Iterative Sparse Channel Estimation and Data Detection for Underwater Acoustic Communications Using Partial Interval Demodulation

Arunkumar K.P. and Chandra R. Murthy, *Senior Member, IEEE*

***Abstract*—We present an iterative scheme for sparse-channel recovery and data detection in cyclic-prefix orthogonal frequency division multiplex (CP-OFDM) communication over doubly-spread underwater acoustic channels. We consider the sequence of observations from partial interval demodulators (PIDs), and cast them into an observation model amenable for sparse channel recovery. We propose a two-stage iterative algorithm for channel estimation and data detection. In the first stage, we recover the channel from pilot only observations and estimate the unknown data symbols from the post-combined PID outputs. In the second stage, we use the data symbols estimated in the first stage to reconstruct the dictionary matrix corresponding to a full interval demodulator, re-estimate the channel using the entire observations including the data subcarriers, and use it to detect the unknown data symbols from the PID outputs. We also propose a computationally attractive algorithm for sparse signal recovery, based on the minimum variance principle, that may be of independent interest. Theoretically, we show that the PID outputs help in tracking the time-varying channel better by providing additional measurements to estimate the ICI due to Doppler spread compared to full interval demodulation. Also, we derive the Cramér-Rao lower bound on the mean squared error in channel estimation, and empirically show that the proposed two-stage algorithm meets the bound at high SNR. Numerical studies on simulated channels and publicly available experimental channel data in WATERMARK show that the proposed algorithm considerably improves data detection performance, in terms of bit error rate, over that from a traditional full length demodulator output, in highly Doppler distorted scenarios.**

***Index Terms*—Underwater acoustic communications, sparse channel recovery, partial interval demodulator.**

I. INTRODUCTION

Orthogonal frequency division multiplexing (OFDM) achieves high data rates, even in doubly-distorted channels with large delay spreads, due to its better resilience to inter symbol interference (ISI). However, the orthogonality between subcarriers is lost due to frequency offsets and the Doppler distortion introduced by the channel, leading to inter carrier interference (ICI). The Doppler distortion is particularly severe in multipath channels prevalent in underwater acoustic (UWA) communications, because different paths can potentially have different Doppler shifts, leading to Doppler spread [2]. This makes the problem of channel estimation and data detection particularly challenging. On the other hand, the UWA channel is known to be sparse in the lag-Doppler domain, because there are typically only a few significant multipath components in the channel. Therefore, it is pertinent to develop techniques that exploit the underlying structure in the channel to jointly estimate the channel and data symbols in the presence of severe ICI, which is the goal of this paper.

Sparsity based channel recovery techniques are well known to produce significantly improved channel estimates, and hence lead to better symbol detection on data subcarriers [3]–[6]. UWA channels typically have large delay and Doppler spreads but have only a few dominant paths [6]. This has been used in [3] to characterize the channel impulse response by using a path-based model, thereby facilitating channel estimation using sparse signal recovery techniques.

In [7], an equalization scheme was proposed where an iterative receiver progressively increases the so-called ICI span parameter to improve the channel estimate in severe ICI conditions. Both [3] and [7] require additional pilots to estimate the channel in high Doppler spread environments. An alternative approach to ICI mitigation using several partial interval demodulators (PID), instead of a full interval demodulator (FID), was proposed in [8]. The authors also develop a recursive algorithm to compute the weights for combining the output of PIDs so as to make the post-combined channel matrix close to diagonal. While [8] used non-overlapping rectangular windows over time for PID and applied coherent detection, [9] extended the decomposition to other forms of windowing and applied differentially coherent data detection. The authors propose a stochastic gradient algorithm to estimate the combiner weights. However, the inherent sparsity of the channel is not exploited in [8] and [9].

In this paper, we consider a cyclic-prefix OFDM (CP-OFDM) system [8], [10]. We propose a two-stage iterative approach for channel estimation and data detection. We exploit channel sparsity to estimate the path-dependent delay, Doppler and amplitude parameters of the channel from pilot-only observations of the PID outputs. An approximate dictionary, initially constructed using only the pilot symbols, is used to initiate the channel recovery. We refine the dictionary using estimates of the data symbols and iterate between channel estimation and data detection. The data symbols detected at the end of the first stage are then used to initialize a second stage that makes use of the entire observation vector, consisting of both pilot and data subcarriers, at the output of the FID. The second stage iteratively bootstraps the channel estimation using the detected data symbols to construct the dictionary matrix for the FID output, thereby reducing the channel estimation error, and ultimately leading to better data detection performance. Our specific contributions are as follows:

Arunkumar K. P. is with the Naval Physical Oceanographic Laboratory, Kochi 682 021, India. Chandra R. Murthy is with the Dept. of ECE, Indian Institute of Science, Bangalore 560 012, India (e-mails: arun.drdo.npol@gmail.com, cmurthy@iisc.ac.in).

A part of this paper has been published in [1].

This work was financially supported by a research grant from the Defence Research Development Organisation, Govt. of India.

1) We reformulate problem of estimating the doubly-spread channel from the PID outputs in a manner that is amenable to sparsity-based channel estimation.
2) We propose a two-stage iterative framework that recovers the channel and detects the data symbols. For sparse channel recovery, we present an improved low complexity algorithm based on the minimum variance principle, that may be of independent interest. The algorithm refines the initial estimate produced by the Orthogonal Matching Pursuit (OMP) algorithm, and is observed to converge within a single iteration. Its performance is better than OMP and is comparable to the computationally intensive Sparse Bayesian Learning (SBL) algorithm.
3) We analytically show that using the PID outputs increases the effective number of measurements compared to using only the FID output. Further, in the context of ℓ_1 based sparse signal recovery, we show that our scheme minimizes a joint cost function of the channel estimation and data detection error, and establish its convergence.
4) We derive a lower bound on the mean square error (MSE) in channel estimation, and numerically show that the bound corresponding to the PID outputs is strictly better than that corresponding to the FID outputs.
5) Through extensive numerical studies, using synthesized and measured channels, we demonstrate that the BER of the proposed scheme is considerably lower than the existing methods, in highly Doppler spread scenarios.

We develop the system model in Section II. In Section III, we describe the two-stage iterative algorithm for data detection and channel estimation. In Section IV, we show how our proposed algorithm improves the channel estimation accuracy. In the same section, we examine the Cramér Rao Bound (CRB) for channel estimators that make use of the observations at the output of the PID and contrast it with those that use only the FID output. We present the results of our numerical studies in Section V and conclude in Section VI.

II. System Model

A. Transmitted and Received Signal

We consider a CP-OFDM system as in [8]. Let T denote the OFDM symbol duration and T_g the guard interval (duration of the cyclic prefix). When using a carrier frequency f_c and K subcarriers, the kth subcarrier is at frequency

$$f_k = f_c + k/T,\ k = -K/2, \ldots, K/2-1. \tag{1}$$

The transmitted symbol at the kth subcarrier is denoted by $s[k]$. The disjoint sets of data subcarriers S_D, pilot subcarriers S_P, and null subcarriers[1] S_N partition the K available subcarriers. The transmitted signal is given by

$$\tilde{x}(t) = \frac{1}{\sqrt{T}}\text{Re}\left\{\left[\sum_{k\in S_D\cup S_P} s[k]e^{j2\pi\frac{k}{T}t}\right]e^{j2\pi f_c t}\right\}, \quad t\in[-T_g, T]. \tag{2}$$

[1]The null subcarriers are not crucial for the performance, but they are useful for estimating the ICI, since the doubly-spread channel is no longer diagonal in the frequency domain. They can also be used to estimate the noise variance.

