Efficient Communication over Highly Spread Underwater Acoustic Channels*

Sung-Jun Hwang
Dept. ECE, The Ohio State University
2015 Neil Ave., Columbus OH 43210
hwangsu@ece.osu.edu

Philip Schniter
Dept. ECE, The Ohio State University
2015 Neil Ave., Columbus OH 43210
schniter@ece.osu.edu

ABSTRACT

In this paper we propose a novel method for communication over underwater acoustic channels for which the product of delay-spread and Doppler-spread is large, such as those pertaining to the surf zone. In particular, we propose the use of pulse-shaped multicarrier modulation to convert the doubly dispersive channel into an inter-carrier interference (ICI) channel with small ICI spread. We then propose a novel joint ICI-estimation/data-detection strategy which performs near optimally yet at low complexity, due to the use of an efficient tree search algorithm and the ability to leverage sparseness in the delay-power profile. The delay-power profile, which itself varies in time, can be readily tracked via pilots. Numerical simulations show that our technique gives performance close to genie-aided bounds over highly spread channels.

Categories and Subject Descriptors

C.3 Computer Systems Organization [**SPECIAL-PURPOSE AND APPLICATION-BASED SYSTEMS**]: Signal processing systems

General Terms

Algorithms, Design, Theory, Performance

1. INTRODUCTION

Underwater networks are becoming increasingly important in facilitating systems that perform remote measurement, autonomous monitoring, littoral anti-submarine warfare, mine countermeasure, and other functions. The impracticality of cables and the impossibility of radio transmission imply that communication between network nodes must take place acoustically. The difficulty of this task should not be underestimated, since experts consider the underwater acoustic channel (UAC) to be "quite possibly nature's more unforgiving wireless medium" [2]. In fact, reliable high-rate underwater acoustic communication remains a principal challenge in the construction of underwater networks.

The physical characteristics of the UAC are highly dependent on the distance and relative movement between the transmitter and receiver; the proximity, roughness, and motion of the scattering surfaces; and the presence of ambient interference. However, the UAC characteristics that pose the primary challenges for data communication can be summarized as *simultaneously large delay- and Doppler-spreads*, *limited bandwidth*, and *limited receiver complexity*. These challenges can be understood as follows. High delay-spread implies that single-carrier communication will be plagued by inter-symbol interference (ISI) that, for practical signal bandwidths, spans hundreds of symbols. High Doppler-spread then implies that this ISI response will change quickly in time. Since optimal mitigation of this long and quickly-varying ISI response becomes computationally infeasible, practitioners have resorted to simple sub-optimal strategies such as the adaptive decision-feedback equalization (DFE) [19, 25]. However, these implementable single-carrier techniques perform far short of optimal and fail altogether in very highly spread environments such as the surf zone [20].

As an alternative, multi-carrier modulation (MCM) has been proposed to increase the symbol interval and thereby decrease the ISI span. While a number of MCM proposals for the UAC have been made over the years (see, e.g., the references in [13, 18, 24]), none seem to have been successful enough to displace single-carrier/DFE as the practical method of choice. The primary difficulty in applying MCM to the doubly dispersive UAC is that, as the symbol interval is increased (to reduce ISI span), the subcarrier spacing must be decreased (to preserve data rate), making the system more susceptible to Doppler-spread-induced inter-carrier interference (ICI). While ISI and ICI could be simultaneously reduced via long symbols and distant subcarriers, this would necessitate low spectral efficiency, which is impractical given the UAC's already limited bandwidth.

The MCM proposals recently presented at the IEEE/MTS OCEANS-2006 conference help to illuminate the challenges in applying MCM to the UAC. For example, in [13, 24], the authors proposed the use of classical ZP-OFDM signaling schemes and assumed Doppler-spreads small enough to induce negligible ICI. For these schemes, the universal

*This work was supported by the National Science Foundation under grant 0237037 and the Office of Naval Research under grant N00014-07-1-0209.

WUWNet'07 September 14, 2007, Montréal, Québec, Canada

ICI bound[1] in [15] implies that an ICI power of -25 dB (which we consider "negligible") occurs when $f_{\mathsf{D}}T_s = 0.03$, where f_{D} denotes single-sided Doppler-spread and T_s symbol duration. Since T_s was chosen as $7T_h$ (as $3.4T_h$) in [13] (in [24]), where T_h denotes channel delay-spread, we deduce that these schemes can handle UACs with delay/Doppler spreading products of at most $f_{\mathsf{D}}T_h = 0.004$ ($f_{\mathsf{D}}T_h = 0.009$). The surf-zone channels described in [20], however, yield $f_{\mathsf{D}}T_h \approx 0.05$, which is more than 10 times (5 times) as severe. As another example, in the non-traditional MCM approach [18], the symbol length was chosen shorter than that needed for perfect ISI-suppression, in order to tolerate high Doppler-spread while keeping ICI negligible. The resulting ISI-span was short enough to enable the use of sophisticated joint estimation/detection techniques (i.e., LMS/Viterbi per-survivor processing [21]), which were shown to significantly outperform the traditional adaptive DFE. But [18] only demonstrated the ability to handle $f_{\mathsf{D}}T_h = 0.0035$ via simulation and $f_{\mathsf{D}}T_h \approx 0.002$ experimentally, which are over 10 times milder than the surf-zone channels discussed in [20].

Communication over doubly dispersive channels, i.e., those with large delay/Doppler-spread product, is a challenging problem that has received significant attention from researchers over the last two decades. Most researchers have approached this problem through the design of MCM pulse-shapes which minimize total ISI/ICI power [11, 12, 16, 26], with the goal that ISI/ICI is rendered negligible. But even with optimized pulses, ISI/ICI remains non-negligible for surf-zone-like channels unless the pulses are designed to sacrifice about 50% in spectral efficiency, a fundamental consequence of the Balian-Low theorem [26]. Given the bandwidth limitations of the UAC, such sacrifices should be avoided. An alternative approach, suggested by the authors, is to allow a small ICI span (e.g., 2-3 subcarriers) and design pulses which minimize *residual* ISI/ICI power [4, 22]. Here, near-perfect residual-ISI/ICI suppression can be accomplished *without* loss of spectral efficiency, and high-performance dominant-ICI mitigation can be accomplished with low complexity. We propose to employ these ICI-shaping pulses for communication over the UAC.

The aforementioned ISI/ICI mitigation schemes require (implicitly or explicitly) accurate channel state information (CSI). Maintaining this CSI is especially difficult when the channel is doubly dispersive, due to the number and rate-of-change of the channel coefficients. Pilot-aided transmission (PAT) is a practical means of aiding data reception in the presence of channel uncertainty [27]. When PAT is used with *decoupled* estimation/detection (DED) schemes, a channel estimate is first obtained via pilots and later used for coherent data detection. With DED, it is important to keep the channel estimates free of interference from unknown data (especially at high SNR), though doing so with a doubly dispersive channel requires time/frequency guard-bands which reduce spectral efficiency [10]. When PAT is used with *joint* estimation/detection (JED), there is no need to keep pilots and data separate, allowing the overall system to achieve the high-SNR ergodic capacity of the doubly dispersive channel, i.e., $(1 - 2f_{\mathsf{D}}T_h)\log_2 \mathsf{SNR}$ bits/sec/Hz [10]. The bandwidth-limited nature of the UAC motivates us to consider PAT with JED.

As one would expect, the complexity of optimal JED is prohibitive. Hence, practical JED is often based on the Viterbi algorithm (VA) with per-survivor processing (PSP), whereby a separate channel estimate is calculated for each surviving path in the trellis [21]. Because the number of trellis states (and hence paths) can be quite large, PSP channel estimation is often accomplished using the simple LMS algorithm. With rapidly varying channels, however, PSP-VA-LMS performance floors at medium SNR, and enhancements based on the list-Viterbi algorithm (LVA) and Kalman channel tracking are needed to keep PSP performing well throughout the practical SNR range [3].

