Near-Optimal Noncoherent Sequence Detection for Doubly Dispersive Channels

Sung-Jun Hwang and Philip Schniter
The Ohio State University, Dept. of Electrical and Computer Engineering,
2015 Neil Avenue, Columbus, OH 43210.
E-mail: {hwangsu, schniter}@ece.osu.edu

***Abstract*—We propose a scheme for near-optimal sequence detection (SD) of uncoded block transmissions over unknown doubly dispersive (DD) channels. Starting with a noncoherent maximum likelihood (ML) metric that leverages a basis expansion model (BEM) for the channel's time-variation, we propose an efficient noncoherent SD strategy based on suboptimal tree search with a fast metric update. Our scheme yields performance within a fraction-of-a-dB from ML sequence detection with genie-aided channel estimates, and maintains complexity that is only quadratic in the block length.**[1]

I. Introduction

In this paper, we consider uncoded block transmission through a doubly dispersive (DD) channel, i.e., a quickly time-varying inter-symbol interference (ISI) channel. When the receiver has perfect channel state information (CSI), maximum-likelihood (ML) sequence detection (SD)—known to minimize the probability of sequence-error, can be implemented using the classical Viterbi algorithm (VA) [1] with complexity $\mathcal{O}(N|\mathcal{S}|^{N_h+1})$, where N denotes the block length, $|\mathcal{S}|$ the alphabet size, and N_h the discrete channel length.

Here we focus on the case that the receiver knows the channel statistics but not the channel state, yielding a noncoherent sequence detection problem. In addition, we focus on Gaussian channels. In this case, MLSD requires a brute-force search among all sequences [2] and implicitly computes a minimum mean-squared error (MMSE) channel estimate for each. It is important to note that the VA cannot be used for MLSD since the ML branch metrics in trellis-based SD are a function of all past and future trellis states. Even when the channel is Gauss-Markov, the ML branch metrics are a function of all past states. Per-survivor processing (PSP) [3], however, can be employed to yield an accurate VA-based approximation of MLSD. In PSP-VA, channel estimates are computed for each surviving path, using past symbol hypotheses in a decision-directed manner. With a Gauss-Markov channel, for instance, the Kalman algorithm can be employed for recursive MMSE channel estimation [2]. As shown in [4], the gap between PSP-VA and MLSD can be made quite small by adopting a list-Viterbi algorithm (LVA), which retains $L \geq 1$ partial paths at each trellis state. Assuming a first-order Gauss-Markov channel model, the complexity of Kalman-PSP-LVA is $\mathcal{O}(N|\mathcal{S}|^{N_h+1}N_h^3 L)$, which is attractive for long blocks due to a linear dependence on block length N. PSP-VA has also been proposed in conjunction with LMS [5] and RLS [3], which may be convenient when the channel statistics are unknown.

Viewing PSP-VA (or PSP-LVA) as a form of suboptimal tree search, the question remains as to whether a different form of suboptimal tree search could offer a superior performance/complexity tradeoff, say, for longer channels. Towards this aim, we propose near-ML noncoherent SD that leverages a simplified ML metric, suboptimal breadth-first search via the M-algorithm [6], and recursive MMSE estimation of basis expansion model (BEM) [7] coefficients. In Section IV, we show that our algorithm performs within a fraction-of-a-dB from MLSD with genie-aided MMSE channel estimates; to our knowledge, this performance is state-of-the-art. In addition, we demonstrate that our algorithm is robust to over-estimation of the channel fading rate. Since our algorithm is $\mathcal{O}(N^2|\mathcal{S}|N_h N_b M)$, where N_b denotes the basis size, it is cheaper than Kalman-PSP-LVA for moderate channel lengths (e.g., $N_h \geq 4$) and block lengths (e.g., $N < 100$). A detailed complexity comparison is provided in Section III.

We now discuss related work. To our knowledge, the Kalman-PSP-LVA [4] is the highest-performance *practical* means of noncoherent sequence detection in the doubly dispersive environment; it has been shown [4], [8] to outperform both LMS-PSP-LVA and RLS-PSP-LVA, as well as iterative methods based on expectation-maximization (EM). Hence, we use it as a baseline for comparison. The estimation of BEM coefficients, in place of the time-varying impulse response, was proposed in [9] in the context of PSP-VA. But BEM estimation does not appear to work well in conjunction with the path-pruning of PSP-VA; our experiments, as well as those in [9], show an early error floor. Though the suggestion to use generic tree-search algorithms for joint channel/data estimation can be found, e.g., in [3], we are not aware of existing strategies that are particularly well-suited to the doubly dispersive environment.

II. System Model

We consider a discrete-time complex-baseband system with transmitted symbols $\{s_n\}$ drawn from a finite QAM alphabet $\mathcal{S}$, and with a channel described by the time-varying discrete

[1]This work was supported by the National Science Foundation CAREER grant CCR-0237037 and the Office of Naval Research.

impulse response $h_{n,l}$. The received samples are written as

$$r_n = \sum_{l=0}^{N_h-1} h_{n,l} s_{n-l} + v_n, \tag{1}$$

where N_h is the channel length and $\{v_n\}$ is zero-mean circular white Gaussian noise (CWGN) with covariance σ^2.