At the receiver, the signal is resampled by a factor $\hat{a}$ corresponding to a coarse Doppler estimate, leading to a baseband received signal $\tilde{z}(t)$ given by [3], [8], [11]–[13]

$$\tilde{z}(t) = \sum_{p=1}^{N_p} A_p \tilde{x}\left((1+b_p)\,t - \tau_p\right) + \tilde{n}\left(\frac{t}{1+\hat{a}}\right). \tag{3}$$

where A_p and τ_p are the amplitude and delay, respectively, of the p^{th} path, N_p is the number of significant paths in the channel, and $\tilde{n}(t)$ is the additive noise. The above assumes that the path amplitudes are constant within the OFDM symbol duration T, and that the time variation of the path delays due to Doppler rate a_p can be approximated as $\tau_p - a_p t$, as in [3]. The resampled signal is equivalent to a received signal obtained through a channel with Doppler rate $b_p = \left(\frac{a_p-\hat{a}}{1+\hat{a}}\right)$. Note that, in this work, the effect of Doppler spread is modeled as a corresponding path dependent time compression/dilation.

We use the above $\tilde{z}(t)$ to formulate the input-output model and the sparse channel recovery model pertaining to the PID output as follows. We divide the OFDM symbol interval $[0, T]$ into M consecutive partial intervals of duration T/M each. The output $z_k^{(m)}$ on the k^{th} subcarrier, upon performing demodulation for the m^{th} partial interval, $(m-1)\frac{T}{M} \le t \le m\frac{T}{M}, m = 1, 2, \cdots, M$, is given by

$$z_k^{(m)} = \sum_{p=1}^{N_p} A_p \sum_{l\in S_D\cup S_P} \varrho_{k,l}^{(p)}[m] e^{-j2\pi f_l \tau_p} s[l] + \eta_k^{(m)}, \tag{4}$$

where $\eta_k^{(m)}$ is the additive noise, and

$$\varrho_{k,l}^{(p)}[m] = \text{sinc}\left(\frac{\beta_{k,l}^{(p)}T}{M}\right) e^{j\pi\beta_{k,l}^{(p)}(2m-1)\frac{T}{M}}, \tag{5}$$

$$\beta_{k,l}^{(p)} = (l-k)\frac{1}{T} + b_p f_l. \tag{6}$$

The above equations combine the models in [3] and [8]. The channel model in the above is the same as in [3], which used only the FID outputs and did not consider PID. On the other hand, [8] used the PID outputs but did not parameterize the channel in the delay-Doppler plane as in [3].

B. Input-Output Data Model

By stacking the received symbols across all the subcarriers into $\mathbf{z}^{(m)} \in \mathbb{C}^K$, the data symbols into $\mathbf{s} \in \mathbb{C}^K$, and the noise into $\mathbf{v}^{(m)} \in \mathbb{C}^K$, we get the channel input-output equation as:

$$\mathbf{z}^{(m)} = \mathbf{H}^{(m)}\mathbf{s} + \mathbf{v}^{(m)}, \tag{7}$$

for $m = 1, \ldots, M$, where the channel matrix $\mathbf{H}^{(m)}$ can be expressed as $\mathbf{H}^{(m)} = \sum_{p=1}^{N_p} A_p \mathbf{\Lambda}_p^{(m)} \mathbf{\Gamma}_p$. Here, $\mathbf{\Lambda}_p^{(m)}$ is a $K \times K$ matrix with $(k,l)^{\text{th}}$ entry $[\mathbf{\Lambda}_p]_{k,l}^{(m)} = \varrho_{k,l}^{(p)}[m]$, and $\mathbf{\Gamma}_p$ is a $K \times K$ diagonal matrix with $(k,k)^{\text{th}}$ entry $[\mathbf{\Gamma}_p]_{k,k} = e^{-j2\pi f_k \tau_p}$.

The output from a FID is obtained by summing up all $\mathbf{z}^{(m)}, m = 1, 2, \cdots, M$: $\mathbf{z} = \sum_{m=1}^{M} \mathbf{z}^{(m)} = \mathbf{H}\mathbf{s} + \mathbf{v}$, where $\mathbf{H}$ is the channel matrix corresponding to the FID output given by $\mathbf{H} = \sum_{m=1}^{M} \mathbf{H}^{(m)}$, and $\mathbf{v} = \sum_{m=1}^{M} \mathbf{v}^{(m)}$.

If the additive noise in (3) is zero mean circularly symmetric white Gaussian distributed, then the noise in FID output is

also zero mean Gaussian with covariance $N_0\mathbf{I}$, where N_0 is the variance of each component of $\mathbf{v}$ and $\mathbf{I}$ is the $K \times K$ identity matrix. The noise in (7) is not white, but it is also zero mean Gaussian distributed with covariance given by:

$$E\left[\eta_k^{(m)}\eta_l^{(m)*}\right] = \frac{N_0}{M}e^{-\frac{j2\pi(k-l)(2m-1)}{2M}}\text{sinc}\left(\frac{k-l}{M}\right), \quad (8)$$

and $E\left[\eta_k^{(m_1)}\eta_l^{(m_2)*}\right] = 0$ for $m_1 \neq m_2$.

It is straightforward to see that when $b_p = 0$, the channel matrix $\mathbf{H}$, as seen by the FID, is diagonal. When b_p's are nonzero and high, due to heavy Doppler spread, the channel matrix is no longer diagonal and this results in the mixing of the symbols at the output corresponding to each subcarrier. Within a partial interval, $b_p t$ can be approximated by $b_p t_m$, where $t_m \triangleq (2m-1)\frac{T}{2M}$ is the mid-point of the m^{th} partial interval [8]. Then, we can write the channel matrices as

$$\mathbf{H}^{(m)} = \mathbf{J}^{(m)}\sum_{p=1}^{N_p} A_p\mathbf{\Gamma}_p^{(m)}, \quad (9)$$

where $\mathbf{\Gamma}_p^{(m)}$ is a diagonal matrix whose $(k,k)^{\text{th}}$ entry is given by $[\mathbf{\Gamma}_p]_{k,k}^{(m)} = e^{-j2\pi f_k(\tau_p - b_p t_m)}$, and $\mathbf{J}^{(m)}$ is a matrix with $(k,l)^{\text{th}}$ element $\mathbf{J}_{k,l}^{(m)} = \frac{1}{M}e^{-\frac{j2\pi(k-l)(2m-1)}{2M}}\text{sinc}\left(\frac{k-l}{M}\right)$, for $k,l = 1,2,\ldots,K$. Henceforth, we consider the data model for the PID outputs expressed by (7), where the channel matrix is given by (9) and the noise vector is zero-mean with a covariance matrix given by (8).

C. Sparse Channel Recovery Model

In the formulation above, the channel matrix $\mathbf{H}^{(m)}$ is defined by N_p triplets (A_p, b_p, τ_p). In this section, we cast the problem of estimating these N_p triplets as a sparse channel recovery problem, given the sequence of PID outputs $\mathbf{z}^{(m)}, m = 1,2,\ldots,M$. To this end, we construct a dictionary consisting of the signals parameterized by a representative selection of possible parameter values [3]. Since parameter values that are not part of the solution will have the corresponding coefficient as zero and a large number of parameter values are needed to construct an accurate dictionary, the vector of coefficients is sparse, thus making sparse vector recovery algorithms applicable. The representative values of (b_p, τ_p) are chosen as

$$\tau_p \in \left\{\frac{T}{\lambda K}, \frac{2T}{\lambda K}, \ldots, \frac{N_\tau T}{\lambda K}\right\}, \quad (10)$$

$$b_p \in \{-b_{\max}, -b_{\max} + \Delta b, \ldots, b_{\max}\}. \quad (11)$$

The time resolution for τ_p is chosen as a multiple, λ, of the sampling time T/K, with N_τ candidate delays such that $\frac{N_\tau T}{\lambda K}$ is larger than the maximum delay spread of the channel. For b_p, we consider $N_b = 2b_{\max}/(\Delta b)+1$ candidate Doppler rates. Defining the coefficient vector corresponding to all delays associated with Doppler scale b_i as $\mathbf{x}_A^{(i)} = \left[A_1^{(i)}, \ldots, A_{N_\tau}^{(i)}\right]^T$, the stacked coefficient vector corresponding to all candidate delays and Doppler rates is given by

$$\mathbf{x} = \left[\left(\mathbf{x}_A^{(1)}\right)^T, \ldots, \left(\mathbf{x}_A^{(N_b)}\right)^T\right]^T. \quad (12)$$

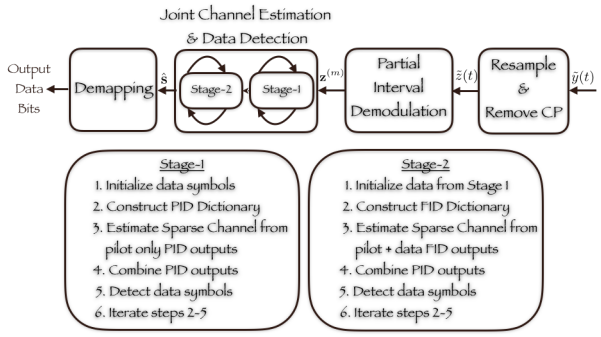

Figure 1. Block diagram of the receiver processing chain and the dual stage algorithm for iterative channel estimation and data detection.