Motivated by the high complexity of PSP-LVA-Kalman, the authors recently proposed a block-based technique, based on a tree-search algorithm which leverages a basis-expansion model (BEM) for the time-varying channel, that was found to outperform PSP-LVA-Kalman at much lower complexity [8]. In particular, the per-symbol complexity of this JED scales as $\mathcal{O}(N_h^2 D^2)$, where $N_h = \frac{T_h}{T_s/N}$ denotes the delay-spread (in chips) and $D = \lceil f_{\mathsf{D}}T_s \rceil$ the single-sided Doppler spread (in subcarriers). But with surf-zone UAC parameters, even the complexity of even this tree-search JED may be prohibitive. For example, if 256 subcarriers were used to transmit a 24 kHz bandwidth signal over the surf-zone UAC from [20], one can expect $N_h \approx 168$ and $f_{\mathsf{D}}T_s \approx 0.15$, for which the N_h^2 dependence becomes problematic.

The solution to the complexity riddle may lie in the sparse nature of realistic UAC responses [5, 6, 14]. For example, if only 1/3 of the channel's N_h delay taps are significant, then a reception algorithm whose complexity is quadratic in the *active* delay taps will save in processing by a factor of 9. But the design of a high-performance system which leverages sparseness is non-trivial. For example, most sparseness-leveraging algorithms are based on single-carrier transmission with adaptive DFE reception (e.g., [14]), whose performance is known to fall far short of optimal, and it is not clear how to exploit sparseness in the recently proposed MCM schemes [13, 18, 24]. To further complicate matters, there exists the challenge of accurately tracking the sparseness structure. In UACs, for example, it has been observed that the tap delay locations can change almost as quickly as the tap gains [20]. While clever order-recursive matching-pursuit algorithms have been proposed to track the structure of sparse UACs (e.g., [14]), their complexity remains quadratic in the dictionary length (i.e., N_h), which is no better than our *non-sparse* tree-search algorithm. So, we need a tracking algorithm that is much less complex.

In this paper, we propose a tree-search-based JED strategy, building on our earlier work [8], that operates on pulse-shaped MCM transmissions and that takes full advantage of sparseness in the channel delay profile. In addition, we propose a simple yet accurate means of tracking the channel's quickly-varying delay-power profile (DPP). The performance of our algorithm is ascertained numerically using simulated channels whose sparseness and overall delay/Doppler-spread product $f_{\mathsf{D}}T_h$ mimic those of the surf-zone channels in [20]. The proposed scheme is found to exhibit BER performance close to genie-aided bounds while maintaining high spectral efficiency and low computational complexity.

Notation: We use $(\cdot)^T$ to denote transpose, $(\cdot)^*$ conjugate, and $(\cdot)^H$ conjugate transpose, and we use $[\boldsymbol{B}]_{m,n}$ to denote the element in the m^{th} row and n^{th} column of matrix $\boldsymbol{B}$, where row/column indices begin with zero. We also use $\mathcal{D}(\boldsymbol{b})$ to denote the diagonal matrix created from vector $\boldsymbol{b}$,

[1] [15] confirms that this bound is tight for $f_{\mathsf{D}}T_s = 0.3$.

$\boldsymbol{I}_K$ the $K \times K$ identity matrix, and $\{\delta_k\}$ the Kronecker delta sequence. Finally, we use $\odot$ to denote element-wise multiplication and $\mathrm{E}\{\cdot\}$ expectation.

2. SYSTEM MODEL

2.1 Pulse-Shaped MCM

We consider discrete-time complex-baseband multi-carrier modulation (MCM) with N subcarriers, where the i^{th} MCM symbol is composed of QAM symbols $\{s_k^{(i)}\}_{k=0}^{N-1}$ drawn from the finite alphabet $\mathcal{S}$. As shown in (1), the transmitted sequence $\{t_n\}$ is generated by transforming the i^{th} QAM symbol vector via N-point inverse discrete Fourier transform (DFT), applying an N_α-point modulation pulse $\{\alpha_n\}_{n=0}^{N_\alpha-1}$ to its cyclic extension, and superimposing the result N samples behind the contribution from the $(i-1)^{th}$ MCM symbol.

$$t_n \;=\; \sum_{i=-\infty}^{\infty} \alpha_{n-iN} \cdot \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} s_k^{(i)} e^{j\frac{2\pi}{N}k(n-iN)}. \qquad (1)$$

(See [4] for details.) A noisy linear time-varying channel then produces the received samples

$$r_n \;=\; \sum_{l=0}^{N_h-1} h_{n,l} t_{n-l} + v_n, \qquad (2)$$

where $\{h_{n,l}\}_{l=0}^{N_h-1}$ denotes the length-N_h discrete impulse response at time n, and where $\{v_n\}$ is zero-mean circular white Gaussian noise (CWGN) with covariance σ^2. Focusing on demodulation of the i^{th} MCM symbol, we define $r_n^{(i)} := r_{iN+n}$, $v_n^{(i)} := v_{iN+n}$, and $h_{n,l}^{(i)} := h_{iN+n,l}$ and rewrite (2) as

$$r_n^{(i)} \;=\; \sum_{l=0}^{N_h-1} h_{n,l}^{(i)} \sum_{\ell=-\infty}^{\infty} \alpha_{\ell N+n-l} \times \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} s_k^{(i-\ell)} e^{j\frac{2\pi}{N}k(n-l+\ell N)} + v_n^{(i)}. \qquad (3)$$

For demodulation, the receiver applies the length-N_β pulse $\{\beta_n\}_{n=0}^{N_\beta-1}$ prior to an N-point DFT, eventually yielding the frequency-domain observations $\{x_d^{(i)}\}_{d=0}^{N-1}$:

$$x_d^{(i)} \;=\; \frac{1}{\sqrt{N}} \sum_{n=0}^{N_\beta-1} r_n^{(i)} \beta_n e^{-j\frac{2\pi}{N}dn}. \qquad (4)$$

Putting (1)-(4) together, it is straightforward to show that

$$x_d^{(i)} \;=\; \sum_{\ell=-\infty}^{\infty} \sum_{k=0}^{N-1} H_{d-k,k}^{(i,\ell)} s_k^{(i-\ell)} + w_d^{(i)}, \qquad (5)$$

where

$$H_{d,k}^{(i,\ell)} \;:=\; \frac{1}{N} \sum_{n=0}^{N_\beta-1} \sum_{l=0}^{N_h-1} h_{n,l}^{(i)} \beta_n \alpha_{\ell N+n-l} e^{-j\frac{2\pi}{N}[dn+k(l-\ell N)]} \qquad (6)$$

$$w_d^{(i)} \;:=\; \frac{1}{\sqrt{N}} \sum_{n=0}^{N_\beta-1} \beta_n v_n^{(i)} e^{-j\frac{2\pi}{N}dn}. \qquad (7)$$

In writing (5), we used the fact that $H_{d,k}^{(i,\ell)}$ is N-cyclic in the indices d and k. Defining $\boldsymbol{x}^{(i)} := [x_0^{(i)}, \ldots, x_{N-1}^{(i)}]^T$, $\boldsymbol{s}^{(i)} := [s_0^{(i)}, \ldots, s_{N-1}^{(i)}]^T$, and $\boldsymbol{w}^{(i)} := [w_0^{(i)}, \ldots, w_{N-1}^{(i)}]^T$, (5) can be written in vector form as

$$\boldsymbol{x}^{(i)} \;=\; \sum_{\ell=-\infty}^{\infty} \sum_{k=0}^{N-1} \mathcal{D}_k(\boldsymbol{s}^{(i-\ell)}) \boldsymbol{H}_{-k}^{(i,\ell)} + \boldsymbol{w}^{(i)} \qquad (8)$$

$$\boldsymbol{H}_d^{(i,\ell)} \;:=\; [H_{d,-d}^{(i,\ell)}, H_{d,-d+1}^{(i,\ell)}, \ldots, H_{d,N-1-d}^{(i,\ell)}]^T \in \mathbb{C}^N, \qquad (9)$$

where $\mathcal{D}_k(\cdot)$ denotes the diagonal matrix created from the k-place cyclic downward shift of its vector argument.