The receiver approximates the channel response, over the block of N samples, by the N_b-size basis expansion

$$h_{n,l} \approx \sum_{p=0}^{N_b-1} b_{n,p}\theta_{p,l} \quad \text{for } n \in \{0,\dots,N-1\}. \tag{2}$$

Here, N_b and $\{b_{n,p}\}$ are design parameters and $\{\theta_{p,l}\}$ are random channel coefficients. An error free approximation is possible with $N_b = N$, while significant reduction in detection complexity is possible with $N_b \ll N$. As we shall see, the detection performance loss due to channel modeling error can be made relatively small through proper choice of basis $\{b_{n,p}\}$, even for $N_b \ll N$. Under approximation (2), the system model (1) can be written, for $n \in \{0,\dots,N-1\}$, as

$$r_n = \boldsymbol{b}_n^H \sum_{l=0}^{N_h-1} s_{n-l}\boldsymbol{\theta}_l + v_n, \tag{3}$$

for $\boldsymbol{b}_n := [b_{n,0},\dots,b_{n,N_b-1}]^H$ and $\boldsymbol{\theta}_l := [\theta_{0,l},\dots,\theta_{N_b-1,l}]^T$. For $n < N$, we can write $\boldsymbol{r}_n := [r_n,\dots,r_0]^T$ as

$$\boldsymbol{r}_n = \boldsymbol{B}_n \boldsymbol{S}_0^n \boldsymbol{\theta} + \boldsymbol{v}_n \tag{4}$$

$$r_{n+1} = \boldsymbol{b}_{n+1}^H \boldsymbol{S}_{n+1}^{n+1}\boldsymbol{\theta} + v_{n+1}, \tag{5}$$

where $\boldsymbol{\theta} := [\boldsymbol{\theta}_0^T,\dots,\boldsymbol{\theta}_{N_h-1}^T]^T$, $\boldsymbol{v}_n := [v_n,\dots,v_0]^T$, and

$$\boldsymbol{B}_n := \begin{bmatrix} \boldsymbol{b}_n^H & & \\ & \ddots & \\ & & \boldsymbol{b}_0^H \end{bmatrix} \tag{6}$$

$$\boldsymbol{S}_m^n := \begin{bmatrix} s_n\boldsymbol{I}_{N_b} & \cdots & s_{n-N_h+1}\boldsymbol{I}_{N_b} \\ \vdots & & \vdots \\ s_m\boldsymbol{I}_{N_b} & \cdots & s_{m-N_h+1}\boldsymbol{I}_{N_b} \end{bmatrix}. \tag{7}$$

The following abbreviations will prove useful in the sequel:

$$\boldsymbol{A}_{\boldsymbol{s}_n} := \boldsymbol{B}_n \boldsymbol{S}_0^n \tag{8}$$

$$\boldsymbol{a}_{\boldsymbol{s}_{n+1}}^H := \boldsymbol{b}_{n+1}^H \boldsymbol{S}_{n+1}^{n+1}. \tag{9}$$

III. Fast Noncoherent Sequence Detection

In this section, we describe a fast algorithm to decode $\{s_n\}_{n=0}^{N-1}$ from the observations $\{r_n\}_{n=0}^{N-1}$ in the presence of channel uncertainty. In doing so, we assume that $\{s_n\}_{n<0}$ are zero or otherwise known. Our algorithm is sequential in nature, in that it estimates the partial-sequence $\boldsymbol{s}_n := [s_n,\dots,s_0]^T$ for $n = 0,1,2,\dots$, ultimately estimating the full sequence $\boldsymbol{s}_{N-1}$. In deriving our algorithm, we employ the BEM approximation (2) and treat $\boldsymbol{\theta}$ as zero-mean circular complex Gaussian with known autocovariance $\boldsymbol{R}_\theta$. Furthermore, we assume that the BEM is chosen to make $\boldsymbol{R}_\theta$ full rank and diagonal. In the performance evaluation of Section IV, however, we will not assume that the BEM approximation (2) holds perfectly, and hence will see the effect of BEM mismatch.

A. Noncoherent Sequence Detection Metric

The MLSD estimate is defined as

$$\hat{\boldsymbol{s}}_n = \arg\max_{\boldsymbol{s}_n} p(\boldsymbol{r}_n|\boldsymbol{s}_n), \tag{10}$$

where, marginalizing over the channel,

$$\begin{aligned} &p(\boldsymbol{r}_n|\boldsymbol{s}_n) \\ &= \int_{\boldsymbol{\theta}} p(\boldsymbol{r}_n|\boldsymbol{s}_n,\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta} \\ &= \int_{\boldsymbol{\theta}} \frac{1}{(\pi\sigma^2)^n}\exp\left\{-\tfrac{1}{\sigma^2}\|\boldsymbol{r}_n - \boldsymbol{A}_{\boldsymbol{s}_n}\boldsymbol{\theta}\|^2\right\} p(\boldsymbol{\theta})d\boldsymbol{\theta}. \end{aligned}$$

