Wavelet-Based Separating Kernels for Sequence Estimation with Unknown Rapidly Time-Varying Channels

Massimiliano (Max) Martone, *Member, IEEE*

***Abstract*— A new method for blind maximum-likelihood sequence estimation is proposed. The unknown channel time variations are decomposed using optimal unconditional bases such as orthonormal wavelet bases. It is shown in this letter that, it is possible to represent the channel in a reduced-order dimensional space by matching the scattering function of the multipath channel to its decomposition and obtain an approach to per-survivor processing that is effective in fast fading environments such as those practically found in macrocell wireless communication applications.**

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

I. INTRODUCTION

THE FUNDAMENTAL model assumption in many of the known approaches to joint channel-sequence estimation is that the channel is static over a certain number of symbols so that a generalized likelihood (GL) argument can be applied [1]–[4]. Slow variations of the channel are then compensated by using adaptive algorithms that ultimately force the estimate to be constantly in search of a convergence point. If the channel coefficients variations in time are fast with respect to the convergence time of the adaptive algorithm, significant degradation may result. In this work we depart from this typical approach in search of new kernels that more accurately can characterize the time varying nature of the estimation problem and focus on a multiresolution representation of the fading process in each component of the channel response, elaborating the ideas of [5] (and in a sense of [6]) as they apply to the blind sequence estimation problem.

II. SYSTEM MODEL

Assuming a baseband continuous-time representation we model the received signal as $\tilde{r}(t) = \sum_n a_n \tilde{h}(t, t-nT) + \tilde{n}(t)$, a_n are the complex symbols defining the signal constellation sed for the particular digital modulation scheme, $\tilde{h}(t, \tau)$ is the convolution of $f(t, \tau)$, the impulse response of a frequency selective fading multipath channel and $g(\tau)$, the overall shaping filter, $\tilde{n}(t)$ is additive white Gaussian noise. The channel is assumed wide sense stationary uncorrelated scattering (WSSUS)[1] [7]. The fundamental idea is to apply a multiresolution decomposition such as the wavelet transform (in reality we use the discrete-time discrete wavelet transform)[2] to $\tilde{h}(t, \tau)$ at any τ and to disregard in the representation components of the decomposition that are unimportant for the reproduction of the time-varying process. The received signal $\tilde{r}(t)$ is the input to an ideal low-pass filter 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 maximum-likelihood sequence estimation $y_n = \sum_{l=0}^{NR-1} h(n, n-l)\tilde{a}_l + \eta_n$ where $\tilde{a}_n = \sum_l \delta(n - lR)a_l$ is the sequence of N symbols interleaved with $R-1$ zeros, $\delta(n)$ is the discrete time delta function, $h(n, k) = \tilde{h}[(nT/R), (kT/R)]$, 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 $g(t)$ 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 $(D+1)T/R = (J+K+1)T$ so that

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

with $\mathbf{a}_n = [\tilde{a}_n, \cdots, \tilde{a}_{n-D}]^T$ and $\mathbf{h}_n = [h(n, 0), \cdots, h(n, D)]^T$. According to the maximum-likelihood principle the optimal metric corresponding to the hypothesis that the sequence $[\alpha_0, \cdots, \alpha_{RN-1}]^T$ was transmitted is

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

where $\boldsymbol{\alpha}_n = [\alpha_n, \cdots, \alpha_{n-D}]^T$, and $l(\boldsymbol{\alpha}_n, \mathbf{h}_n)$ is the branch metric. In practice the time varying coefficients $\mathbf{h}_n$ are unknown and must be identified.

[1]The WSSUS assumption implies $E\{f^*(t_1, \tau_1)f(t_2, \tau_2)\} = P_f(\tau_1, t_2 - t_1)\delta(\tau_2 - \tau_1)$ and the expectation is over the channel ensemble. The scattering function of the multipath channel is related to $P_f(\tau, t)$ by the Fourier transform, first with respect to τ, and then with respect to t.

[2]The wavelet transform is an atomic decomposition that represents a signal [in our case $h(t, \tau)$] in terms of shifted/dilated versions of a prototype bandpass wavelet function and shifted versions of a low-pass scaling function [8].