The techniques in [4] can be used to design pulses $\{\alpha_n\}$ and $\{\beta_n\}$ that yield both negligible ISI:

$$H_{d,k}^{(i,\ell)} \;\approx\; 0 \quad \text{for} \quad \ell \neq 0, \qquad (10)$$

as well as negligible ICI beyond a radius of $D := \lceil f_{\mathrm{D}} T_c N \rceil$ subcarriers:

$$H_{d,k}^{(i,\ell)} \;\approx\; 0 \quad \text{for} \quad D < d < N-D. \qquad (11)$$

Here, f_{D} denotes the single-sided Doppler spread in Hz and T_c denotes the sampling (or "chip") interval in seconds. Under these conditions, the system model (8) simplifies to

$$\boldsymbol{x}^{(i)} \;=\; \sum_{k=-D}^{D} \mathcal{D}_k(\boldsymbol{s}^{(i)}) \boldsymbol{H}_{-k}^{(i)} + \boldsymbol{w}^{(i)} \qquad (12)$$

$$\boldsymbol{H}_d^{(i)} \;:=\; [H_{d,-d}^{(i,0)}, H_{d,-d+1}^{(i,0)}, \ldots, H_{d,N-1-d}^{(i,0)}]^T \in \mathbb{C}^N. \qquad (13)$$

2.2 A Sparse BEM

We now develop a sparse basis expansion model (BEM) for the frequency-domain channel (13) and use it to rewrite the system model in a form convenient for sequential detection.

Using (6), $\boldsymbol{H}_d^{(i)}$ can be written as

$$\boldsymbol{H}_d^{(i)} \;=\; \boldsymbol{F} \breve{\boldsymbol{\theta}}_d^{(i)} \qquad (14)$$

$$\breve{\boldsymbol{\theta}}_d^{(i)} \;:=\; \mathcal{D}(\boldsymbol{f}_d^* \sqrt{N})(\boldsymbol{\mathcal{H}}^{(i)} \odot \boldsymbol{\mathcal{P}})^T \underline{\boldsymbol{f}}_d \in \mathbb{C}^N, \qquad (15)$$

where $\boldsymbol{F} \in \mathbb{C}^{N\times N}$, $\boldsymbol{\mathcal{H}}^{(i)} \in \mathbb{C}^{N_\beta \times N}$, and $\boldsymbol{\mathcal{P}} \in \mathbb{C}^{N_\beta \times N}$ are defined element-wise as

$$[\boldsymbol{F}]_{n,m} \;:=\; \frac{1}{\sqrt{N}} e^{-j\frac{2\pi}{N}nm} \qquad (16)$$

$$[\boldsymbol{\mathcal{H}}^{(i)}]_{n,l} \;:=\; h_{n,l}^{(i)} \qquad (17)$$

$$[\boldsymbol{\mathcal{P}}]_{n,l} \;:=\; \beta_n \alpha_{n-l}, \qquad (18)$$

where $\boldsymbol{f}_d \in \mathbb{C}^N$ denotes the d^{th} column of the DFT matrix $\boldsymbol{F}$, and $\underline{\boldsymbol{f}}_d \in \mathbb{C}^{N_\beta}$ denotes its N_β-length (cyclic) extension:

$$\underline{\boldsymbol{f}}_d \;:=\; [e^{-j\frac{2\pi}{N}0}, e^{-j\frac{2\pi}{N}1}, \ldots, e^{-j\frac{2\pi}{N}(N_\beta-1)}]^T. \qquad (19)$$

Equation (14) can be recognized as an N^{th}-order (i.e., exact) BEM for the channel vector $\boldsymbol{H}_d^{(i)}$: the columns of $\boldsymbol{F}$ are the basis vectors and elements of $\breve{\boldsymbol{\theta}}_d^{(i)}$ are the basis coefficients.

We now show how sparseness in the impulse response can be leveraged to reduce the BEM order. Notice that if only $N_a < N_h$ taps of the impulse response $\{h_{n,l}^{(i)}\}_{l=0}^{N_h-1}$ are non-zero over the time duration $n \in \{0, \ldots, N_\beta - 1\}$, then only N_a columns of $\boldsymbol{\mathcal{H}}^{(i)}$ will be non-zero, implying that only N_a BEM coefficients in $\breve{\boldsymbol{\theta}}_d^{(i)}$ will be non-zero. In this case, the BEM order can be reduced to N_a without loss. To make this more precise, let $\mathcal{L}^{(i)}$ denote the set of taps which are active during the i^{th} multicarrier symbol interval,

$$\mathcal{L}^{(i)} \;=\; \big\{l : h_{n,l}^{(i)} \neq 0 \text{ for some } n \in \{0, \ldots, N_\beta - 1\}\big\} \qquad (20)$$

where $|\mathcal{L}^{(i)}| = N_a$. Constructing $\boldsymbol{B}^{(i)} \in \mathbb{C}^{N\times N_a}$ from the columns of $\boldsymbol{F}$ with indices in $\mathcal{L}^{(i)}$, and constructing $\boldsymbol{\theta}_d^{(i)} \in \mathbb{C}^{N_a}$ from the corresponding elements of $\breve{\boldsymbol{\theta}}_d^{(i)}$, equation (14) can be restated as

$$\boldsymbol{H}_d^{(i)} = \boldsymbol{B}^{(i)}\boldsymbol{\theta}_d^{(i)}. \tag{21}$$

Using this sparse BEM, (12) can be rewritten as

$$\boldsymbol{x}^{(i)} = \left[\mathcal{D}_D(\boldsymbol{s}^{(i)})\boldsymbol{B}^{(i)}, \cdots, \mathcal{D}_{-D}(\boldsymbol{s}^{(i)})\boldsymbol{B}^{(i)}\right]\boldsymbol{\theta}^{(i)} + \boldsymbol{w}^{(i)} \tag{22}$$

$$\boldsymbol{\theta}^{(i)} := [\boldsymbol{\theta}_{-D}^{(i)T}, \ldots, \boldsymbol{\theta}_{D}^{(i)T}]^T. \tag{23}$$

2.3 Modifications for Sequence Detection

In Section 3, we will describe how tree-search can be used to recover the data symbols $\boldsymbol{s}^{(i)}$ from the frequency-domain observations $\boldsymbol{x}^{(i)}$. Before describing the tree-search, we place a few restrictions on the model (22).

From (22), it can be seen that every element in $\boldsymbol{x}^{(i)}$ sees contributions from $2D+1$ unknown data symbols. For tree search, we would like that the first observation contains a contribution from only one unknown symbol, the second contains contributions from only two unknown symbols, the third from only three unknown symbols, and so on. An easy way to ensure this is to set $\{s_k^{(i)}\}_{k=-D+1}^{D-1} = 0$, i.e., "to turn off" the first and last D subcarriers—a technique commonly used to prevent adjacent-channel interference. Note that the loss in spectral efficiency will be small when $2D \ll N$.