Since $\boldsymbol{\theta} \sim \mathcal{CN}(\boldsymbol{0},\boldsymbol{R}_\theta)$, we can write

$$\begin{aligned} &p(\boldsymbol{r}_n|\boldsymbol{s}_n) \\ &= C_1\int_{\boldsymbol{\theta}} \exp\left\{-\tfrac{1}{\sigma^2}\|\boldsymbol{r}_n - \boldsymbol{A}_{\boldsymbol{s}_n}\boldsymbol{\theta}\|^2 - \boldsymbol{\theta}^H\boldsymbol{R}_\theta^{-1}\boldsymbol{\theta}\right\}d\boldsymbol{\theta} \\ &= C_2\int_{\boldsymbol{\theta}} \exp\left\{-\tfrac{1}{\sigma^2}\left\|\boldsymbol{\theta} - \boldsymbol{\Sigma}_{\boldsymbol{s}_n}^{-1}\boldsymbol{A}_{\boldsymbol{s}_n}^H\boldsymbol{r}_n\right\|_{\boldsymbol{\Sigma}_{\boldsymbol{s}_n}}^2\right\}d\boldsymbol{\theta} \\ &\quad\times\exp\left\{-\tfrac{1}{\sigma^2}\left(\boldsymbol{r}_n^H\boldsymbol{r}_n - \boldsymbol{r}_n^H\boldsymbol{A}_{\boldsymbol{s}_n}\boldsymbol{\Sigma}_{\boldsymbol{s}_n}^{-1}\boldsymbol{A}_{\boldsymbol{s}_n}^H\boldsymbol{r}_n\right)\right\} \\ &= \frac{C_3}{\det(\sigma^{-2}\boldsymbol{\Sigma}_{\boldsymbol{s}_n})}\exp\left\{-\tfrac{1}{\sigma^2}\left(\boldsymbol{r}_n^H\boldsymbol{r}_n - \boldsymbol{r}_n^H\boldsymbol{A}_{\boldsymbol{s}_n}\boldsymbol{\Sigma}_{\boldsymbol{s}_n}^{-1}\boldsymbol{A}_{\boldsymbol{s}_n}^H\boldsymbol{r}_n\right)\right\} \end{aligned}$$

where

$$\boldsymbol{\Sigma}_{\boldsymbol{s}_n} := \boldsymbol{A}_{\boldsymbol{s}_n}^H\boldsymbol{A}_{\boldsymbol{s}_n} + \sigma^2\boldsymbol{R}_\theta^{-1}, \tag{11}$$

and where $\{C_i\}$ are constants irrelevant to the maximization in (10). Using the monotonicity of $\log(\cdot)$, we can write

$$\begin{aligned} \hat{\boldsymbol{s}}_n = \arg\max_{\boldsymbol{s}_n}\Big\{&\boldsymbol{r}_n^H\sigma^{-2}\big(\boldsymbol{A}_{\boldsymbol{s}_n}\boldsymbol{\Sigma}_{\boldsymbol{s}_n}^{-1}\boldsymbol{A}_{\boldsymbol{s}_n}^H - \boldsymbol{I}_{n+1}\big)\boldsymbol{r}_n \\ &-\log\det(\sigma^{-2}\boldsymbol{\Sigma}_{\boldsymbol{s}_n})\Big\}. \end{aligned} \tag{12}$$

As reported elsewhere (e.g., [10]), the maximization in (12) can be simplified by ignoring the bias term $\log\det(\sigma^{-2}\boldsymbol{\Sigma}_{\boldsymbol{s}_n})$. Remarkably, experiments in Section IV show that the resulting performance loss is negligible over the SNR range of interest. The simplified detection rule reads as

$$\hat{\boldsymbol{s}}_n \approx \arg\min_{\boldsymbol{s}_n}\mu(\boldsymbol{s}_n) \tag{13}$$

with

$$\mu(\boldsymbol{s}_n) := \sigma^2\boldsymbol{r}_n^H\boldsymbol{\Phi}_{\boldsymbol{s}_n}\boldsymbol{r}_n \tag{14}$$

$$\boldsymbol{\Phi}_{\boldsymbol{s}_n} := \sigma^{-2}\big(\boldsymbol{I}_{n+1} - \boldsymbol{A}_{\boldsymbol{s}_n}\boldsymbol{\Sigma}_{\boldsymbol{s}_n}^{-1}\boldsymbol{A}_{\boldsymbol{s}_n}^H\big) \tag{15}$$

$$= \big(\boldsymbol{A}_{\boldsymbol{s}_n}\boldsymbol{R}_\theta\boldsymbol{A}_{\boldsymbol{s}_n}^H + \sigma^2\boldsymbol{I}_{n+1}\big)^{-1}, \tag{16}$$

where (16) follows from the matrix inversion lemma.