Hence, (7) now takes the form:

$$\mathbf{z}^{(m)} = \mathbf{A}^{(m)}\mathbf{x} + \mathbf{v}^{(m)}, \quad (13)$$

$$\text{where } \mathbf{A}^{(m)} = \mathbf{J}^{(m)}\left[\mathbf{\Gamma}_1^{(m)}\mathbf{s}, \ldots, \mathbf{\Gamma}_N^{(m)}\mathbf{s}\right], \quad (14)$$

for $m = 1,\ldots,M$, with $N = N_\tau N_b$ representing the total number of grid points used in the delay-Doppler plane.

Although the channel vector $\mathbf{x}$ is sparse, the construction of the dictionary matrix $\mathbf{A}^{(m)}$ requires knowledge of the transmitted symbol vector $\mathbf{s}$, which is unknown at the receiver. In the next section, we propose a two-stage iterative algorithm to recover the channel vector $\mathbf{x}$ and the data vector $\mathbf{s}$.

III. Channel Estimation and Data Detection

Our proposed two stage algorithm works as follows. In stage 1, we use PID measurements from only the pilot subcarriers to estimate the channel, and subsequently use the estimated channel to detect the unknown data symbols. In stage 2, we use the FID outputs on both the data and pilot subcarriers, with the data symbols initialized using the outcome of stage 1, to further reduce the channel estimation error and improve the data detection performance. Figure 1 shows a block diagram for the proposed approach.

A. Stage 1

We start by constructing a vector $\hat{\mathbf{s}} \in \mathbb{C}^K$ by placing the known pilot symbol at the pilot subcarrier locations and zeros at the null subcarrier locations. Further, we initialize the unknown data symbols at subcarrier locations S_D to zero.[2] By defining $\mathbf{z}_{S_P}^{(m)} \in \mathbb{C}^{|S_P|}$ to be a sub-vector of $\mathbf{z}^{(m)} \in \mathbb{C}^K$ that collects the symbols corresponding to the pilot subcarrier locations S_P, from (13) we have,

$$\mathbf{z}_{S_P}^{(m)} = \hat{\mathbf{A}}_{S_P}^{(m)}\mathbf{x} + \mathbf{e}_{S_P}^{(m)}, \quad (15)$$

where

$$\hat{\mathbf{A}}_{S_P}^{(m)} = \mathbf{I}_{S_P}\mathbf{J}^{(m)}\left[\mathbf{\Gamma}_1^{(m)}\hat{\mathbf{s}}, \ldots, \mathbf{\Gamma}_N^{(m)}\hat{\mathbf{s}}\right] \in \mathbb{C}^{|S_P|\times N}, \quad (16)$$

$\mathbf{I}_{S_P} \in \mathbb{R}^{|S_P|\times K}$ is the submatrix of the $K \times K$ identity matrix consisting of its rows indexed by S_P, and $\mathbf{e}_{S_P}^{(m)} \triangleq$

[2]We find in our simulation studies that a random initialization of the data symbols also works equally well.

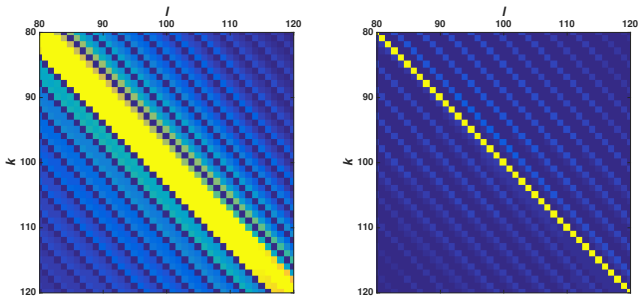

Figure 2. Left: Channel matrix seen at FID output. Right: Post-combined channel matrix at (combined) PID output. (k, l) are (row, column) indices of the channel matrix. Only indices from 80-120 are shown for clarity. Bright yellow indicates high magnitude while dark black indicates low magnitude. These matrices are obtained for the CP-OFDM system I in Table II, Section V, and residual Doppler spread in $[-b_{\max}, b_{\max}]$, where $b_{\max} = 5 \times 10^{-4}$.

$\Delta\mathbf{A}_{S_P}^{(m)}\mathbf{x} + \mathbf{v}_{S_P}^{(m)}$ is the effective noise that includes the dictionary estimation error given by:

$$\Delta\mathbf{A}_{S_P}^{(m)} = \mathbf{I}_{S_P}\mathbf{J}^{(m)}\left[\mathbf{\Gamma}_1^{(m)}(\mathbf{s}-\hat{\mathbf{s}}), \ldots, \mathbf{\Gamma}_N^{(m)}(\mathbf{s}-\hat{\mathbf{s}})\right]. \quad (17)$$

The error $\mathbf{s}-\hat{\mathbf{s}}$ affects $\mathbf{z}_{S_P}^{(m)}$ in as much as there is ICI captured by the matrix $\mathbf{J}^{(m)}$. Note that, the presence of null subcarriers around each pilot subcarrier mitigates the ICI.

To recover the sparse channel vector $\mathbf{x}$ in (15), we propose an improved algorithm that bootstraps from the channel estimate produced by OMP and refines this estimate. We call our proposed algorithm the Minimum Variance Recovery (MVR) algorithm, and provide its details in Section III-C.

Once the sparse channel vector $\mathbf{x}$ is recovered, the channel matrix $\hat{\mathbf{H}}^{(m)}$ for each PID output is constructed using (9). The estimates $\hat{\mathbf{H}}^{(m)}, m = 1, 2, \ldots, M$, are then weighted and combined so as to make the post-combined channel matrix close to diagonal, as follows:

$$\hat{\mathbf{H}}_{k,:} = \sum_{m=1}^{M} w_k^{(m)*}\hat{\mathbf{H}}_{k,:}^{(m)}, \quad (18)$$

where the weights $\mathbf{w}_k = \left[w_k^{(1)}, \cdots, w_k^{(M)}\right]^T \in \mathbb{C}^M$ are chosen for each subcarrier k to minimize the interference from other subcarriers as $\mathbf{w}_k = \hat{\mathbf{R}}_{\mathbf{z}_k}^{-1}\hat{\mathbf{h}}_{k,k}/\|\hat{\mathbf{R}}_{\mathbf{z}_k}^{-1}\hat{\mathbf{h}}_{k,k}\|_2$, with $\hat{\mathbf{h}}_{k,l} \triangleq \left[\hat{\mathbf{H}}_{k,l}^{(1)}, \cdots, \hat{\mathbf{H}}_{k,l}^{(M)}\right]^T$ and $\hat{\mathbf{R}}_{\mathbf{z}_k} \triangleq \sum_{l=1}^{K}\hat{\mathbf{h}}_{k,l}\hat{\mathbf{h}}_{k,l}^H + \frac{N_0}{M}\mathbf{I}_M$ [8]. The post-combined observation $\tilde{z}_k$ is obtained by weighting and adding the k^{th} entry, $z_k^{(m)}$, of $\mathbf{z}^{(m)}$:

$$\tilde{z}_k = \sum_{m=1}^{M} w_k^{(m)*} z_k^{(m)}. \quad (19)$$

Figure 2 contrasts an instance of the channel matrix obtained by simply summing up the PID outputs with that obtained by their weighted combination. The latter is clearly closer to diagonal than the former, which simplifies the subsequent data detection step, described next.