To proceed further, we find it convenient to define the D-shifted quantities $\breve{s}_k^{(i)} = s_{\langle k+D\rangle_N}^{(i)}$ and $\breve{\boldsymbol{s}}^{(i)} := [\breve{s}_0^{(i)}, \ldots, \breve{s}_{N-1}^{(i)}]^T$, noticing that the last $2D$ elements in $\breve{\boldsymbol{s}}^{(i)}$ constitute a zero-valued guard interval. Since $\mathcal{D}_k(\breve{\boldsymbol{s}}^{(i)}) = \mathcal{D}_{k+D}(\boldsymbol{s}^{(i)})$ for any k, we can rewrite (22) as

$$\boldsymbol{x}^{(i)} = \boldsymbol{A}^{(i)}\boldsymbol{\theta}^{(i)} + \boldsymbol{w}^{(i)} \tag{24}$$

$$\boldsymbol{A}^{(i)} := \left[\mathcal{D}_{2D}(\breve{\boldsymbol{s}}^{(i)})\boldsymbol{B}^{(i)}, \ldots, \mathcal{D}_0(\breve{\boldsymbol{s}}^{(i)})\boldsymbol{B}^{(i)}\right] \tag{25}$$

and see that, for each $k \in \{0, \ldots, N-1\}$, the observations $\{x_d^{(i)}\}_{d=0}^k$ depend only on $\{\breve{s}_d^{(i)}\}_{d=0}^k$. The now "causal" nature of the ICI channel allows us to write the partial observation vector $\boldsymbol{x}_k^{(i)} := [x_0^{(i)}, \ldots, x_k^{(i)}]^T$ as

$$\boldsymbol{x}_k^{(i)} = \boldsymbol{A}_k^{(i)}\boldsymbol{\theta}^{(i)} + \boldsymbol{w}_k^{(i)}, \tag{26}$$

where $\breve{\boldsymbol{w}}_k^{(i)} := [w_0^{(i)}, \ldots w_k^{(i)}]^T$ and where $\boldsymbol{A}_k^{(i)}$ appends a new row with each k:

$$\boldsymbol{A}_k^{(i)} = \begin{bmatrix} \boldsymbol{a}_0^{(i)H} \\ \vdots \\ \boldsymbol{a}_k^{(i)H} \end{bmatrix} \tag{27}$$

$$\boldsymbol{a}_k^{(i)H} = [\breve{s}_{k-2D}^{(i)}\boldsymbol{b}_k^{(i)H}, \ldots, \breve{s}_k^{(i)}\boldsymbol{b}_k^{(i)H}]. \tag{28}$$

In (28), the k-indexing on $\breve{s}_k^{(i)}$ is performed modulo-N and $\boldsymbol{b}_k^{(i)H}$ denotes the k^{th} row of $\boldsymbol{B}^{(i)}$.

We now investigate $\boldsymbol{R}_{\theta_m,\theta_n}^{(i)} := \mathrm{E}\{\boldsymbol{\theta}_m^{(i)}\boldsymbol{\theta}_n^{(i)H}\}$, a key parameter in tree-search. If we assume that the channel obeys the wide-sense stationary uncorrelated scattering (WSSUS) assumption over the duration of one MCM symbol, i.e.,

$$\mathrm{E}\{h_{n,l}^{(i)}h_{n-k,l-m}^{(i)*}\} = \rho_k\sigma_l^2\delta_m \text{ for } n \in \{0, \ldots, N_\beta - 1\}, \tag{29}$$

where $\{\sigma_l^2\}_{l=0}^{N_h-1}$ denotes the delay-power profile (DPP) and $\{\rho_k\}$ denotes the tap autocorrelation sequence (normalized so that $\rho_0 = 1$), then $\boldsymbol{R}_{\theta_m,\theta_n}^{(i)}$ is a diagonal matrix with

$$\left[\boldsymbol{R}_{\theta_m,\theta_n}^{(i)}\right]_{\nu,\nu} = \sigma_{l_\nu^{(i)}}^2 \sum_{k=-N_\beta+1}^{N_\beta-1} \rho_k\gamma_{m,n,k,\nu}^{(i)}. \tag{30}$$

In (30), $l_\nu^{(i)}$ denotes the index of the ν^{th} sparse tap (i.e., $\mathcal{L}^{(i)} = \{l_0^{(i)}, l_1^{(i)}, \ldots, l_{N_a-1}^{(i)}\}$) and

$$\gamma_{m,n,k,\nu}^{(i)} = \frac{1}{N}\sum_{p,q:p-q=k} \beta_p\breve{\alpha}_{m,p-l_\nu^{(i)}}\beta_q^*\breve{\alpha}_{n,q-l_\nu^{(i)}}^*$$

$$\breve{\alpha}_{m,p} := \alpha_p e^{-j\frac{2\pi}{N}mp}.$$

The autocovariance matrix $\boldsymbol{R}_{\theta_m}^{(i)} := \mathrm{E}\{\boldsymbol{\theta}_m^{(i)}\boldsymbol{\theta}_m^{(i)H}\}$ is also diagonal with

$$\left[\boldsymbol{R}_{\theta_m}^{(i)}\right]_{\nu,\nu} = \frac{\sigma_{l_\nu^{(i)}}^2}{N}\sum_{k=-N_\beta+1}^{N_\beta-1}\rho_k\sum_{p,q:p-q=k}\beta_p\alpha_{p-l_\nu^{(i)}}\beta_q^*\alpha_{q-l_\nu^{(i)}}^* e^{-j\frac{2\pi}{N}mk} \tag{31}$$

which corresponds to the power spectral density of the pulse-shaped channel.

3. SEQUENCE DETECTION

In this section, we describe a fast algorithm to decode $\breve{\boldsymbol{s}}^{(i)}$ from the observations $\boldsymbol{x}^{(i)}$ assuming knowledge of the sparse basis $\boldsymbol{B}^{(i)}$, but not of the BEM coefficients $\boldsymbol{\theta}^{(i)}$. Estimation of $\boldsymbol{B}^{(i)}$ (i.e., tracking of active taps) will be discussed in Section 4.1. Our algorithm is sequential in that it estimates the partial sequence $\breve{\boldsymbol{s}}_k^{(i)} := [\breve{s}_0^{(i)}, \ldots, \breve{s}_k^{(i)}]^T$ for $k = 0, 1, 2, \ldots$, eventually estimating the full sequence $\breve{\boldsymbol{s}}^{(i)} = \breve{\boldsymbol{s}}_{N-1}^{(i)}$. In deriving our algorithm, we employ the sparse model (26) and assume that $\boldsymbol{\theta}^{(i)}$ is zero-mean circular Gaussian with a full-rank autocovariance matrix $\boldsymbol{R}_\theta^{(i)} := \mathrm{E}\{\boldsymbol{\theta}^{(i)}\boldsymbol{\theta}^{(i)H}\}$ whose contents can be inferred from (23) and (30). To keep the development concise, we drop the symbol index i.

3.1 The Noncoherent Metric

The MLSD estimate is defined as

$$\hat{\boldsymbol{s}}_{\mathsf{ML},k} = \arg\max_{\breve{\boldsymbol{s}}_k} p(\boldsymbol{x}_k|\breve{\boldsymbol{s}}_k), \tag{32}$$

where, marginalizing over the channel (26),

$$\begin{aligned} &p(\boldsymbol{x}_k|\breve{\boldsymbol{s}}_k) \\ &= \int_{\boldsymbol{\theta}} p(\boldsymbol{x}_k|\breve{\boldsymbol{s}}_k,\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta} \\ &= \int_{\boldsymbol{\theta}} \frac{1}{(\pi\sigma^2)^k}\exp\left\{-\frac{1}{\sigma^2}\|\boldsymbol{x}_k - \boldsymbol{A}_k\boldsymbol{\theta}\|^2\right\}p(\boldsymbol{\theta})d\boldsymbol{\theta}. \end{aligned}$$

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

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

where

$$\boldsymbol{\Sigma}_k := \boldsymbol{A}_k^H \boldsymbol{A}_k + \sigma^2 \boldsymbol{R}_\theta^{-1}, \tag{33}$$

and where $\{C_i\}$ are constants irrelevant to the maximization in (32). Using the monotonicity of $\log(\cdot)$, we can write

$$\hat{\boldsymbol{s}}_{\mathsf{ML},k} = \arg\max_{\check{\boldsymbol{s}}_k} \Big\{ \sigma^{-2} \boldsymbol{x}_k^H \big(\boldsymbol{A}_k \boldsymbol{\Sigma}_k^{-1} \boldsymbol{A}_k^H - \boldsymbol{I}_{k+1} \big) \boldsymbol{x}_k - \log\det(\sigma^{-2}\boldsymbol{\Sigma}_k) \Big\}. \tag{34}$$