We will now show that the detection rule (13) performs implicit minimum mean-squared error (MMSE) estimation of $\boldsymbol{\theta}$. Denoting by $\hat{\boldsymbol{\theta}}_{\boldsymbol{s}_n}$ the MMSE estimate of $\boldsymbol{\theta}$ from $\boldsymbol{r}_n$ given knowledge of $\boldsymbol{s}_n$, we have

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{s}_n} = \mathrm{E}\{\boldsymbol{\theta}\boldsymbol{r}_n^H|\boldsymbol{s}_n\}\,\mathrm{E}\{\boldsymbol{r}_n\boldsymbol{r}_n^H|\boldsymbol{s}_n\}^{-1}\boldsymbol{r}_n \tag{17}$$

$$= \boldsymbol{R}_\theta\boldsymbol{A}_{\boldsymbol{s}_n}^H\boldsymbol{\Phi}_{\boldsymbol{s}_n}\boldsymbol{r}_n, \tag{18}$$

where (18) follows from the fact that $\boldsymbol{\Phi}_{\boldsymbol{s}_n} = \mathrm{E}\{\boldsymbol{r}_n \boldsymbol{r}_n^H | \boldsymbol{s}_n\}^{-1}$. Plugging (15) into (18), and then applying (11), we find

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{\boldsymbol{s}_n} &= \sigma^{-2} \boldsymbol{R}_\theta \left(\boldsymbol{\Sigma}_{\boldsymbol{s}_n} \boldsymbol{\Sigma}_{\boldsymbol{s}_n}^{-1} \boldsymbol{A}_{\boldsymbol{s}_n}^H - \boldsymbol{A}_{\boldsymbol{s}_n}^H \boldsymbol{A}_{\boldsymbol{s}_n} \boldsymbol{\Sigma}_{\boldsymbol{s}_n}^{-1} \boldsymbol{A}_{\boldsymbol{s}_n}^H \right) \boldsymbol{r}_n \\ &= \boldsymbol{\Sigma}_{\boldsymbol{s}_n}^{-1} \boldsymbol{A}_{\boldsymbol{s}_n}^H \boldsymbol{r}_n, \end{aligned} \tag{19}$$

Then, from (14), (15), (19), and (11), we find

$$\begin{aligned} \mu(\boldsymbol{s}_n) &= \boldsymbol{r}_n^H \boldsymbol{r}_n - \boldsymbol{r}_n^H \boldsymbol{A}_{\boldsymbol{s}_n} \boldsymbol{\Sigma}_{\boldsymbol{s}_n}^{-1} \boldsymbol{A}_{\boldsymbol{s}_n}^H \boldsymbol{r}_n \\ &= \boldsymbol{r}_n^H \boldsymbol{r}_n - \hat{\boldsymbol{\theta}}_{\boldsymbol{s}_n}^H \boldsymbol{A}_{\boldsymbol{s}_n}^H \boldsymbol{r}_n - \boldsymbol{r}_n^H \boldsymbol{A}_{\boldsymbol{s}_n} \hat{\boldsymbol{\theta}}_{\boldsymbol{s}_n} + \hat{\boldsymbol{\theta}}_{\boldsymbol{s}_n}^H \boldsymbol{\Sigma}_{\boldsymbol{s}_n} \hat{\boldsymbol{\theta}}_{\boldsymbol{s}_n} \\ &= \|\boldsymbol{r}_n - \boldsymbol{A}_{\boldsymbol{s}_n} \hat{\boldsymbol{\theta}}_{\boldsymbol{s}_n}\|^2 - \hat{\boldsymbol{\theta}}_{\boldsymbol{s}_n}^H \boldsymbol{A}_{\boldsymbol{s}_n}^H \boldsymbol{A}_{\boldsymbol{s}_n} \hat{\boldsymbol{\theta}}_{\boldsymbol{s}_n} + \hat{\boldsymbol{\theta}}_{\boldsymbol{s}_n}^H \boldsymbol{\Sigma}_{\boldsymbol{s}_n} \hat{\boldsymbol{\theta}}_{\boldsymbol{s}_n} \\ &= \|\boldsymbol{r}_n - \boldsymbol{A}_{\boldsymbol{s}_n} \hat{\boldsymbol{\theta}}_{\boldsymbol{s}_n}\|^2 + \sigma^2 \hat{\boldsymbol{\theta}}_{\boldsymbol{s}_n}^H \boldsymbol{R}_\theta^{-1} \hat{\boldsymbol{\theta}}_{\boldsymbol{s}_n}. \end{aligned} \tag{20}$$

Equation (20) shows that the noncoherent metric $\mu(\boldsymbol{s}_n)$ can be written as the sum of a "coherent metric" with a term that reconciles the implicit channel estimate against the priors. (A similar observation was made in [11].)