Since the weights in (18) are normalized to have unit 2-norm, it follows from (8) and (19) that the variance of the noise at the post-combiner output is $E\left\{|\tilde{\eta}_k|^2\right\} = N_0$. A minimum mean square error (MMSE) receiver is applied for data demodulation, as follows:

$$\hat{\mathbf{s}} = \text{dec}\left[\left(\hat{\mathbf{H}}^H\hat{\mathbf{H}} + N_0\mathbf{I}\right)^{-1}\hat{\mathbf{H}}^H\tilde{\mathbf{z}}\right] \quad (20)$$

where $\tilde{\mathbf{z}} \in \mathbb{C}^K$ is obtained by stacking $\tilde{z}_k, k = 1, \ldots, K$, and $\text{dec}(\cdot)$ is the hard-thresholding operation to the signal constellation. We find in our simulation studies that zeroing out all but diagonal entries of the post-combined channel matrix, $\hat{\mathbf{H}}$, has negligible effect on data detection accuracy. On the positive side, this approximation significantly reduces the computational complexity involved in the matrix inversion in (20). Using this approximation, the unknown data symbols are estimated as follows:

$$\hat{s}_k \approx \text{dec}\left[\frac{\hat{H}_{k,k}^*}{|\hat{H}_{k,k}|^2 + N_0}\tilde{z}_k\right], k \in S_D, \quad (21)$$

where $\hat{H}_{k,k}$ is the kth diagonal entry of $\hat{\mathbf{H}}$. Using $\hat{\mathbf{s}}$, we reconstruct the dictionary matrix $\hat{\mathbf{A}}_{S_P}^{(m)}$, for $m = 1, 2, \cdots, M$, and iterate through channel estimation and data symbol detection.

In stage 1, the channel vector is recovered using only the pilot subcarriers in the output of the PID, and the data symbols are detected from the post-combined demodulator output. The detection of data symbols helps in accurately estimating the ICI, which in turn helps in reducing $\|\Delta\mathbf{A}_{S_P}^{(m)}\mathbf{x}\|_2$ in (17), leading to better channel estimates. However, the observations on the data subcarriers are not used for channel estimation in stage 1. In stage 2, we make use of the measurements from both data and pilot subcarriers for channel estimation.

B. Stage 2

We start by constructing the dictionary matrix corresponding to the FID as $\mathbf{A} = \sum_{m=1}^{M}\mathbf{A}^{(m)}$, using the data symbols $\hat{\mathbf{s}}$ estimated at the end of stage 1. Then we proceed to re-estimate the channel vector from:

$$\mathbf{z} = \sum_{m=1}^{M}\mathbf{z}^{(m)} = \mathbf{A}\mathbf{x} + \mathbf{v}, \quad (22)$$

as in stage 1, but now using the full set of observations including data subcarriers at the output of the FID. From the estimated channel vector, we construct the post-combined channel matrix $\hat{\mathbf{H}}$, and the corresponding post-combined measurement $\tilde{\mathbf{z}}$, using (18) and (19) respectively, and apply the MMSE receiver in (21) for data demodulation using the PID outputs, as in stage 1. Further, we iterate between channel estimation and data detection until convergence or till a fixed number of iterations (N_{iter}) have elapsed.

C. Minimum Variance Recovery (MVR) Algorithm

The proposed MVR algorithm, given in Algorithm 1, is inspired by the minimum variance spectrum estimation principle [14] and works by refining the channel estimate produced by OMP. We construct an estimate of the measurement covariance matrix from a thresholded[3] version of the sparse channel

[3]The thresholding operation sets the coefficients whose magnitude is smaller than the threshold to zero. Empirically, we find that performance is insensitive to the value of the threshold, so its choice is not critical.

Algorithm 1 Minimum Variance Recovery

1: **function** MVR($\mathbf{A}, \mathbf{z}, \hat{\mathbf{x}}_0, \sigma$) ▷
$\mathbf{A} \in \mathbb{C}^{m\times n}$: dictionary matrix, $\mathbf{z} \in \mathbb{C}^m$: measurement vector, $\hat{\mathbf{x}}_0 \in \mathbb{C}^n$: initial estimate of the sparse channel vector (obtained from OMP), σ: noise standard deviation
2: Threshold: $\gamma = \sigma\sqrt{\frac{m}{n}}$
3: Initial support: $\mathbb{S} = \{i \in [1, 2 \dots n] : |\hat{\mathbf{x}}_{0,i}| > \gamma\}$
4: Initial solution: $\hat{\mathbf{x}}_{\mathbb{S}} = \mathbf{A}^{\dagger}_{:,\mathbb{S}}\mathbf{z}$, $\hat{\mathbf{x}}_{\mathbb{S}^c} = 0$
5: Cardinality: $k = |\mathbb{S}|$
6: Measurement covariance: $\mathbf{R} = \sum_{i\in\mathbb{S}} |\hat{x}_i|^2 \mathbf{a}_i\mathbf{a}_i^H + \delta I$
7: Weight vector: $\mathbf{w}_i = \frac{\mathbf{R}^{-1}\mathbf{a}_i}{\mathbf{a}_i^H\mathbf{R}^{-1}\mathbf{a}_i}$
8: Minimum variance signal: $\tilde{\mathbf{x}} = \mathbf{W}^H\mathbf{z}$
9: $\mathbb{S}$ = Set of indices of the largest k entries of $|\tilde{\mathbf{x}}|$
10: Update solution: $\hat{\mathbf{x}}_{\mathbb{S}} = \mathbf{A}^{\dagger}_{:,\mathbb{S}}\mathbf{z}$, $\hat{\mathbf{x}}_{\mathbb{S}^c} = 0$
11: Output: $\hat{\mathbf{x}}$
12: **end function**

estimate provided by OMP, and use it to compute an adaptive minimum variance weight vector. We use the weight vector to re-estimate each entry of the channel vector, and update the solution to contain the indices of the k largest entries of the minimum variance solution, where k is the cardinality of the thresholded version of the initial channel estimate. This form of support update helps to minimize the leakage of interference from other nonzero coefficients while identifying the location of a nonzero coefficient in the sparse channel vector. Finally, a refined solution vector is constructed using the updated support. We note that, in the context of synthetic aperture radar imaging, an algorithm similar in flavor called the iterative adaptive algorithm has been proposed in [15].

In Figure III-C, we show the phase transition curves of OMP, SBL and MVR on the m/n-k/n plane. At all operating points below the curve, the signal to reconstruction error ratio (SRR) exceeds 15 dB for at least 90% of the 1000 trial runs, where m is the number of measurements, k is the number of nonzero entries in the sparse vector and $n = 50$ is the ambient dimension of the sparse vector. The $m \times k$ measurement matrices are drawn independent and identically distributed (i.i.d.) from a standard Gaussian distribution with i.i.d. entries. The sparse vector is generated with k nonzeros at uniformly random locations and with i.i.d. entries, uniformly distributed on $[-2, -1]\cup[1, 2]$. The additive noise is zero mean Gaussian with a standard deviation chosen to achieve a signal to noise ratio (SNR) of 30 dB. The phase transition curves of SBL and MVR nearly coincide, and are superior to OMP. However, SBL is a computationally demanding algorithm and it takes a large number of iterations to converge. For the MVR algorithm, while steps from lines 6 to 10 in Algorithm 1 can be iterated multiple times, we find, in our simulations, that the channel estimation error reduces significantly in just one pass of the algorithm. Thus, MVR offers a computationally attractive alternative to SBL with comparable performance for our application. Empirically, we have observed a similar relative performance behavior at other SNRs also.

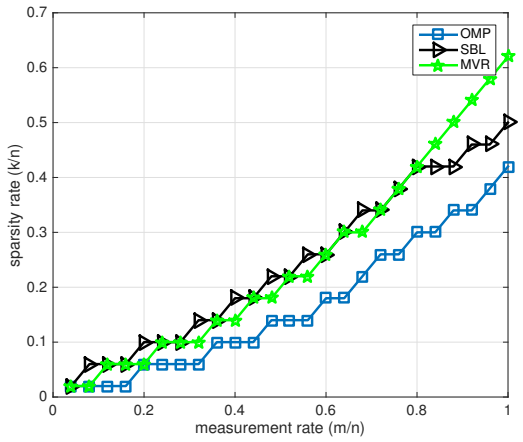

Figure 3. OMP, SBL & MVR phase transition curves at SNR = 30 dB.