As reported elsewhere (e.g., [8, 17]), the maximization in (34) can be simplified by ignoring the bias term $\log\det(\sigma^{-2}\boldsymbol{\Sigma}_k)$, yielding the quasi-ML detection rule

$$\hat{\boldsymbol{s}}_k = \arg\max_{\check{\boldsymbol{s}}_k} \boldsymbol{x}_k^H \boldsymbol{A}_k \boldsymbol{\Sigma}_k^{-1} \boldsymbol{A}_k^H \boldsymbol{x}_k. \tag{35}$$

We will now show that the detection rule (35) performs implicit minimum mean-squared error (MMSE) estimation of $\boldsymbol{\theta}$. Using $\hat{\boldsymbol{\theta}}_k$ to denote the MMSE estimate of $\boldsymbol{\theta}$ from $\boldsymbol{x}_k$ under hypothesis $\check{\boldsymbol{s}}_k$, we have

$$\hat{\boldsymbol{\theta}}_k = \mathrm{E}\{\boldsymbol{\theta}\boldsymbol{x}_k^H | \check{\boldsymbol{s}}_k\} \, \mathrm{E}\{\boldsymbol{x}_k \boldsymbol{x}_k^H | \check{\boldsymbol{s}}_k\}^{-1} \boldsymbol{x}_k \tag{36}$$

$$= \boldsymbol{R}_\theta \boldsymbol{A}_k^H \boldsymbol{\Phi}_k \boldsymbol{x}_k, \tag{37}$$

where we used the definition

$$\boldsymbol{\Phi}_k := \mathrm{E}\{\boldsymbol{x}_k \boldsymbol{x}_k^H | \check{\boldsymbol{s}}_k\}^{-1}$$

$$= \Big(\boldsymbol{A}_k \boldsymbol{R}_\theta \boldsymbol{A}_k^H + \sigma^2 \boldsymbol{I}_{k+1} \Big)^{-1} \tag{38}$$

$$= \sigma^{-2} \Big(\boldsymbol{I}_{k+1} - \boldsymbol{A}_k \boldsymbol{\Sigma}_k^{-1} \boldsymbol{A}_k^H \Big), \tag{39}$$

where (39) follows from the matrix inversion lemma. Plugging (39) into (37), and then applying (33), we find

$$\hat{\boldsymbol{\theta}}_k = \sigma^{-2} \boldsymbol{R}_\theta \Big(\boldsymbol{\Sigma}_k \boldsymbol{\Sigma}_k^{-1} \boldsymbol{A}_k^H - \boldsymbol{A}_k^H \boldsymbol{A}_k \boldsymbol{\Sigma}_k^{-1} \boldsymbol{A}_k^H \Big) \boldsymbol{x}_k \tag{40}$$

$$= \boldsymbol{\Sigma}_k^{-1} \boldsymbol{A}_k^H \boldsymbol{x}_k. \tag{41}$$

Comparing (41) to (35), it is clear that the noncoherent search (35) involves implicit estimation of $\hat{\boldsymbol{\theta}}_k$. This observation will help in deriving a fast search algorithm.

3.2 Fast Metric Update

Since the sequence detector (35) must evaluate the metric

$$\mu(\check{\boldsymbol{s}}_k) = \boldsymbol{x}_k^H \boldsymbol{A}_k \boldsymbol{\Sigma}_k^{-1} \boldsymbol{A}_k^H \boldsymbol{x}_k \tag{42}$$

at least once per k, a fast algorithm to compute $\boldsymbol{\Sigma}_k^{-1}$ from $\boldsymbol{\Sigma}_{k-1}^{-1}$ is clearly of interest. Due to the rank-one update $\boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}_{k-1} + \boldsymbol{a}_k \boldsymbol{a}_k^H$, the matrix inversion lemma yields

$$\boldsymbol{\Sigma}_k^{-1} = \boldsymbol{\Sigma}_{k-1}^{-1} - \eta_k \boldsymbol{d}_k \boldsymbol{d}_k^H \tag{43}$$

with

$$\boldsymbol{d}_k := \boldsymbol{\Sigma}_{k-1}^{-1} \boldsymbol{a}_k \tag{44}$$

$$\eta_k := (1 + \boldsymbol{a}_k^H \boldsymbol{d}_k)^{-1}. \tag{45}$$

The partitions $\boldsymbol{x}_k = [\boldsymbol{x}_{k-1}^T, x_k]^T$ and $\boldsymbol{A}_k = [\boldsymbol{A}_{k-1}^H, \boldsymbol{a}_k]^H$ then allow the recursive metric update

$$\mu(\check{\boldsymbol{s}}_k) = (\boldsymbol{a}_k x_k + \boldsymbol{A}_{k-1}^H \boldsymbol{x}_{k-1})^H \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{a}_k x_k + \boldsymbol{A}_{k-1}^H \boldsymbol{x}_{k-1})$$

$$= \mu(\check{\boldsymbol{s}}_{k-1}) + x_k^* \boldsymbol{a}_k^H \boldsymbol{\Sigma}_k^{-1} \boldsymbol{a}_k x_k - \eta_k \hat{\boldsymbol{\theta}}_{k-1}^H \boldsymbol{a}_k \boldsymbol{a}_k^H \hat{\boldsymbol{\theta}}_{k-1} + 2\Re\{ x_k^* \boldsymbol{a}_k^H \hat{\boldsymbol{\theta}}_{k-1} - \eta_k x_k^* \boldsymbol{a}_k^H \boldsymbol{d}_k \boldsymbol{a}_k^H \hat{\boldsymbol{\theta}}_{k-1} \} \tag{46}$$

$$= \mu(\check{\boldsymbol{s}}_{k-1}) + \boldsymbol{u}_k^H \boldsymbol{M}_k \boldsymbol{u}_k, \tag{47}$$

with

$$\boldsymbol{u}_k := \begin{bmatrix} x_k \boldsymbol{a}_k \\ \hat{\boldsymbol{\theta}}_{k-1} \end{bmatrix} \tag{48}$$

$$\boldsymbol{M}_k := \begin{bmatrix} \boldsymbol{\Sigma}_k^{-1} & \boldsymbol{I}_{N_a N_b} - \eta_k \boldsymbol{d}_k \boldsymbol{a}_k^H \\ \boldsymbol{I}_{N_a N_b} - \eta_k \boldsymbol{a}_k \boldsymbol{d}_k^H & -\eta_k \boldsymbol{a}_k \boldsymbol{a}_k^H \end{bmatrix}, \tag{49}$$

where $N_b := 2D + 1$. Notice that $\hat{\boldsymbol{\theta}}_k$ can also be calculated recursively. Plugging (43) into (41), we get

$$\hat{\boldsymbol{\theta}}_k = (\boldsymbol{I}_{N_a N_b} - \eta_k \boldsymbol{d}_k \boldsymbol{a}_k^H) \hat{\boldsymbol{\theta}}_{k-1} + (1 - \eta_k \boldsymbol{d}_k^H \boldsymbol{a}_k) x_k \boldsymbol{d}_k. \tag{50}$$

Thus, initializing with

$$\big[\mu(\check{\boldsymbol{s}}_k),\ \boldsymbol{\Sigma}_k^{-1},\ \hat{\boldsymbol{\theta}}_k\big]_{k=-1} = \big[0,\ \sigma^{-2}\boldsymbol{R}_\theta,\ \boldsymbol{0}\big], \tag{51}$$

we can efficiently compute $\{\mu(\check{\boldsymbol{s}}_k), \boldsymbol{\Sigma}_k^{-1}, \hat{\boldsymbol{\theta}}_k\}$ using the inputs $\{x_k, \boldsymbol{a}_k\}$ and the previously calculated quantities $\{\mu(\check{\boldsymbol{s}}_{k-1}), \boldsymbol{\Sigma}_{k-1}^{-1}, \hat{\boldsymbol{\theta}}_{k-1}\}$ at each time $k \in \{0, \ldots, N-1\}$. It is straightforward to show that this update can be performed using $2(2D+1)^2 N_a^2 + 9(2D+1)N_a + 8$ multiplications.