B. Fast Metric Update

Since the sequence detection algorithm must compute $\mu(\boldsymbol{s}_n)$ at each time n, a fast algorithm to compute $\boldsymbol{\Sigma}_{\boldsymbol{s}_{n+1}}^{-1}$ from $\boldsymbol{\Sigma}_{\boldsymbol{s}_n}^{-1}$ is clearly of interest. Due to the rank-one update $\boldsymbol{\Sigma}_{\boldsymbol{s}_{n+1}} = \boldsymbol{\Sigma}_{\boldsymbol{s}_n} + \boldsymbol{a}_{\boldsymbol{s}_{n+1}} \boldsymbol{a}_{\boldsymbol{s}_{n+1}}^H$, the matrix inversion lemma yields

$$\boldsymbol{\Sigma}_{\boldsymbol{s}_{n+1}}^{-1} = \boldsymbol{\Sigma}_{\boldsymbol{s}_n}^{-1} - \frac{(\boldsymbol{\Sigma}_{\boldsymbol{s}_n}^{-1} \boldsymbol{a}_{\boldsymbol{s}_{n+1}})(\boldsymbol{\Sigma}_{\boldsymbol{s}_n}^{-1} \boldsymbol{a}_{\boldsymbol{s}_{n+1}})^H}{1 + \boldsymbol{a}_{\boldsymbol{s}_{n+1}}^H (\boldsymbol{\Sigma}_{\boldsymbol{s}_n}^{-1} \boldsymbol{a}_{\boldsymbol{s}_{n+1}})}. \tag{21}$$

With the aid of (21), it is straightforward to show that $\mu(\boldsymbol{s}_n)$ can be computed using $2N_b N_h(3+N_b N_h)+(n+1)(N_b N_h+1)$ multiplications.

C. Suboptimal Tree Search

We propose to perform an approximate minimization in (13) via tree search. While many options exist, we choose breadth-first search via the M-algorithm [6] since it offers near-optimum performance at low (and channel/SNR-independent) complexity. We now outline the M-algorithm assuming that all symbols in $\{s_n\}_{n=0}^{N-1}$ are unknown; a modification for known pilot/guard symbols will be described in the next section. Say that, at the n^{th} detection stage, the M-algorithm has a record of the M surviving length-n partial paths, where M is a design parameter. The M-algorithm then computes the metric (20) for each length-$(n+1)$ extension of these M paths and keeps only the best M of these extensions as survivors for the next stage. At the final stage, the best survivor is chosen as the full sequence estimate. Thus, $M|\mathcal{S}|$ metrics need to be computed at stage n. Using the fast algorithm of Section III-B, the total number of multiplications required to compute $\hat{\mathbf{s}}_{N-1}$ is $[2NN_b N_h(3+N_b N_h) + \frac{1}{2}N(N+1)(N_b N_h+1)]M|\mathcal{S}|$.

Figure 1 compares the complexity of the proposed technique to Kalman-PSP-LVA over a range of block lengths N and channel lengths N_h. In particular, Fig. 1 shows contours of the complexity ratio of Kalman-PSP-LVA relative to the proposed algorithm, where contour label p indicates that Kalman-PSP-LVA is 10^p times as complex. Thus, for (N, N_h) located above the "0"-labeled contour, the proposed algorithm is cheaper.

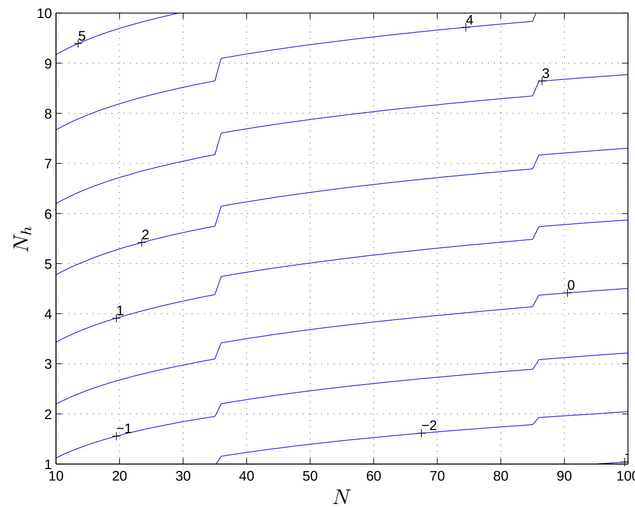

Fig. 1. Complexity of proposed scheme relative to Kalman-PSP-LVA for various combinations of channel length and block length. Contour label p indicates that Kalman-PSP-LVA is 10^p times as complex.

Roughly speaking, this corresponds to $N_h \geq 4$ and $N < 100$. In computing the figure, we assumed QPSK and $f_d T_s = 0.005$. The BEM size N_b (a function of N) was set equal to the number of significant eigenvalues of the channel correlation matrix. To evaluate Kalman-PSP-LVA complexity, we counted the multiplies required of the algorithm in [4], assuming $P \times P$ matrix inversion costs $\frac{4}{3}P^3 - 2P^2 + \frac{2}{3}P$ multiplies (e.g., using Gaussian elimination), and noting that [4] assumed a first-order Gauss-Markov channel model. For the M-algorithm, we used $M = 4$, and, for LVA, we used $L = 4$.