Manuscript received October 10, 1998. The assoicate editor coordinating the review of this letter and approving it for publication was Prof. N. C. Beaulieu. This work was supported by the Watkins-Johnson Company, Gaithersburg, MD.

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

Publisher Item Identifier S 1089-7798(99)02643-5.

III. The New Kernel

One approach to implement the discrete wavelet transform is to use a binary subband tree structure that is constructed using stages of two-channel filterbanks [9]. Define $c_i(n)\ i = 0, 1$ as a dyadic perfect reconstruction filter bank [9] [in the z-domain $C_i(z)$]. At a generic resolution depth P we can represent $h(n, k)$ as [5]

$$h(n, k) = \sum_{m=0}^{NR/2^P-1} \zeta_{P,m}(k) c_0^{(P)}(2^P m - n) + \sum_{l=1}^{P} \sum_{m=0}^{NR/2^l-1} \xi_{l,m}(k) c_1^{(l)}(2^l m - n) \quad (3)$$

where $\zeta_{P,m}(k)$ and $\xi_{l,m}(k)$ are discrete wavelet transform (DWT) coefficients and the filters $c_0^{(P)}(n)$ and $c_1^{(l)}(n)$ (with real coefficients) are given in the z-domain by $C_0^{(P)}(z) = \prod_{i=1}^{P-1} C_0(z^{2^i})$ and $C_1^{(l)}(z) = C_1(z^{2^l-1}) \prod_{i=1}^{l-2} C_0(z^{2^i})$. Using vector notation it is possible to express (3) simply as

$$h(n, k) = \mathbf{c}(P, n)^T \mathbf{w}(P, k) \quad (4)$$

where the organization of the wavelet coefficients in $\mathbf{w}(P, k)$ is $\mathbf{w}(P, k) = [\boldsymbol{\zeta}_P(k)^T \boldsymbol{\xi}_1(k)^T \cdots \boldsymbol{\xi}_P(k)^T]^T$, with[3] $[\boldsymbol{\zeta}_P(k)]_m = \zeta_{P,m}(k)$, $[\boldsymbol{\xi}_l(k)]_m = \xi_{l,m}(k)$. Evidently $\mathbf{c}(P, n) = [\mathbf{c}_0^{(P)}(n)^T, \mathbf{c}_1^{(1)}(n)^T, \cdots, \mathbf{c}_1^{(P)}(n)^T]^T$ where $[\mathbf{c}_i^{(l)}(n)]_m = c_i^{(l)}(2^l m - n)$ for $i = 0, 1$. We assume that the discrete time autocorrelation function of the channel $E\{h^*(n, k_1)h(n + l, k_2)\} = R_h(l, k_1, k_2)$ is known.[4] Using (3) we have

$$R_h(l, k_1, k_2) = \mathbf{c}(P, n)^T E\{\mathbf{w}^*(P, k_1)\mathbf{w}^T(P, k_2)\} \cdot \mathbf{c}(P, n + l). \quad (5)$$

Observe that due to the orthonormality of the DWT, defining $\mathbf{h}(k) = [h(0, k), h(1, k), \cdots, h(NR-1, k)]^T$, we also have $E\{\mathbf{w}^*(P, k_1)\mathbf{w}^T(P, k_2)\} = E\{\mathbf{h}^*(k_1)\mathbf{h}^T(k_2)\}$ which is a matrix constructed from $R_h(l, k_1, k_2)$. We can validate our model (3) using (5). In other words we can determine which components of the expansion (3) can be neglected (i.e., zeroed) without compromising the parsimony of the wavelet-based representation. Define $(P, k)_M$ as the vector obtained zeroing the last $\sum_{m=1}^{M} NR/2^m$ elements of $\mathbf{w}(P, k)$ and $\mathbf{c}(P, n)_M$ as the vector obtained zeroing the last $\sum_{m=1}^{M} NR/2^m$ elements of $\mathbf{c}(P, n)$. A possible indication of the parsimony in the representation (3) retaining only $NR/2^M$ DWT coefficients, $M = 0, 1, 2, \cdots, P$ is $\gamma(P, M) = \sum_{k_1=0}^{D} \sum_{k_2=0}^{D} ||\mathbf{r}(k_1, k_2) - \mathbf{v}_M(k_1, k_2)||^2$ where

