bmatrix}.$$

It is then evident that by writing down (39) for

$$\boldsymbol{B}_1 = \begin{bmatrix}\boldsymbol{V}^{(n+1)^{-H}}\\ \tilde{\boldsymbol{g}}(n+1)^T\end{bmatrix} \qquad \boldsymbol{B}_2 = \begin{bmatrix}\boldsymbol{V}^{(n+1)}\\ \boldsymbol{0}^T\end{bmatrix}$$

we get

$$\begin{aligned}&\begin{bmatrix}\boldsymbol{V}^{(n)^{-1}}\lambda^{-1} & \boldsymbol{0}^T\end{bmatrix}\begin{bmatrix}\lambda\boldsymbol{V}^{(n)}\\ \boldsymbol{z}(\boldsymbol{\alpha}_{n+1})^T\end{bmatrix}\\ &\quad = \begin{bmatrix}\boldsymbol{V}^{(n+1)^{-1}} & \tilde{\boldsymbol{g}}(n+1)^*\end{bmatrix}\begin{bmatrix}\boldsymbol{V}^{(n+1)} & \boldsymbol{0}^T\end{bmatrix}\end{aligned} \tag{41}$$

that is, $\boldsymbol{V}^{(n)^{-1}}\boldsymbol{V}^{(n)} = \boldsymbol{V}^{(n+1)^{-1}}\boldsymbol{V}^{(n+1)} = \boldsymbol{I}_Q$. □

Fact 2: $\partial\boldsymbol{\theta}_n$ is obtained as

$$\partial\boldsymbol{\theta}_n = -\tilde{\boldsymbol{g}}(n+1)^* r_{n+1} \tag{42}$$

where $\tilde{\boldsymbol{g}}(n+1)$ is obtained from (40) and r_{n+1} is obtained as shown in Table II.

Proof: We use again the matrix factorization lemma with $\boldsymbol{A}_1$ and $\boldsymbol{B}_1$ defined as in the proof of Theorem 1 while we use now

$$\boldsymbol{A}_2 = \begin{bmatrix}\boldsymbol{0}\\ \epsilon_{n+1}\end{bmatrix} \qquad \boldsymbol{B}_2 = \begin{bmatrix}\overline{\boldsymbol{Y}}^{(n+1)}\\ r_{n+1}\end{bmatrix}.$$

Using again (39) (because obviously (38) is verified), we can write

$$\begin{aligned}&\begin{bmatrix}\boldsymbol{V}^{(n)^{-1}}\lambda^{-1} & \boldsymbol{0}\end{bmatrix}\begin{bmatrix}0\\ \epsilon_{n+1}\end{bmatrix}\\ &\quad = \begin{bmatrix}\boldsymbol{V}^{(n+1)^{-1}} & \tilde{\boldsymbol{g}}(n+1)^*\end{bmatrix}\begin{bmatrix}\overline{\boldsymbol{Y}}^{(n+1)}\\ r_{n+1}\end{bmatrix}\end{aligned} \tag{43}$$

which gives

$$\begin{aligned}&\overbrace{\boldsymbol{V}^{(n+1)^{-1}}\overline{\boldsymbol{Y}}^{(n+1)}}^{\partial\boldsymbol{\theta}_n} + \tilde{\boldsymbol{g}}(n+1)^* r_{n+1}\\ &\quad = \lambda^{-1}\boldsymbol{V}^{(n)^{-1}}\boldsymbol{0} + \epsilon_{n+1}\boldsymbol{0} = \boldsymbol{0}.\end{aligned} \tag{44}$$

This last expression proves (42). □

ACKNOWLEDGMENT

Many superb engineers at Watkins–Johnson contributed to the hardware experiments described in this paper. The technical support from L. Jugler, W. Tariq, M. Oh, A. Faheem, H. Fan, J. Tsay, L. Ruck, J. Taylor, G. Rowe, P. Floyd, B. Hiller, D. Woodruff, D. Bell, S. Wilbur, C. Welch, K. Burr, D. Littke, S. Shaffer, B. Goettsch, T. Dziwulski, T. Potter, P. Kline, J. Vucetic, and J. Rugolo is gratefully acknowledged.

REFERENCES

[1] K. M. Chugg and A. Polydoros, “MLSE for an unknown channel—Part I: Optimality Considerations,” *IEEE Trans. Commun.*, vol. 44, pp. 836–846, July 1996.

[2] M.-C. Chiu and C.-C. Chao, “Analysis of LMS-adaptive MLSE equalization on multipath fading channels,” *IEEE Trans. Commun.*, vol. 44, pp. 1684–1692, Dec. 1996.