D. On Pilot/Guard Symbols

Because the metric (13) lacks an absolute phase reference, decoding ambiguities will exist for rotationally symmetric alphabets (e.g., QAM or PSK). Furthermore, gain ambiguities may exist for non-constant-modulus alphabets. These scalar ambiguities, however, can be resolved by the use of a single pilot. While we do not classify the methods described in this paper as "pilot-aided" per se (e.g., they unambiguously decode non-symmetric alphabets [12]), their performance (and the performance of noncoherent MLSD) can be significantly enhanced through judicious embedding of known symbols.

For example, the achievable diversity is strongly influenced by how the last $N_h - 1$ symbol locations are used. To see why, notice that a data symbol at index $N - k$ would contribute to the observations $\{r_n\}_{n=0}^{N-1}$ through the k channel coefficients $\{h_{:,l}\}_{l=0}^{k-1}$ and, hence, see at most k^{th}-order diversity. To prevent symbols transmitted in the last $N_h - 1$ locations from achieving less than N_h^{th}-order diversity—thereby dominating high-SNR error-rate, we advocate the use of length-$(N_h - 1)$ zero-padding. As another example, decoding complexity can be strongly influenced by how the first N_h symbol locations are used. If $N_p \geq N_h$ pilot symbols are placed at the beginning of the sequence, then they can be used to compute a reasonable initial channel estimate, allowing "calibration" of the noncoherent metric (20) before the M-algorithm discards partial paths. Such a calibration allows good M-algorithm performance with a small value of M (or, similarly, good PSP-LVA performance with a small value of L).

Note that a simple modification of the M-algorithm suffices to handle the case of arbitrary pilot/guard symbols: When the M-algorithm encounters a known symbol, each surviving path is given a single (rather than $|\mathcal{S}|$-ary) extension.

IV. Numerical Results

To generate DD channel realizations, we used a wide-sense stationary uncorrelated scattering (WSSUS) Jake's model [13] with uniform delay-power profile, for which $\mathrm{E}\{h_{n,l}h_{n-m,l-\ell}^*\} = \rho_m \sigma_l^2 \delta_\ell$ and $\sigma_l^2 = 1/N_h$. For the case of Rayleigh fading, $\rho_m = J_0(2\pi f_\mathsf{d} T_s m)$, where $f_\mathsf{d} T_s$ denotes the normalized single-sided Doppler spread and $J_0(\cdot)$ denotes the 0^{th}-order Bessel function of the first kind. As suggested by Section III-D, the symbol sequence included N_p leading pilots and $N_h - 1$ trailing zeros.

We consider two choices of receiver BEM: Karhunen Lòeve (KL) [14] and oversampled complex exponential (OCE) [15]. For the KL-BEM, $b_{n,p} = [\boldsymbol{V}]_{n,p}$, where the columns of $\boldsymbol{V}$ are the eigenvectors of the channel covariance matrix corresponding to the N_b largest eigenvalues. For the OCE-BEM basis, we start with $b_{n,p} = \frac{1}{\sqrt{N}} e^{j\frac{2\pi}{NK}(p - \frac{N_b-1}{2})n}$ with oversampling factor $K = 2$, determine the statistics $\boldsymbol{R}_\theta$ that best match the channel covariance, then de-correlate the basis vectors so that $\boldsymbol{R}_\theta$ becomes diagonal.

A. Suboptimality of Metric Simplification and M-Algorithm

Figure 2 compares the frame error rate (FER) of exact noncoherent MLSD in (12) to that of the simplified SD criterion in (13). To facilitate exact MLSD, the system parameters were chosen as $N = 10$, $N_h = 2$, $f_\mathsf{d} T_s = 0.005$, $N_p = 1$, and $\mathcal{S} =$ BPSK, where a KL-BEM with $N_b = 2$ was assumed at the receiver. It can be seen that the performance loss due to ML-metric simplification is negligible throughout the SNR range of interest. Figure 2 also shows the FER when the M-algorithm is used to approximately minimize (13), under various choices of M. It can be seen that choosing $M \geq 4$ results in performance that is within 0.2 dB of optimal.

B. Performance Versus SNR

Figures 3–4 compare the FER of the proposed noncoherent sequence detector (with $M = 6$) to Kalman-PSP-LVA (with $L = 4$) and to two genie-aided schemes: MLSD with perfectly known $\{h_{n,l}\}$, and MLSD with a genie-aided MMSE estimate of $\boldsymbol{\theta}$. In both figures, subplot (a) corresponds to $f_\mathsf{d} T_s = 0.002$ while subplot (b) corresponds to $f_\mathsf{d} T_s = 0.005$. For all experiments, $N = 64$, $N_h = 3$, $\mathcal{S} =$ BPSK, and $N_p = 9$. The length-N_p pilot sequence was the "spectrally efficient Kronecker delta" (SEKD) scheme from [16]—a cascade of N_b Kronecker delta sequences, each of length N_h.

The Kalman-PSP-LVA assumed a first-order Gauss-Markov channel whose parameters were experimentally optimized to minimize FER. The initial Kalman estimate was computed with the aid of the N_p-length pilot sequence. For genie-aided MLSD, $\boldsymbol{\theta}$ was estimated using a full frame of (randomly-generated) known symbols transmitted over an identical channel realization. We reason that MLSD with genie-aided MMSE-$\hat{\boldsymbol{\theta}}$ acts as a tight performance upper bound for any BEM-based noncoherent sequence detection scheme. In Fig. 3, the KL-BEM was employed, while in Fig. 4 the OCE-BEM was employed. All traces used BEM order $N_b = 3$.