$$[\mathbf{v}_M(k_1, k_2)]_l = \frac{1}{NR} \sum_{n=0}^{NR-1} \mathbf{c}(P, n)_M^T E\{\mathbf{h}^*(k_1)\mathbf{h}^T(k_2)\} \cdot \mathbf{c}(P, n + l)_M$$

and $[\mathbf{r}(k_1, k_2)]_l = R_h(l, k_1, k_2)$. Once a particular M has been selected, we can write

$$\mathbf{h}_n \simeq (\mathbf{I}_D \otimes \overline{\mathbf{c}}(n)^T)[\overline{\mathbf{w}}(0)^T, \cdots, \overline{\mathbf{w}}(D)^T]^T = \mathbf{C}(n)\mathbf{w} \quad (6)$$

where[5] $\overline{\mathbf{c}}(n) = \mathbf{I}_{[q, \overline{q}]}\mathbf{c}(P, n)$, $\overline{\mathbf{w}}(k) = \mathbf{I}_{[q, \overline{q}]}\ \mathbf{w}(P, k)$ with $q = NR2^{-M}$, $\overline{q} = NR - q = NR \sum_{m=1}^{M} 2^{-m}$, and $\otimes$ is the Kronecker product. The metric in (2) can be expressed as $L_N \simeq \sum_{n=0}^{NR-1} |y_n - \tilde{\boldsymbol{\alpha}}_n^T \mathbf{w}|^2$ where $\tilde{\boldsymbol{\alpha}}_n^T = \boldsymbol{\alpha}_n^T \mathbf{C}(n)$. We can then write

$$L_N = ||\mathbf{y} - \tilde{\mathbf{A}}\hat{\mathbf{w}}||^2 = \mathbf{y}^H \tilde{\mathbf{Q}}_A \mathbf{y} \quad (7)$$

where $\tilde{\mathbf{Q}}_A = \mathbf{I} - \tilde{\mathbf{P}}_A$ and $\tilde{\mathbf{P}}_A = \tilde{\mathbf{A}}(\tilde{\mathbf{A}}^H\tilde{\mathbf{A}})^{-1}\tilde{\mathbf{A}}^H$ is the new separating kernel with $\tilde{\mathbf{A}} = [\tilde{\boldsymbol{\alpha}}_0, \tilde{\boldsymbol{\alpha}}_1, \cdots \tilde{\boldsymbol{\alpha}}_{NR-1}]$. Evidently $\tilde{\mathbf{P}}_A$ is the projection matrix for the signal space $\tilde{\mathcal{A}}$ spanned by the columns of $\tilde{\mathbf{A}}$.

IV. Adaptive Implementation

Real-time algorithms to compute the metric (7) on a sample by sample basis can be easily derived defining $\hat{\mathbf{w}}_n$ as the estimated DWT vector at time step n and $l(\boldsymbol{\alpha}_n, \hat{\mathbf{w}}_n) = |y_n - \tilde{\boldsymbol{\alpha}}_n^T \hat{\mathbf{w}}_n|^2$ as the corresponding branch metric.[6] Possible approaches are gradient-based methods [like the least mean squares (LMS)] and recursive least squares (RLS)-based methods. The typical recursion in both cases is

$$\hat{\mathbf{w}}_{n+1} = \hat{\mathbf{w}}_n + \mathbf{p}_n(\tilde{\boldsymbol{\alpha}}_n, \hat{\mathbf{w}}_n) \quad (8)$$

where $\mathbf{p}_n(\tilde{\boldsymbol{\alpha}}_n, \hat{\mathbf{w}}_n) = \mu\epsilon_n(\tilde{\boldsymbol{\alpha}}_n, \hat{\mathbf{w}}_n)\tilde{\boldsymbol{\alpha}}_n^*$ if the algorithm of choice is the LMS, μ is a step-size parameter that controls the rate of adjustment, and $\epsilon_n(\tilde{\boldsymbol{\alpha}}_n, \hat{\mathbf{w}}_n) = y_n - \tilde{\boldsymbol{\alpha}}_n^T \hat{\mathbf{w}}_n$.

Remark: It is important to observe that the true DWT vector $\mathbf{w}$ is time-invariant, so the task of the adaptive algorithm in the proposed approach is to converge to the channel parameters as opposed to *track* them.