Table I
SUMMARY OF COMPUTATIONAL COMPLEXITY.

Computation	Stage 1	Stage 2
Dictionary Matrix	$\mathcal{O}(M\|S_P\|KN)$	$\mathcal{O}(K^2N)$
OMP (per iteration)	$\mathcal{O}(M\|S_P\|N)$	$\mathcal{O}(KN)$
SBL (per iteration)	$\mathcal{O}(M^2\|S_P\|^2N)$	$\mathcal{O}(K^2N)$
MVR	$\mathcal{O}(M^2\|S_P\|^2N)$	$\mathcal{O}(K^2N)$
Channel Matrix	$\mathcal{O}(K^2\hat{N}_p)$	$\mathcal{O}(K^2\hat{N}_p)$
Combining Weights	$\mathcal{O}\left(KM^3\right)$	$\mathcal{O}(KM^3)$
Post-Combined Channel Matrix	$\mathcal{O}(K^2M)$	$\mathcal{O}(K^2M)$
Data Demodulation	$\mathcal{O}(K)$	$\mathcal{O}(K)$

D. Computational Complexity

Table I shows the computational complexity, based on floating point operation (FLOP) count for matrix-vector operations [16], [17], per iteration, of stage 1 and stage 2. Note that the near-diagonal nature of the post-combined channel matrix was utilized to reduce the computational complexity of data demodulation in stage 1. We iterate OMP for $\hat{N}_p$ times, where $\hat{N}_p$ is an integer exceeding the number of paths, SBL for a fixed number of times, κ, and MVR just once. Stage 1 has a complexity of $\mathcal{O}(K^2(\hat{N}_p+M)+KM^3+(K+\hat{N}_p)NM|S_P|)$ per iteration when using OMP, $\mathcal{O}(K^2(\hat{N}_p+M)+KM^3+KNM|S_P|+\kappa M^2|S_P|^2N)$ per iteration when using SBL, and $\mathcal{O}(K^2(\hat{N}_p+M)+KM^3+(K+\hat{N}_p)NM|S_P|+M^2|S_P|^2N)$ per iteration when using MVR. Stage 2 has a complexity of $\mathcal{O}(K^2(\hat{N}_p+M)+KM^3)$ per iteration when using OMP and MVR, and $\mathcal{O}(K^2(\hat{N}_p+M)+\kappa KM^3)$ per iteration when using SBL for sparse channel recovery.

In the proposed algorithm, the construction of dictionary matrices in (14) consumes most of the CPU time, especially for large K and N. Also, the dictionary matrix must be recomputed in every iteration during stage 1 and stage 2 since the data symbols get updated after the data detection step. We perform a recursive update of the dictionary matrix during the ν^{th} iteration according to

$$\mathbf{A}^{(m)}(\nu) = \mathbf{A}^{(m)}(\nu-1) + \Delta\mathbf{A}^{(m)}(\nu), \tag{23}$$

where $\Delta\mathbf{A}^{(m)}(\nu) = \mathbf{J}^{(m)}\left[\mathbf{\Gamma}_1^{(m)}\Delta\mathbf{s}^{(\nu)}, \dots, \mathbf{\Gamma}_N^{(m)}\Delta\mathbf{s}^{(\nu)}\right]$ and $\Delta\mathbf{s}^{(\nu)} = \mathbf{s}^{(\nu)} - \mathbf{s}^{(\nu-1)}$ is the change in the data symbols

between iterations. As the iteration progresses, $\Delta\mathbf{s}^{(\nu)}$ will contain many zero elements and the corresponding columns of $\Delta\mathbf{A}^{(m)}(\nu)$ will be all zero vectors. Thus, only the columns of $\mathbf{A}^{(m)}(\nu)$ corresponding to the nonzero entries in $\Delta\mathbf{s}^{(\nu)}$ need to be computed, making the dictionary update fast.

IV. Analysis

Recall that the PID outputs are obtained from the same OFDM receiver output as the FID outputs. Yet, using the PID outputs in stage 1 leads to better data detection performance. First, we analyze this behavior and show that using the PID leads to a larger number of effective measurements compared to the FID, which explains its better performance.

A. Effective Number of Measurements

We compare the number of linearly independent rows in the measurement matrices corresponding to the PID and the FID outputs for pilot only measurements in the stage 1. It is easy to see that:

$$\text{rank}\left(\sum_{m=1}^{M}\mathbf{A}_{S_P}^{(m)}\right) \leq \text{rank}\left(\tilde{\mathbf{A}}_{S_P}\right) \leq \min\left(M|S_P|, N\right), \tag{24}$$

where $\tilde{\mathbf{A}}_{S_P}$ is the augmented matrix obtained by vertically stacking $\mathbf{A}_{S_P}^{(m)}, m = 1, 2, \cdots, M$. Therefore, there are at least as many linearly independent pilot only measurements at the output of the PID, whose sensing matrix is $\tilde{\mathbf{A}}_{S_P}$, as the FID, whose sensing matrix is $\mathbf{A}_{S_P} \triangleq \sum_{m=1}^{M}\mathbf{A}_{S_P}^{(m)}$. We proceed to show that it is possible to obtain strictly more linearly independent measurements from the PIDs than from the FID.

In the first iteration of stage 1, we set all the entries of $\hat{\mathbf{s}}$, except pilot locations, to zero. If all the pilot symbols are equal to, say, s_0, then, the dictionary matrix constructed in the first iteration of stage 1 is given by:

$$\mathbf{A}_{S_P}^{(m)} = s_0\mathbf{I}_{S_P}\mathbf{J}^{(m)}\mathbf{I}_{S_P}^{T}\mathbf{I}_{S_P}\mathbf{\Gamma}^{(m)}, \tag{25}$$

where,

$$\mathbf{\Gamma}^{(m)} \triangleq \left[\mathbf{\Gamma}_1^{(m)}\mathbf{1}_K \ldots \mathbf{\Gamma}_{N_\tau N_b}^{(m)}\mathbf{1}_K\right], \tag{26}$$

$\mathbf{1}_K \in \mathbb{R}^K$ is a column vector of K ones and $\mathbf{I}_{S_P}$ is as defined after (16). The matrix $\mathbf{\Gamma}^{(m)}$ can be decomposed as:

$$\begin{aligned}
\mathbf{\Gamma}^{(m)} &= \mathbf{\Gamma}_b^{(m)} \otimes \mathbf{\Gamma}_\tau, &(27)\\
\mathbf{\Gamma}_b^{(m)} &= e^{j2\pi\mathbf{f}\mathbf{b}^T t_m} \in \mathbb{C}^{K\times N_b}, &(28)\\
\mathbf{\Gamma}_\tau &= e^{-j2\pi\mathbf{f}\tau^T} \in \mathbb{C}^{K\times N_\tau}, &(29)
\end{aligned}$$

where $\otimes$ denotes the Khatri-Rao product,[4] $e^{\mathbf{B}}$ finds element-wise exponentiation of a matrix $\mathbf{B}$, and

$$\begin{aligned}
\mathbf{f} &= [f_1, \ldots, f_K]^T \in \mathbb{R}^K, &(30)\\
\tau &= \left[\frac{T}{\lambda K}, \frac{2T}{\lambda K}, \ldots, \frac{N_\tau T}{\lambda K}\right]^T \in \mathbb{R}^{N_\tau}, &(31)\\
\mathbf{b} &= [-b_{\max}, -b_{\max}+\Delta b, \ldots, b_{\max}]^T \in \mathbb{R}^{N_b}. &(32)
\end{aligned}$$

[4]The Khatri-Rao product is formed by taking the row-wise Kronecker products of $\mathbf{\Gamma}_b^{(m)} \in \mathbb{C}^{K\times N_b}$ and $\mathbf{\Gamma}_\tau \in \mathbb{C}^{K\times N_\tau}$.