Figures 3–4 show that, when the KL-BEM is used, the proposed noncoherent sequence detector comes within a fraction-of-a-dB from MLSD with genie-aided MMSE-$\hat{\boldsymbol{\theta}}$, and within about 1 dB from MLSD with perfect CSI, for all tested values of SNR and $f_\mathsf{d} T_s$. In comparison, Kalman-PSP-AR performs about 2 dB worse. When OCE-BEM is used, the proposed scheme performs similar to to Kalman-PSP-AR at low SNR, but seems to avoid (or postpone) the error floor exhibited by Kalman-PSP-AR at high SNR. From the high-SNR slope of the error traces, it can be seen that the proposed algorithm attains a diversity order of $N_h = 3$ for both BEMs.

C. Performance Versus $f_\mathsf{d} T_s$

In Fig. 5, the proposed KL-BEM and Kalman-PSP-LVA detectors were designed for fixed $f_\mathsf{d} T_s = 0.005$ but then tested on differing values of $f_\mathsf{d} T_s$. (The simulation parameters were the same as for Figures 3–4.) Figure 5 demonstrates the robustness of the proposed algorithm (as well as Kalman-PSP-LVA) to over-estimation of the Doppler spread. This is an important practical consideration because known statistics have been assumed throughout.

V. Conclusion

We proposed a scheme for near-optimal sequence detection of uncoded block transmissions over unknown doubly dispersive channels, focusing on the case of Gaussian channels with known statistics. Our scheme was based on suboptimal tree search of a simplified noncoherent ML metric, and it leveraged the M-algorithm in conjunction with fast update of MMSE-estimated BEM parameters. The inclusion of a few pilots, while not strictly necessary, was encouraged to facilitate the use of rotationally symmetric symbol alphabets, the use of facilitate low-complexity decoding, and the achievement of N_h^{th}-order diversity. Numerical experiments showed that the proposed algorithm performed within a fraction-of-a-dB from MLSD with genie-aided MMSE channel estimation, and about 1 dB from MLSD with perfect CSI. Furthermore, the proposed algorithm was found to be robust to mismatch in the assumed Doppler spread. The experiments also showed that the proposed algorithm outperformed Kalman-PSP with list-Viterbi, while at the same time requiring less complexity (assuming channel length $N_h \geq 4$ and block length $N < 100$).

References

[1] G. Forney, "Maximum-likelihood sequence estimation of digital sequences in the presence of inter symbol interference," *IEEE Trans. on Information Theory*, vol. 18, pp. 363–378, May 1972.

[2] R. A. Iltis, "A Bayesian maximum-likelihood sequence estimation algorithm for a priori unknown channels and symbol timing," *IEEE Journal on Selected Areas In Communications*, vol. 10, pp. 579–588, Apr. 1992.

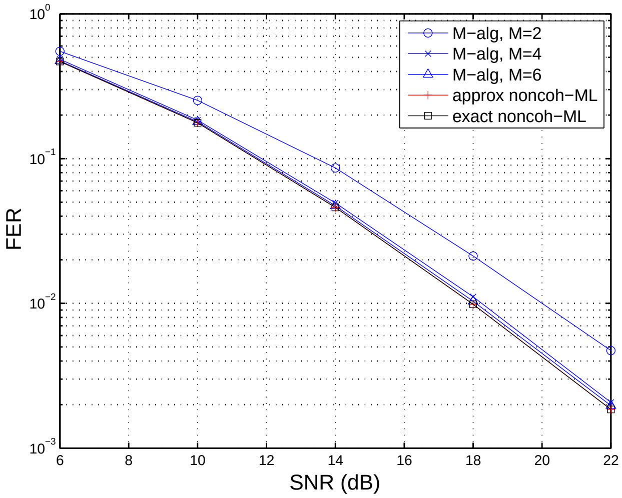

Fig. 2. FER of MLSD and various suboptimal schemes.

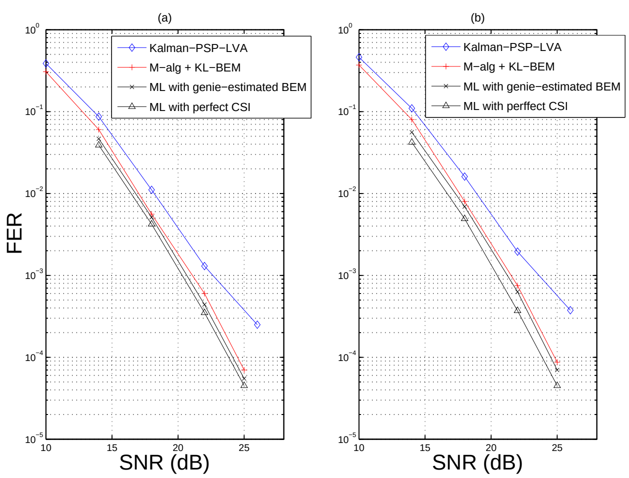

Fig. 3. FER for KL-BEM at (a) $f_d T_s = 0.002$, (b) $f_d T_s = 0.005$.