3.3 Suboptimal Tree Search

Because exact maximization of (35) is computationally challenging, we propose to perform approximate maximization of (35) via tree search. While many options exist, we choose breadth-first search via the M-algorithm [1] since it offers a reasonably good performance/complexity tradeoff while keeping its complexity invariant of SNR and channel realization (unlike, e.g., sphere decoders [9]). We now briefly summarize the M-algorithm assuming that the symbols $\{\check{s}_k\}_{k=0}^{N-1}$ are all unknown; a modification to handle pilot/guard symbols will be described in Section 3.4.

At the k^{th} detection stage, the M-algorithm has a record of the M "best" surviving length-k partial paths, where M is a design parameter. The M-algorithm then computes the metric (42) for each length-$(k+1)$ 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. From this description, it can be seen that $M|\mathcal{S}|$ metrics need to be computed at each stage k. Thus, using the fast algorithm of Section 3.2, the total number of multiplications required to compute $\hat{\boldsymbol{s}}_{N-1}$ is only $\approx 2NM|\mathcal{S}|(2D+1)^2 N_a^2$.

3.4 Incorporating Pilot Symbols

With the previously described noncoherent sequence detector, a single pilot symbol is sufficient to resolve gain/phase ambiguities (e.g., [7]). But the judicious use of several pilots can dramatically improve the performance/complexity tradeoff of suboptimal tree-search. For example, if $(2D+1)N_a$ pilot symbols are placed at the beginning of the sequence $\boldsymbol{s}$, they can be used to compute a reasonable initial estimate of $\boldsymbol{\theta}$ before the M-algorithm is forced to discard partial paths, allowing high performance with even small M. In fact, the same approach can be taken with an arbitrarily placed block of contiguous pilots after cyclically shifting both $\boldsymbol{s}$ and $\boldsymbol{x}$ so that the pilot block appears at the beginning of the data sequence. The possibility of cyclic shifts can be recognized from (5) and the fact that $H_{d,k}^{(i,\ell)}$ is N-cyclic in the indices d and k.

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

4. TRACKING SPARSENESS

So far, we have assumed that the "active-tap" locations $\mathcal{L}^{(i)}$ are known. Since, in practice, these locations must be learned/tracked, Section 4.1 suggests an efficient means of doing so. We have also assumed that the inactive taps are zero-valued. Since, in practice, they will be small but non-zero, Section 4.2 suggests a simple means of compensation.

4.1 Active-Tap Identification

To estimate the locations of active taps, we first compute a pilot-aided MMSE estimate of the BEM coefficient vector $\breve{\boldsymbol{\theta}}^{(i)} := [\breve{\boldsymbol{\theta}}_{-D}^{(i)H}, \ldots, \breve{\boldsymbol{\theta}}_{D}^{(i)H}]$ for each symbol index i, as detailed below. Next, we estimate each component of the delay-power profile (DPP) $\{\sigma_l^2\}_{l=0}^{N_h-1}$ by summing the energies of the $2D+1$ Doppler coefficients. In particular, the l^{th} DPP element during the i^{th} MCM symbol interval is estimated as $\hat{\sigma_l^2}^{(i)} = \sum_{d=-D}^{D} \left|[\breve{\boldsymbol{\theta}}_d^{(i)}]_l\right|^2$. Note that this approach permits accurate DPP estimation even when the channel gains $\{h_{n,l}^{(i)}\}$ vary significantly over the MCM symbol interval. Once the DPP for a given symbol index has been estimated, there are several ways to assign N_a "active" channel taps from it. If a fixed receiver complexity was important, then N_a could be fixed and $\mathcal{L}^{(i)}$ could be set as the largest N_a taps in the DPP. If, instead, a fixed performance was preferred, then all taps in the DPP above a particular threshold could be assigned to $\mathcal{L}^{(i)}$. Note that the orthogonality of the basis $\boldsymbol{F}$ in (14) makes possible the simple assignment of $\mathcal{L}^{(i)}$ from the DPP; a matching-pursuit approach [14] is not needed.

For pilot-aided BEM-coefficient estimation, we chose the pilot pattern illustrated in Fig. 1, which repeats every P MCM symbols. There, each MCM symbol has a contiguous block of $N_p = N/P$ subcarriers dedicated to pilots and separated from the data by D zero-valued guard subcarriers on either side. The location of the pilot block is rotated so that every subcarrier is used as a pilot once every P MCM symbols. The role of the guard interval is twofold. First, it facilitates tree-search by making the ICI channel (24) "causal," as described in Section 2.3. Second, it ensures that N_p elements in $\boldsymbol{x}^{(i)}$ contain contributions from pilots and not from (unknown) data. In this case, the MMSE estimator of $\boldsymbol{\theta}^{(i)}$ from the "pilot-only" observations can be implemented with a fixed matrix multiplication. In contrast, decision-directed MMSE estimation of $\boldsymbol{\theta}^{(i)}$ would require the inversion of data-dependent matrices, which would be much more expensive. Recall that the role of the pilots are also twofold. First, they improve the performance-complexity tradeoff of suboptimal tree-search, as discussed in Section 3.4. Second, they facilitate tracking of active taps, which is the subject of this section. Though not investigated here, pilots could also be used for bulk adjustment of timing clock offsets and carrier frequency offsets (e.g., by keeping the delay and Doppler spreads within the window of interest). Small time/frequency offsets are subsumed by the doubly dispersive channel response, obviating the need for adjustement.

We now detail the MMSE estimation step. In our approach, the pilot-only observations from $\{\boldsymbol{x}^{(i)}, \ldots, \boldsymbol{x}^{(i+P-1)}\}$ are used to estimate the coefficients $\{\breve{\boldsymbol{\theta}}^{(i+1)}, \ldots, \breve{\boldsymbol{\theta}}^{(i+P-2)}\}$. Notice that estimates of $\{\breve{\boldsymbol{\theta}}^{(i)}, \breve{\boldsymbol{\theta}}^{(i+P-1)}\}$ are not computed from these observations because it is difficult to do so accurately. Instead, an estimate of $\breve{\boldsymbol{\theta}}^{(i)}$ will have been generated using the group of P observation vectors shifted back in time by $P-2$ symbol indices, i.e., $\{\boldsymbol{x}^{(i-P+2)}, \ldots, \boldsymbol{x}^{(i+1)}\}$ while an estimate of $\breve{\boldsymbol{\theta}}^{(i+P-1)}$ will be generated using the group of P observation vectors shifted forward in time by $P-2$ symbol indices, i.e., $\{\boldsymbol{x}^{(i+P-2)}, \ldots, \boldsymbol{x}^{(i+2P-3)}\}$. Notice that, in each estimation step, a total of $P \cdot \frac{N}{P} = N$ scalar observations are used to estimate $(P-2)(2D+1)N_h$ scalar BEM coefficients. Since MMSE estimation boils down to a $(P-2)(2D+1)N_h \times N$ matrix multiplication, the cost of channel estimation is only $(2D+1)N_hN$ multiplications per MCM symbol. Note that this is smaller than the cost of noncoherent decoding by the factor $2N_a^2(2D+1)/N_h$.