It readily follows from (27) and the definition of the Khatri-Rao product that,

$$\mathbf{I}_{S_P}\mathbf{\Gamma}^{(m)} = \mathbf{I}_{S_P}\mathbf{\Gamma}_b^{(m)} \otimes \mathbf{I}_{S_P}\mathbf{\Gamma}_\tau. \tag{33}$$

Let $\tilde{\mathbf{\Gamma}}_{S_P} \in \mathbb{C}^{M|S_P|\times N_\tau N_b}$ and $\tilde{\mathbf{\Gamma}}_{b,S_P} \in \mathbb{C}^{M|S_P|\times N_b}$ denote the augmented matrices obtained by vertically stacking $\mathbf{I}_{S_P}\mathbf{\Gamma}^{(m)}, m = 1, \ldots, M$, and $\mathbf{I}_{S_P}\mathbf{\Gamma}_b^{(m)}, m = 1, \ldots, M$, respectively. Similarly, let $\tilde{\mathbf{\Gamma}}_{\tau,S_P} \in \mathbb{C}^{M|S_P|\times N_\tau}$ be the augmented matrix obtained by vertically stacking $\mathbf{I}_{S_P}\mathbf{\Gamma}_\tau$ M-times. We then have:

$$\tilde{\mathbf{\Gamma}}_{S_P} = \tilde{\mathbf{\Gamma}}_{b,S_P} \otimes \tilde{\mathbf{\Gamma}}_{\tau,S_P}. \tag{34}$$

We now state two properties of the Khatri-Rao product that are useful in the sequel. We denote k-rank$(\mathbf{B})$ to be the row Kruskal-rank[5] of a matrix $\mathbf{B}$.

Lemma 1. *The rank (k-rank) of the Khatri-Rao product of two matrices, both having at least one of the columns with all its entries nonzero, is never less than the rank (k-rank) of the two matrices, i.e., if $\mathbf{B} \in \mathbb{C}^{r\times p}$ and $\mathbf{C} \in \mathbb{C}^{r\times q}$ are two matrices such that for some $n \in \{1, 2, \ldots, p\}$ and $n' \in \{1, 2, \ldots, q\}$, we have $[\mathbf{B}]_{m,n} \neq 0$ and $[\mathbf{C}]_{m,n'} \neq 0$ for every $m \in \{1, 2, \ldots, r\}$, then:*

$$\begin{aligned}
\text{rank}\left(\mathbf{B}\otimes\mathbf{C}\right) &\geq \max\left(\text{rank}\left(\mathbf{B}\right), \text{rank}\left(\mathbf{C}\right)\right), &(35)\\
k\text{-rank}\left(\mathbf{B}\otimes\mathbf{C}\right) &\geq \max\left(k\text{-rank}\left(\mathbf{B}\right), k\text{-rank}\left(\mathbf{C}\right)\right). &(36)
\end{aligned}$$

Proof. See Appendix. □

Lemma 2. *If $\mathbf{B} \in \mathbb{C}^{r\times p}$ and $\mathbf{C} \in \mathbb{C}^{r\times q}$ then*

$$k\text{-rank}\left(\mathbf{B}\otimes\mathbf{C}\right) \geq \min\left(k\text{-rank}\left(\mathbf{B}\right) + k\text{-rank}\left(\mathbf{C}\right) - 1, r\right). \tag{37}$$

Proof. See Appendix. □

All entries of the matrices $\mathbf{I}_{S_P}\mathbf{\Gamma}_b^{(m)}$ and $\mathbf{I}_{S_P}\mathbf{\Gamma}_\tau$, in (34), have unit magnitude and therefore satisfy the conditions in Lemma 1. Also, it is straightforward to see that

$$\begin{aligned}
\text{rank}\left(\tilde{\mathbf{\Gamma}}_{\tau,S_P}\right) &= \text{rank}\left(\mathbf{I}_{S_P}\mathbf{\Gamma}_\tau\right), &(38)\\
k\text{-rank}\left(\tilde{\mathbf{\Gamma}}_{\tau,S_P}\right) &= 1, &(39)
\end{aligned}$$

and therefore it follows from Lemma 1 that

$$\begin{aligned}
\text{rank}\left(\tilde{\mathbf{\Gamma}}_{S_P}\right) &\geq \max\left(\text{rank}\left(\tilde{\mathbf{\Gamma}}_{b,S_P}\right), \text{rank}\left(\mathbf{I}_{S_P}\mathbf{\Gamma}_\tau\right)\right), &(40)\\
k\text{-rank}\left(\tilde{\mathbf{\Gamma}}_{S_P}\right) &\geq k\text{-rank}\left(\tilde{\mathbf{\Gamma}}_{b,S_P}\right). &(41)
\end{aligned}$$

We are now ready to state a *sufficient* condition that results in strictly larger number of independent measurements from the PID output than the FID. We say that a collection of subspaces $\mathcal{S} = \{S_i \in \mathbb{V} : i = 1, 2, \cdots, n\}$ of a vector space $\mathbb{V}$ forms a *virtually disjoint partition* if $\mathbb{V}$ is a *direct sum* of its subspaces $S_i, i = 1, 2, \cdots, n$, i.e., if $\oplus_{i=1}^{n} S_i = \mathbb{V}$ and $\cap_{i=1}^{n} S_i = \{\mathbf{0}\}$, where the operator $\oplus$ denotes the subspace sum [18].

[5]The row Kruskal-rank (k-rank) of a matrix is r if every subset of its r rows is linearly independent and at least one subset of $r+1$ rows is linearly dependent.

Theorem 1. *If the number of grids N_b used for representing the Doppler parameter, and the augmented matrix $\tilde{\mathbf{\Gamma}}_{b,S_P}$ generated by the representative Doppler values forming the grid, satisfy:*

$$N_b + N_\tau \geq M|S_P| + 1, \tag{42}$$

$$k\text{-rank}\left(\tilde{\mathbf{\Gamma}}_{b,S_P}\right) = \min\left(N_b, M|S_P|\right), \tag{43}$$

and the pilot locations in S_P are chosen such that

$$\text{rank}\left(\mathbf{I}_{S_P}\mathbf{J}^{(m)}\mathbf{I}_{S_P}^T\right) = |S_P|, m = 1, \ldots, M, \tag{44}$$

then:

$$|S_P| = \text{rank}\left(\sum_{m=1}^{M}\mathbf{A}_{S_P}^{(m)}\right) < \text{rank}\left(\tilde{\mathbf{A}}_{S_P}\right) = M|S_P|. \tag{45}$$

Proof. See Appendix. □

Discussion: The result in Theorem 1 indicates that the pilot only measurements from the PID outputs can potentially lead to better channel estimates compared to that from the FID. The conditions (42) and (43) are not necessary; we find in our simulation studies that far fewer number of grid points N_b in the Doppler parameter leads to (45) being satisfied. To illustrate this point, Figures 4 and 5 show the distribution of singular values of the stacked up dictionary matrix $\tilde{\mathbf{A}}_{S_P}$ corresponding to the PID output for pilot subcarriers when the Doppler spreads are $b_{\max} = 5 \times 10^{-4}$ and $b_{\max} = 10^{-3}$, respectively, and when pilot carriers are chosen as in the numerical case study in Section V. Also shown in this figure is a plot of the singular values of the dictionary matrix $\mathbf{A}_{S_P}$ corresponding to FID observation of pilot subcarriers. It is clear that the numerical rank of the stacked up dictionaries of the PID is greater than that of the FID, especially, when the dictionaries are designed for high Doppler spread. Therefore, using the sequence of observations from PIDs, which is tantamount to oversampling the OFDM receiver output, helps estimate the channel better.

We have shown that the sparse channel estimation from the PID output provides a good initial estimate of the channel matrix. Next, we justify that the proposed two stage algorithm that iterates between the sparse channel estimation and data detection steps in each stage can only improve the channel estimation and data detection accuracy with every iteration.