V. Performance Analysis Results

The cellular TDMA system under analysis is IS-136. Hardware experiments are performed using the Watkins–Johnson wideband dual-mode (AMPS and IS-136) base-station system Base_2 [12], [13]. The slots are generated of dyadic length with $R = 2$, $(NR = 256)$ the symbol period is 41.2 ms and $C_0(z)C_1(z)$ are Daubechies filters of order 3 as in [5]. We assume [10] that $R_h(l, k_1, k_2) = \sigma_k^2 J_0(2\pi l f_D T/R)$ for $k_1 = k_2 = k$ where $J_0(x)$ is the Bessel function of order zero, f_D is Doppler spread which depends on the velocity of the mobile transmitter, σ_k^2 is the variance of the fading process. It is then straightforward to verify (using the described validation method) that for speeds of the mobile up to 300 Km/hr and $P = 3$ or $P = 4$ it is possible to impose $\xi_{l,m}(k) = 0$ in (3) for any l, m, k without significant penalty in the matching metric $\gamma(P, M)$, for $M = P$. This means that we can have only $NR/2^M = NR/2^P = 32$ or $NR/2^M = NR/2^P =$

[3] The notation $[\mathbf{v}]_k$ is used for the kth element of vector $\mathbf{v}$.

[4] In fact $R_h(l, k_1, k_2)$ can be computed from the knowledge of $P_f(\tau, lT/R)$ and $g(\tau)$.

[5] The notation $\mathbf{I}_A$ identifies the $A \times A$ identity matrix while the notation $\mathbf{I}_{[A, B]}$ identifies the $A \times (A + B)$ matrix whose first A columns are the columns of $\mathbf{I}_A$, and the last B columns have elements equal to zero.

[6] Observe that even if $\tilde{\boldsymbol{\alpha}}_n$ is a $q = NR2^{-M}$ vector, it still depends on $D + 1$ symbols only. So the same trellis search typically implemented using (2) is applicable to (7).

TABLE I
COMPUTATIONAL COMPLEXITY WITH TWO ANTENNAS, FOUR STATES VITERBI ALGORITHM $R = 2$, $D = 3$, $NR = 256$

Algorithm	MIPS (millions of instructions per second)
PSPLMS	37
DWTPSPLMS(3)	54
DWTPSPLMS(4)	49
DWTPSPLMS(5)	45
DWTPSPLMS(6)	39

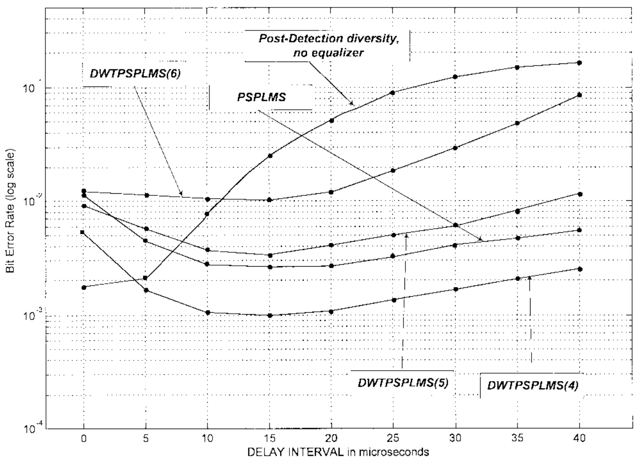

Fig. 1. Bit error rate at SNR per bit equal to 25 dB and 100 km/h, versus delay interval (of the two-ray fading channel). Training is used for the first 14 symbols (IS-136 standard) and slots have a (reverse digital traffic channel) RDTC-like format with 128 symbols.

16 wavelet coefficients in $\mathbf{w}$. In other words the excellent "compacting" properties (see [11]) of the wavelet transform are able to compress most of the energy of the time variations of the channel in the low resolution representation of the fading process and this makes the approach very attractive in practice. There are two antennas and the metric has to be modified for the two independent fading branches as $L_N = \sum_{n=0}^{NR-1} \sum_{i=1}^{2} |y_n^{(i)} - \tilde{\boldsymbol{\alpha}}_n^T \hat{\mathbf{w}}_n^{(i)}|^2$, where $y_n^{(i)}$, $\hat{\mathbf{w}}_n^{(i)}$ are signal samples and DWT channel vector from the ith antenna $i = 1, 2$, respectively. More details on the hardware set-up of the experiments and the hardware architecture of the receiver can be found in [12]. The trellis has four states ($D = 3$) because the span of the ISI channel is two symbols and the Viterbi algorithm is used in the trellis search. We compare the following approaches:

`PSPLMS` One channel estimator per survivor using the PSP technique [3].