The MMSE estimation matrix can be designed as follows. Say that we collect the pilot-only observations from $\{\boldsymbol{x}^{(i)}, \ldots, \boldsymbol{x}^{(i+P-1)}\}$ into the vector $\underline{\boldsymbol{x}}^{(i)}$, and that we collect all impulse-response coefficients $\{h_{n,l}\}$ and noise samples $\{v_n\}$ affecting these observations into the vectors $\underline{\boldsymbol{h}}^{(i)}$ and $\underline{\boldsymbol{v}}^{(i)}$, respectively. Then, as seen from (5)-(7), we can write $\underline{\boldsymbol{x}}^{(i)} = \boldsymbol{C}\underline{\boldsymbol{h}}^{(i)} + \boldsymbol{D}\underline{\boldsymbol{v}}^{(i)}$, where $\boldsymbol{C}$ contains pilots and MCM pulse coefficients and where $\boldsymbol{D}$ contains MCM pulse coefficients, Likewise, we stack the BEM coefficient vectors $\{\breve{\boldsymbol{\theta}}^{(i+1)}, \ldots, \breve{\boldsymbol{\theta}}^{(i+P-2)}\}$ into the vector $\underline{\boldsymbol{\theta}}^{(i)}$ and see from (6), (13), and (14) that it is possible to write $\underline{\boldsymbol{\theta}}^{(i)} = \boldsymbol{G}\underline{\boldsymbol{h}}^{(i)}$ for some matrix $\boldsymbol{G}$. The MMSE estimate of $\underline{\boldsymbol{\theta}}^{(i)}$ from $\underline{\boldsymbol{x}}^{(i)}$ is then $\hat{\underline{\boldsymbol{\theta}}}^{(i)} = \boldsymbol{G}\boldsymbol{R}_h\boldsymbol{C}^H\left(\boldsymbol{C}\boldsymbol{R}_h\boldsymbol{C}^H + \sigma^2\boldsymbol{D}\boldsymbol{D}^H\right)^{-1}\underline{\boldsymbol{x}}^{(i)}$, where the channel statistics $\boldsymbol{R}_h := \mathrm{E}\{\underline{\boldsymbol{h}}^{(i)}\underline{\boldsymbol{h}}^{(i)H}\}$ are given by the WSSUS model (29) under the a-priori uniform DPP: $\sigma_l^2 = N_h^{-1}$ for $l \in \{0, \ldots, N_h-1\}$. We note that this pilot-aided MMSE channel estimation procedure is the same as that used in [23] except that the pilot pattern is different.

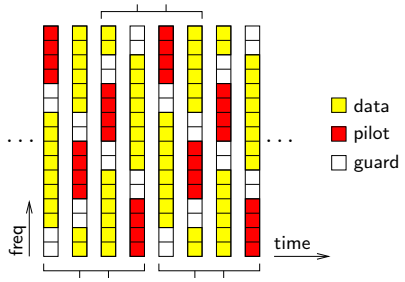

Figure 1: Illustration of multicarrier pilot pattern with $N = 16$, $P = 4$, $N_p = 2$ and $D = 2$.

4.2 Residual-Tap Compensation

Since non-active taps are not considered by the sparse BEM, they effectively add to the background noise. Because the total power of these residual (i.e., non-active) taps varies with symbol index i, it too should be tracked by the detection algorithm and lumped in with the noise power "σ^2," which would then also vary with i. Doing so is relatively easy. For example, after the noncoherent sequence detection algorithm has estimated $\boldsymbol{A}^{(i)}$ and $\boldsymbol{\theta}^{(i)}$, the residual $\hat{\boldsymbol{w}}^{(i)} := \boldsymbol{x}^{(i)} - \hat{\boldsymbol{A}}^{(i)}\hat{\boldsymbol{\theta}}^{(i)}$ can be computed and its energy calculated to retrieve an estimate of σ^2. This estimate could then be employed when decoding the $(i+1)^{th}$ MCM symbol.

5. NUMERICAL RESULTS

We used two types of channel for our simulations: a "perfectly sparse" channel and "sparse" channel. The "perfectly sparse" channel had $N_s < N_h$ nonzero impulse response coefficients, each of which varied randomly according to the Jakes autocorrelation $\rho_n = J_0(2\pi f_{\mathrm{D}} T_c n)$, where $J_0(\cdot)$ denotes the 0^{th}-order Bessel function of the first kind, and with $f_{\mathrm{D}} T_c = 0.0025$. Here, the nonzero impulse response trajectories were uncorrelated but had the same energy, N_s^{-1}. In addition, the delays of each trajectory were varied over time as described below. To generate the more realistic "sparse" channel model, we started with the "perfectly sparse" model but then leaked the energy of each active tap into four neighboring taps. We accomplished this by convolving the impulse response, at each time, with the truncated sinc sequence $[-0.0721, 0.0739, 0.9893, 0.0739, -0.0721]$. As a result, 2% of the dominant coefficient energy was leaked to residual coefficients.

For both channel types, the discrete delays of the active taps were varied in time as follows. We used $N_s = 4$ principal impulse response coefficients. The first and second had delays that were fixed at 2 and 5 chips, the delay of the third switched between 7 and 8 chips every 225 symbol intervals, and the delay of the fourth changed from 15 to 16 to 17 chips, and then back, stopping at each delay for 20 symbol intervals. Thus, the delay spread of the perfectly sparse channel was $N_h = 16$ and that of the sparse channel was $N_h = 20$. Note that this delay/Doppler-spread product equaled $f_{\mathrm{D}} T_h = f_{\mathrm{D}} T_c N_h = 0.05$, which matches that of the surf-zone channel from [20].

In all experiments, we used MCM with $N = 64$ QPSK subcarriers. (Coding and larger constellations will be examined in future work.) For the MCM pulses, we used the "transmitter optimized max-SINR (TOMS)" design from [4], which specifies a smooth modulation pulse of length $N_\alpha = 1.5N$ and a rectangular demodulation pulse of length $N_\beta = N$. With the ICI radius $D = \lceil f_{\mathrm{D}} T c N \rceil = 1$, the TOMS pulses are known for excellent suppression of ISI and out-of-band ICI without introducing noise correlation or wasting bandwidth through the use of a prefix. We used random constant modulus pilot sequences with pilot period $P = 4$, implying the use of $N_p = N/P = 16$ pilots per MCM symbol and a total spectral efficiency of $\frac{N-P-2D}{N} = 0.72$ scalar-symbols/sec/Hz. Since $N_p > (2D+1)N_s$, the noncoherent M-algorithm was able to generate a robust estimate of $\boldsymbol{\theta}^{(i)}$ before discarding partial paths, allowing us to choose $M = 8$. Finally, by assuming $N_a = 8$ active delay taps during reception, we guaranteed fixed receiver complexity.

We compared the BER of the proposed noncoherent M-algorithm to a genie-aided reference: coherent MLSD based on an MMSE estimate of $\boldsymbol{\theta}^{(i)}$ computed from 100% (genie-provided) training symbols. This genie-aided MLSD upper bounds the performance of any noncoherent sequence detection algorithm, but gives a tighter bound than does coherent MLSD under a *perfectly* known channel. Fig. 2 shows BER versus SNR for the "perfectly sparse" channel, where SNR equals σ^2 since the transmitted signal was unit-power and the channel was energy preserving. Note that the performance of the noncoherent M-algorithm is very close to the optimal genie-aided MLSD; less than 2 dB SNR loss (at 10^{-4} BER) can be observed. It can also be seen that detection under estimated $\mathcal{L}^{(i)}$ performed nearly as well as detection under perfectly known $\mathcal{L}^{(i)}$. The BER slopes suggest that a diversity of order-2 is extracted from the ICI; coding and interleaving could be used to extract delay diversity as well, as demonstrated by [4].

Fig. 3 shows BER performance for the "sparse" channel. With $N_a = 8$, some residual channel taps are left unmodeled, and so even the best selection of $\mathcal{L}^{(i)}$ loses about 0.7% of the channel energy. These unmodeled residual taps lead to an unavoidable BER error floor, even for the genie-aided MLSD. However, this error floor could be suppressed by increasing N_a (at the expense of complexity). Again, the proposed noncoherent M-algorithm performs about 2 dB worse than the genie-aided detector before both schemes succumb to the error floor. The detectors suffer only a bit when $\mathcal{L}^{(i)}$ is estimated from pilots.

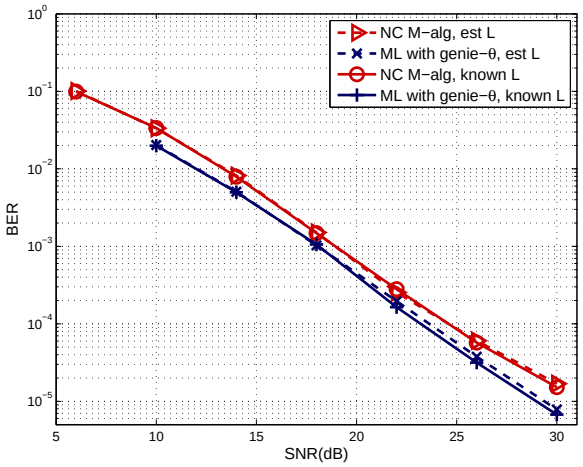

Figure 2: BER versus SNR for a "perfectly sparse" channel.