B. Convergence

Recall that our channel estimation is based on the sparse vector recovery framework expressed by (15) and (22) for stage 1 and stage 2, respectively. We recapitulate the channel estimation framework in the following form:

$$\mathbf{z} = \mathbf{A}(\mathbf{s})\mathbf{x} + \mathbf{v}, \tag{46}$$

where $\mathbf{A}(\mathbf{s})$ indicates the dependence of dictionary matrix on the data symbols $\mathbf{s}$. Similarly, the data detection problem in the stage 1 and stage 2 is given by

$$\mathbf{z} = \mathbf{H}(\mathbf{x})\mathbf{s} + \mathbf{v}, \tag{47}$$

where the dependence of the channel matrix on the vector $\mathbf{x}$ is indicated via $\mathbf{H}(\mathbf{x})$. Note that $\mathbf{A}(\mathbf{s})\mathbf{x} = \mathbf{H}(\mathbf{x})\mathbf{s}$.

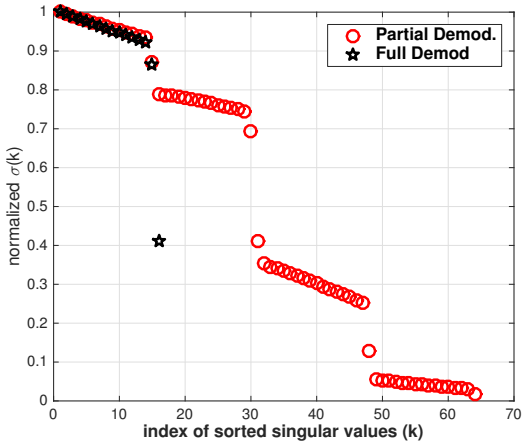

Figure 4. Singular values of the dictionary matrices corresponding to PID and FID outputs. Doppler spread $b_{\max} = 0.5 \times 10^{-3}$.

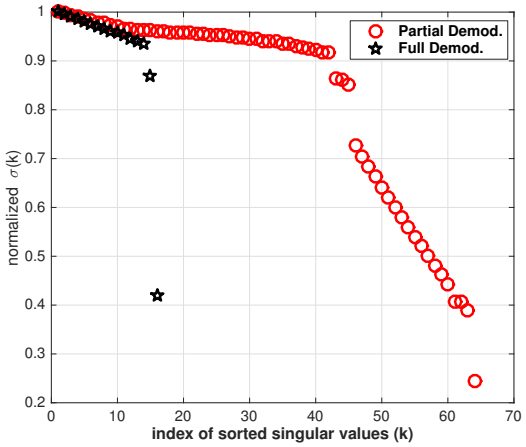

Figure 5. Singular values of the dictionary matrices corresponding to PID and FID outputs. Doppler spread $b_{\max} = 10^{-3}$.

Ideally, for estimating the sparse channel, we would like to minimize $\|\mathbf{x}\|_0$ subject to $\|\mathbf{z} - \mathbf{H}(\mathbf{x})\mathbf{s}\|_2 \leq \eta$. This problem is NP-hard, and, therefore, we consider its convex relaxation:

$$C(\mathbf{x}, \mathbf{s}) = \|\mathbf{x}\|_1 + \lambda\|\mathbf{z} - \mathbf{H}(\mathbf{x})\mathbf{s}\|_2^2, \tag{48}$$

over $\mathbf{x}, \mathbf{s} \in \mathbb{C}$.

Let $\mathbf{x}^{(\nu)}$ denote the sparse channel vector estimated in the ν^{th} iteration. For a given $\mathbf{x}^{(\nu)}$, choosing

$$\mathbf{s}^{(\nu+1)} = \arg\min_{\mathbf{s}\in\mathbb{C}}\|\mathbf{z} - \mathbf{H}(\mathbf{x}^{(\nu)})\mathbf{s}\|_2, \tag{49}$$

ensures that $C(\mathbf{x}^{(\nu)}, \mathbf{s}^{(\nu+1)}) \leq C(\mathbf{x}^{(\nu)}, \mathbf{s}^{(\nu)})$. For a given $\mathbf{s}^{(\nu+1)}$, choosing

$$\mathbf{x}^{(\nu+1)} = \arg\min_{\mathbf{x}\in\mathbb{C}} C(\mathbf{x}, \mathbf{s}^{(\nu+1)}), \tag{50}$$

ensures that $C(\mathbf{x}^{(\nu+1)}, \mathbf{s}^{(\nu+1)}) \leq C(\mathbf{x}^{(\nu)}, \mathbf{s}^{(\nu)})$. Therefore, the cost $C(\mathbf{x}, \mathbf{s})$ reduces with every iteration and is bounded below by 0. Hence, the iterations in stage 1 and stage 2 converge to a local minimum of the ℓ_1 regularized joint cost function. We can show that, by associating a Laplacian prior to the channel vector $\mathbf{x}$, the solution $(\mathbf{x}, \mathbf{s})$ which minimizes $C(\mathbf{x}, \mathbf{s})$ is also a solution to joint channel estimation and maximum *a posteriori* probability (MAP) data detection problem. Note

that the above recipe of iterating between channel estimation and data detection is unlike joint receivers that provide a low complexity approximation to MAP detection [19], [20].

C. Lower Bound on the MSE

We obtain a lower bound on the MSE in the estimate of the channel matrix (see remark below) for the two classes of *unbiased* estimators: one that makes use of the observations at the output of the PID, and the other that uses only the observations from the FID. We have the following theorem.

Theorem 2. *For the PID measurement model in* (13)*, the MSE in the channel matrix* $\mathbf{H}$*, which is a function of the channel vector* $\mathbf{x}$*, can be lower bounded as*

$$\mathbb{E}\{\|\mathbf{H}-\hat{\mathbf{H}}\|_F^2\} \geq \ tr\left(\mathbf{F}_{\mathcal{S},\mathcal{S}}^{-1}\frac{\partial \mathbf{h}}{\partial \mathbf{x}_{\mathcal{S}}}^{H}\frac{\partial \mathbf{h}}{\partial \mathbf{x}_{\mathcal{S}}}\right), \tag{51}$$

$\mathcal{S}$ is the support of the channel vector $\mathbf{x}$*,* $\mathbf{F}_{\mathcal{S},\mathcal{S}} \in \mathbb{C}^{|\mathcal{S}|\times|\mathcal{S}|}$ *is the submatrix of the Fisher Information Matrix (FIM)* $\mathbf{F} \in \mathbb{C}^{N\times N}$*, corresponding to the rows and columns indexed by* $\mathcal{S}$*, for the observation model in* (13)*, given by*

$$\mathbf{F} = \sum_{m=1}^{M} \mathbf{A}^{(m)H}\mathbf{C}_m^{\dagger}\mathbf{A}^{(m)}, \tag{52}$$

$\mathbf{C}_m$ *is the covariance of the noise at the output of the PID, whose entries are given by* (8)*,* $\mathbf{h}(\mathbf{x}) \triangleq vec\,(\mathbf{H}) \in \mathbb{C}^{K^2\times 1}$*, the columns of the matrix* $\frac{\partial \mathbf{h}}{\partial \mathbf{x}} \in \mathbb{C}^{K^2\times N}$ *are given by*

$$\frac{\partial \mathbf{h}(\mathbf{x})}{\partial \mathbf{x}_k} = \begin{cases} vec\left(\sum_{m=1}^{M}\mathbf{J}^{(m)}\mathbf{\Gamma}_k^{(m)}\right), & k \in \mathcal{S}, \\ \mathbf{0}_{K^2}, & otherwise, \end{cases} \tag{53}$$

and $\frac{\partial \mathbf{h}}{\partial \mathbf{x}_{\mathcal{S}}}$ *is the submatrix of* $\frac{\partial \mathbf{h}}{\partial \mathbf{x}}$ *consisting of only the columns indexed by* $\mathcal{S}$*.*