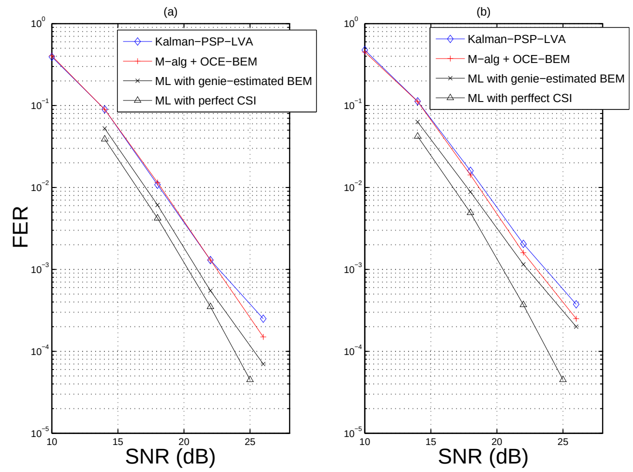

Fig. 4. FER for OCE-BEM at (a) $f_d T_s = 0.002$, (b) $f_d T_s = 0.005$.

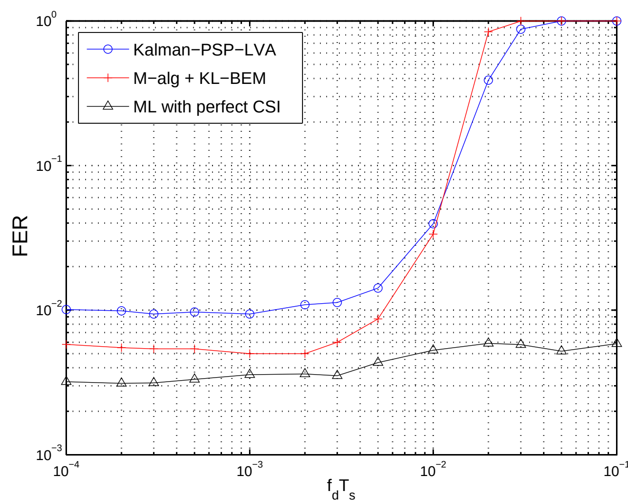

Fig. 5. FER versus $f_d T_s$ for receivers that assume $f_d T_s = 0.005$.

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

[4] H. Chen, K. Buckley, and R. Perry, "Time-recursive maximum likelihood based sequence estimation for unknown ISI channels," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, pp. 1005–1009, 2000.

[5] H. Kubo, K. Murakami, and T. Fujino, "An adaptive maximum-likelihood sequence estimator for fast time-varying intersymbol interference channels," *IEEE Trans. on Communications*, vol. 42, pp. 1872–1880, Feb./Mar./Apr. 1994.

[6] J. B. Anderson and S. Mohan, "Sequential decoding algorithms: A survey and cost analysis," *IEEE Trans. on Communications*, vol. 32, pp. 169–172, 1984.

[7] M. K. Tsatsanis and G. B. Giannakis, "Modeling and equalization of rapidly fading channels," *Internat. Journal of Adaptive Control & Signal Processing*, vol. 10, pp. 159–176, Mar. 1996.

[8] H. Chen, R. Perry, and K. Buckley, "On MLSE algorithms for unknown fast time-varying channels," *IEEE Trans. on Communications*, vol. 51, pp. 730–734, May 2003.

[9] A. E.-S. El-Mahdy, "Adaptive channel estimation and equalization for rapidly mobile communication channels," *IEEE Trans. on Communications*, vol. 52, pp. 1126–1135, July 2004.

[10] A. Mammela and D. P. Taylor, "Bias terms in the optimal quadratic receiver," *IEEE Communications Letters*, vol. 2, pp. 57–58, Feb. 1998.

[11] D. J. Reader and W. G. Cowley, "Blind maximum likelihood sequence detection over fast fading channels," in *Proc. European Signal Processing Conf.*, 1996.

[12] B. D. Hart, "Maximum likelihood sequence detection using a pilot tone," *IEEE Trans. on Vehicular Technology*, vol. 49, pp. 550–560, Mar. 2000.

[13] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 4th ed., 2001.

[14] M. Visintin, "Karhunen-Lòeve expansion of a fast Rayleigh fading process," *Electronics Letters*, vol. 32, no. 18, pp. 1712–1713, 1996.

[15] T. A. Thomas and F. W. Vook, "Multi-user frequency-domain channel identification, interference suppression, and equalization for time-varying broadband wireless communications," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop*, (Boston, MA), pp. 444–448, Mar. 2000.

[16] A. P. Kannu and P. Schniter, "On the spectral efficiency of noncoherent doubly selective channels," in *Proc. Allerton Conf. on Communication, Control, and Computing*, Oct. 2006.