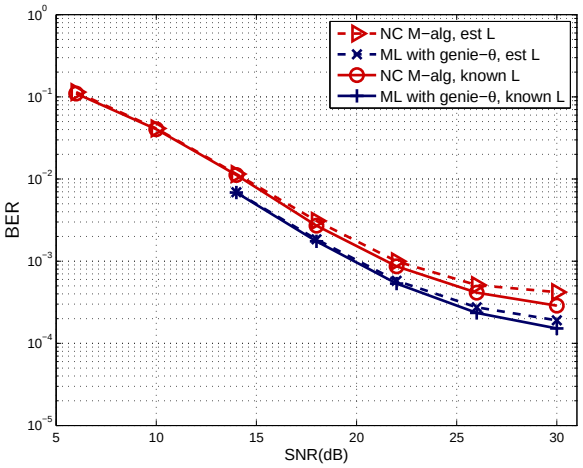

Figure 3: BER versus SNR for a "sparse" channel.

6. CONCLUSION

In this paper we presented a novel multicarrier strategy for communication over UACs with large delay-spread/Doppler-spread products. A multicarrier scheme was chosen to trans-

form an ISI span with tens or hundreds of taps to an ICI span with no more than a few taps. In doing so, a careful choice of multicarrier pulses eliminated the need for bandwidth-consuming time-domain guards. A near-ML symbol detection metric was then derived assuming knowledge of the ICI statistics, but not the ICI realizations, and the M-algorithm was employed to search for the metric-minimizing symbol sequence. Because the metric has a fast recursive update and simplifies under a (known) sparse delay-power profile, the search requires only about $2M|\mathcal{S}|(2D+1)^2N_a^2$ multiplications per scalar symbol. We found that small values of the search parameter M could be tolerated when several subcarriers were dedicated as pilots, even though, strictly speaking, a single pilot subcarrier suffices. Finally, a pilot-aided means of tracking variations in the power-delay profile was suggested. Simulations on highly spread channels (e.g., $f_{\mathrm{D}}T_h = 0.05$) showed that the performance of the proposed algorithm was less than 2 dB away from coherent ML detection using genie-estimated ICI. To prevent an error floor, however, the algorithm must ensure that the power of residual taps remains below that of the noise.

In future work, we plan to incorporate channel coding and the use of multiple transducers, and to test our approach experimentally in highly-spread underwater channels.

7. REFERENCES

[1] J. B. Anderson and S. Mohan. Sequential decoding algorithms: A survey and cost analysis. *IEEE Trans. Commun.*, 32:169–172, 1984.

[2] D. Brady and J. C. Preisig. Underwater acoustic communications. In H. V. Poor and G. W. Wornell, editors, *Wireless Communications: Signal Processing Perspectives*, chapter 8, pages 330–379. Prentice-Hall, 1998.

[3] H. Chen, R. Perry, and K. Buckley. On MLSE algorithms for unknown fast time-varying channels. *IEEE Trans. Commun.*, 51(5):730–734, May 2003.

[4] S. Das and P. Schniter. Max-SINR ISI/ICI-shaping multi-carrier communication over the doubly dispersive channel. *IEEE Trans. Signal Processing*, 55(12):5782–5795, Dec. 2007.

[5] T. H. Eggen, A. B. Baggeroer, and J. C. Preisig. Communication over Doppler spread channels—Part I: Channel and receiver presentation. *IEEE J. Oceanic Eng.*, 25(1):62–71, Jan. 2000.

[6] T. H. Eggen, A. B. Baggeroer, and J. C. Preisig. Communication over Doppler spread channels—Part II: Receiver characterization and practical results. *IEEE J. Oceanic Eng.*, 25(1):612–621, Oct. 2001.

[7] B. D. Hart. Maximum likelihood sequence detection using a pilot tone. *IEEE Trans. Veh. Tech.*, 49(2):550–560, Mar. 2000.

[8] S.-J. Hwang and P. Schniter. Near-optimal noncoherent sequence detection for doubly dispersive channels. In *Proc. Asilomar Conf. Signals, Systems and Computers*, pages 134–138, Nov. 2006.

[9] J. Jalden and B. Ottersten. On the complexity of sphere decoding in digital communications. *IEEE Trans. Signal Processing*, 53(4):1474–1484, 2005.

[10] A. P. Kannu and P. Schniter. On the spectral efficiency of noncoherent doubly selective channels. In *Proc. Allerton Conf. Commun., Control, and Computing*, Oct. 2006.

[11] W. Kozek and A. F. Molisch. Nonorthogonal pulseshapes for multicarrier communications in doubly dispersive channels. *IEEE J. Select. Areas In Commun.*, 16(8):1579–1589, Oct. 1998.

[12] B. Le Floch, M. Alard, and C. Berrou. Coded orthogonal frequency division multiplex. *Proc. IEEE*, 83(6):982–996, June 1995.

[13] B. Li, S. Zhou, M. Stojanovic, and L. Freitag. Pilot-tone based ZP-OFDM demodulation for an underwater acoustic channel. In *Proc. IEEE OCEANS*, Sept. 2006.

[14] W. Li and J. C. Preisig. Estimation and equalization of rapidly varying sparse acoustic communication channels. In *Proc. IEEE OCEANS*, Sept. 2006.

[15] Y. Li, C. N. Georghiades, and G. Huang. Iterative maximum-likelihood sequence estimation for space-time coded systems. *IEEE Trans. Commun.*, 49(6):948–951, June 2001.

[16] K. Liu, T. Kadous, and A. M. Sayeed. Orthogonal time-frequency signaling over doubly dispersive channels. *IEEE Trans. Inform. Theory*, 50(11):2583–2603, Nov. 2004.

[17] A. Mammela and D. P. Taylor. Bias terms in the optimal quadratic receiver. *IEEE Commun. Letters*, 2(2):57–58, Feb. 1998.

[18] A. K. Morozov and J. C. Preisig. Underwater acoustic communications with multi-carrier modulation. In *Proc. IEEE OCEANS*, Sept. 2006.

[19] J. C. Preisig. Performance analysis of adaptive equalization for coherent acoustic communications in the time-varying ocean environment. *J. Acoust. Soc. Am.*, 118:263–278, July 2005.

[20] J. C. Preisig and G. Deane. Surface wave focusing and acoustic communications in the surf zone. *J. Acoust. Soc. Am.*, 116:2067–2080, Oct. 2004.

[21] R. Raheli, A. Polydoros, and C. K. Tzou. Per-survivor processing: A general approach to MLSE in uncertain environments. *IEEE Trans. Commun.*, 43(2/3/4):354–364, Feb./Mar./Apr. 1995.

[22] P. Schniter. A new approach to multicarrier pulse design for doubly dispersive channels. In *Proc. Allerton Conf. Commun., Control, and Computing*, Oct. 2003.

[23] P. Schniter. On doubly dispersive channel estimation for pilot-aided pulse-shaped multi-carrier modulation. In *Proc. Conf. Inform. Science and Systems*, Mar. 2006.

[24] M. Stojanovic. Low complexity OFDM detector for underwater acoustic channels. In *Proc. IEEE OCEANS*, Sept. 2006.

[25] M. Stojanovic, J. A. Catipovic, and J. G. Proakis. Phase-coherent digital communications for underwater acoustic channels. *IEEE J. Oceanic Eng.*, 19(1):100–111, Jan. 1994.

[26] T. Strohmer and S. Beaver. Optimal OFDM design for time-frequency dispersive channels. *IEEE Trans. Commun.*, 51(7):1111–1122, July 2003.

[27] L. Tong, B. M. Sadler, and M. Dong. Pilot-assisted wireless transmissions. *IEEE Signal Processing Mag.*, 21(6):12–25, Nov. 2004.