`DWTPSPLMS(P)` One DWT-channel estimator per survivor using the PSP technique, resolution depth equal to P, and representing the channel in the low-resolution $NR/2^P$-dimensional space.

The complexity of each algorithm is reported in Table I 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 (RDTC). The above

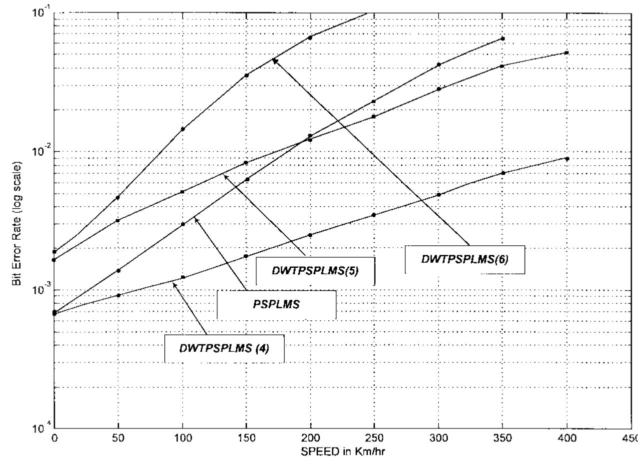

Fig. 2. Bit error rate at SNR per bit equal to 22 dB and 20.6 ms delay interval and versus mobile speed. Training is used for the first 14 symbols (IS-136 standard) and slots have a (reverse digital traffic channel) RDTC-like format with 128 symbols. Observe the different slope of the BER performance of the traditional approach (PSPLMS) with respect to the new wavelet-based approach.

algorithms have been designed for analog devices ADSP-2181 processors (40-MHz clock rate, 16 bits fixed point arithmetic) and have been implemented in simulated fixed-point language, while executed on baseband samples collected from the hardware platform. Bit error rate analysis results are shown in Figs. 1 and 2. Fig. 1 shows results at 100 km/h versus delay interval, while the SNR per bit is 25 dB. Fig. 2 is for delay interval equal to 20.6 μs varying the mobile speed and SNR per bit equal to 22 dB.

REFERENCES

[1] M. Ghosh and C. L. Weber, "Maximum likelihood blind equalization," *Opt. Eng.*, vol. 31, no. 6, pp. 1224–1228, June 1992.

[2] N. Seshadri, "Joint data and channel estimation using trellis search techniques," *IEEE Trans. Commun.*, vol. 42, pp. 80–89, Feb./Mar./Apr. 1994.

[3] 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.

[4] 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.

[5] M. K. Tsatsanis and G. B. Giannakis, "Time-varying system identification and model validation using wavelets," *IEEE Trans. Signal Processing*, vol. 41, pp. 3512–3523, Dec. 1993.

[6] G. B. Giannakis and C. Tepedelenlioglu, "Basis-expansion models and diversity techniques for blind identification and equalization of time-varying channels," *Proc. IEEE*, vol. 86, pp. 1969–1986, Oct. 1998.

[7] P. A. Bello, "Characterization of randomly time variant linear channels," *IEEE Trans. Commun. Syst. Technol.*, vol. COM-11, pp. 360–393, Dec. 1963.

[8] I. Daubechies, *Ten Lectures in Wavelets.* New York: SIAM, 1992.

[9] P. P. Vaidyanathan, *Multirate Systems and Filter Banks.* Englewood Cliffs, NJ: Prentice Hall, 1993.

[10] W. C. Jakes, Jr., *Microwave Mobile Communications.* New York: Wiley, 1974.

[11] D. L. Donoho, "Unconditional bases are optimal bases for data compression and for statistical estimation," *App. Comp. Harmonic Anal.*, **vol.??**, pp. 100–115, Dec. 1993.

[12] 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.

[13] "Wideband base station architecture for digital communications system," U.S. Patent 5 678 268, June 16, 1998.