[3] G. Castellini, F. Conti, E. Del Re, and L. Pierucci, “A continuously adaptive MLSE reciver for mobile communications: Algorithm and performance,” *IEEE Trans. Commun.*, vol. 45, pp. 80–89, Jan. 1997.

[4] B. C. Ng, M. Cedervall, and A. Paulraj, “A structured channel estimator for maximum-likelihood sequence detection,” *IEEE Commun. Lett.*, vol. 1, pp. 52–55, Mar. 1997.

[5] R. D. Nowak, “Penalized least squares estimation of Volterra filters and higher-order statistics,” *IEEE Trans. Signal Processing*, vol. 46, pp. 419–428, Feb. 1998.

[6] ——, “Optimal signal estimation using cross-validation,” Rice Univ., Houston, TX, Rice Univ. Tech. Rep., ECE-TR-9601, Jan. 1998.

[7] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1986.

[8] R. Raheli, A. Polydoros, and C.-K. Tzou, “Per-survivor processing: A general approach to MLSE in uncertain environments,” *IEEE Trans. Commun.*, vol. 43, pp. 354–364, Feb./Mar./Apr. 1995.

[9] M. Martone, “An adaptive algorithm for adaptive antenna array low-rank processing in cellular TDMA base-stations,” *IEEE Trans. Commun.*, vol. 46, pp. 627–643, May 1998.

[10] ——, “Complex scaled tangent rotations (CSTAR) for fast space-time adaptive equalization of wireless TDMA,” *IEEE Trans. Commun.*, vol. 46, pp. 1587–1591, Dec. 1998.

[11] ——, "On MMSE real-time antenna array processing using fourth-order statistics in the US cellular TDMA system," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1396–1410, Oct. 1998.
[12] ——, "Cumulant-based adaptive multi-channel filtering for wireless communication systems with multipath RF propagation using antenna arrays," *IEEE Trans. Veh. Technol.*, vol. 47, pp. 377–391, May 1998.
[13] ——, "Hybrid nonlinear moments subspace processing for wireless communication systems using antenna arrays," *IEEE Trans. Signal Processing*, vol. 47, pp. 1434–1441, May 1999.
[14] L. L. Scharf, *Statistical Signal Processing*. Reading, MA: Addison-Wesley, 1991.
[15] "TDMA cellular/PCS—Radio interface—Mobile station—Base station compatibility—Digital control channel,", TIA/EIA/IS-136.1-A, Oct. 1996.
[16] "TDMA cellular/PCS—Radio interface—Mobile station—Base station compatibility—Traffic channels and FSK control channel,", TIA/EIA/IS-136.2-A, Oct. 1996.
[17] G. Ungerboeck, "Adaptive maximum-likelihood receiver for carrier-modulated data transmission systems," *IEEE Trans. Commun.*, vol. COM-22, May 1974.
[18] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*. New York: Wiley, 1968, pt. I.
[19] G. D. Forney, "Maximum likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 363–378, May 1972.
[20] F. R. Magee, "Adaptive maximum likelihood sequence estimation for digital signaling in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 120–124, Jan. 1973.
[21] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.

Massimiliano (Max) Martone (M'93) was born in Rome, Italy. He received the "Doctor in Electronic Engineering" degree from the University of Rome "La Sapienza," Rome, Italy, in 1990.

From 1990 to 1991, he was with the Italian Air Force, and he consulted in the area of digital signal processing applied to communications for Staer, Inc., S.P.E., Inc., and TRS-Alfa Consult, Inc. In 1991, he joined the technical staff of the On-Board Equipment Division, Alenia Spazio, where he was involved in the design of satellite receivers and spread-spectrum transponders for the European Space Agency. From 1992 to 1994, he collaborated with the digital communications research group of Fondazione Ugo Bordoni. In 1994, he was appointed Visiting Scientist at the ECSE Department of Rensselaer Polytechnic Institute, Troy, NY. He was a wireless communications consultant for ATS, Inc., Waltham, MA, and in 1995, he joined the Telecommunications Group with Watkins–Johnson Company, Gaithersburg, MD, where he is currently head of the Advanced Wireless Development section. He has been a leader in the development of several proprietary signal processing algorithms and digital hardware platforms used in Watkins–Johnson state-of-the-art wireless products. He has published more than 50 papers in international journals and proceedings of international conferences and contributed to one book on cellular communications. He has been listed in Marquis Who's Who in the World, Who's Who in Finance in Industry, and Who's Who in the East. His main interests are in advanced signal processing for wireless receivers implementation, spread-spectrum multiple-access communications, cellular radio architectures, and VLSI implementations.

Dr. Martone is a member of the New York Academy of Sciences and the American Association for the Advancement of Science. He has received 14 Watkins–Johnson Editorial Achievement Awards.