Proof. See Appendix. □

Theorem 3. *A lower bound on the MSE of the unbiased estimators of the channel matrix* $\mathbf{H}$*, that uses the observations from the FID measurement model in* (22)*, is given by*

$$\mathbb{E}\{\|\mathbf{H}-\hat{\mathbf{H}}\|_F^2\} \geq \ tr\left(\mathbf{G}_{\mathcal{S},\mathcal{S}}^{-1}\frac{\partial \mathbf{h}}{\partial \mathbf{x}_{\mathcal{S}}}^{H}\frac{\partial \mathbf{h}}{\partial \mathbf{x}_{\mathcal{S}}}\right), \tag{54}$$

where $\mathcal{S}$ *is the support of the channel vector* $\mathbf{x}$*,* $\mathbf{G} \in \mathbb{C}^{N\times N}$ *is the FIM for the FID observation model in* (22)*, given by:*

$$\mathbf{G} = \frac{1}{N_0}\mathbf{A}^H\mathbf{A}, \tag{55}$$

$\mathbf{h}(\mathbf{x}) \triangleq vec\,(\mathbf{H}) \in \mathbb{C}^{K^2\times 1}$*, and the columns of the matrix* $\frac{\partial \mathbf{h}}{\partial \mathbf{x}} \in \mathbb{C}^{K^2\times N}$ *are given as in Theorem 2.*

Proof. Similar to Theorem 2. □

Remarks:

1) It is tempting to consider the MSE in the estimate of the sparse channel vector $\mathbf{x}$, in (12), defined on the delay-Doppler grid. While the MSE in the channel vector $\mathbf{x}$ relates to the MSE in the channel matrix $\mathbf{H}$, we assert that the latter is more meaningful for our problem. This is because, a small mismatch in support estimation (i.e., when the recovered support returns indices near the true support) can lead to a large MSE in the sparse vector $\mathbf{x}$, but need not translate to a large MSE in the channel. However, the data detection performance primarily depends on the fidelity in the estimation of $\mathbf{H}$, and not as much on $\mathbf{x}$. Hence, we consider the MSE in the channel matrix $\mathbf{H}$ as the performance metric in this work.
2) Theorems 2 and 3 allow us to compare the bounds on the MSE of unbiased estimators that use the PID outputs and those that use the FID outputs. Numerical evaluation shows that the bound on the MSE for the estimators that make use of PID outputs is indeed better. Also, our proposed two-stage data detection and channel estimation algorithm, that makes use of the PID outputs in stage 1 and the FID outputs in stage 2, approaches the lower bound at high SNR. See Section V for details.

V. Numerical Simulations

We simulate the performance of the proposed algorithm for the CP-OFDM system whose parameters are listed in Table II. The specifications of the system matches with the settings used in the SPACE'08 experiment and is widely used for simulation studies in several past works, for example, [3], [7], and [8]. Pilot symbols are spaced uniformly. Half the null carriers are placed at the band edges and the remaining are inserted between the data as specified in [3], [7].

We consider two simulation models for the underwater acoustic communication channel. In model I, adopted from [3], [7], we generate sparse channels with a few discrete paths whose inter-arrival times are exponentially distributed with a mean of 1 ms. The path amplitudes are Rayleigh distributed with the average power decreasing exponentially with delay, where the difference between the beginning and the end of the guard time is 20 dB. The residual Doppler rate for each path is uniformly distributed in $[-b_{\max}, b_{\max}]$, where channels with $b_{\max} = 5\times 10^{-4}$ and 1×10^{-3} are considered to be severely Doppler distorted in the underwater communication literature [21]. In model II, we simulate the time-varying stochastic channel response according to the model proposed in [2]. Model II incorporates the effect of frequency dependent attenuation, the surface/bottom scattering and other random fluctuations in the medium and source-receiver position. The authors report a good match of their theoretical model with the experimental data collected from four different deployment sites of varying degrees of mobility. We use model II to demonstrate the relatively strong performance of the proposed algorithm to mismatches in the model assumptions.

We finally show the performance of our proposed algorithm on the measured time varying channel impulse response data available in WATERMARK [22]. We consider two channel

Table II
CP-OFDM PARAMETERS USED IN THE SIMULATION.

Carrier frequency (f_c)	13 kHz
Bandwidth (B)	9.77kHz
No. of subcarriers (K)	1024
No. of pilots ($\|S_P\|$)	256
No. of nulls ($\|S_N\|$)	96
Symbol duration (T)	104.86 ms
Subcarrier spacing (Δf)	9.54 Hz
Guard interval (T_g)	24.6 ms

measurements corresponding to a low Doppler spread channel and a high Doppler spread channel. These measured responses in WATERMARK include the effect of system hardware and the real world acoustic propagation as well.

We define the signal to noise ratio (SNR) as SNR $= \frac{E\{\|\mathbf{Hs}\|_2^2\}}{E\{\|\mathbf{v}\|_2^2\}}$, and the normalized MSE in channel estimation as NMSE $= \frac{E\{\|\mathbf{H}-\hat{\mathbf{H}}\|_F^2\}}{E\{\|\mathbf{H}\|_F^2\}}$, where $\mathbf{H} \in \mathbb{C}^{K\times K}$ and $\hat{\mathbf{H}} \in \mathbb{C}^{K\times K}$ are the true and the estimated channels, respectively.

A. Simulations using Model I

We first consider coded 16-QAM transmissions, and channels generated according to model I with $N_p = 15$ discrete paths and $b_{\max} = 10^{-3}$. The turbo code uses two rate-1/2 convolutional encoders with feedback and an interleaver of length 232 bits [23]. The bit error rate (BER) and normalized MSE are averaged over 1000 independent instantiations. For constructing the dictionary matrix, grids are formed using $N_b = 15$ points for the Doppler rate and $N_\tau = 480$ points for the delay resolution corresponding to an oversampling factor of $\lambda = 2$ over the guard interval. The PID and FID dictionary matrices are, therefore, of size 1024×7200.

For the sparse channel recovery, we experiment with OMP [24] and MVR. For OMP, we set the number of nonzero entries to be recovered to $\hat{N}_p = 25$. The number of propagation paths need not be known precisely; as long as $\hat{N}_p > N_p$, simulation studies show good channel recovery. The MVR algorithm is iterated only once. We compare our proposed dual stage algorithm that uses PID outputs in stage 1 against the algorithms in [3], [7], the least squares based channel estimation and data detection, and genie-aided data detection which uses the channel state information. For simulating the algorithm in [7], we use the ICI-aware receiver with an ICI depth parameter of $D = 6$.

In Figure 6, we show the normalized MSE in the channel matrix estimate for the different algorithms. For the proposed dual-stage algorithm, we iterate for $N_{\text{iter}} = 3$ through both stage 1 and stage 2. We use $M = 4$ PIDs in our simulations.[6] For the value of M used here, from equation (8), the noise at the output of the PID is uncorrelated among the pilot-only observations. The MVR algorithm leads to a lower MSE in the channel matrix estimate compared to OMP. The normalized MSE performance of the algorithms in [3], [7], that use the FID output for pilot-only measurements, is shown in Figure 6. The algorithm in [7] is similar to [3] but iterates to improve the data symbol detection; we use $N_{\text{iter}} = 6$ in the simulations of the iterative algorithm in [3]. Also included in Figure 6 are the normalized MSE performances of two least square channel estimation algorithms labeled `LS CDD (FID)` and `LS CDD (PID, Optm. Wt.)` in the plot. LS CDD (FID) estimates the channel using pilot measurements at FID output. LS CDD (PID, Optm. Wt.) estimates the channel using pilot measurements at optimally combined PID output. Only the combiner weights, for LS CDD (PID, Optm. Wt.), are computed using the true channel matrix (genie aided). Note that the performance of [8], in which combiner weights are estimated through an adaptive algorithm, cannot be better than LS CDD (PID, Optm. Wt.). Stage 1 recovers the channel better than the existing FID and PID based algorithms and hence provides a good estimate of unknown data symbols to initialize stage 2. The MSE in the channel matrix further reduces at the end of stage 2, especially at high SNR, which leads to better symbol detection performance. Figure 7 shows the reduction in MSE for the iterative algorithms compared above, as the number of iterations is increased to $N_{\text{iter}} = 10$ at an SNR of 16 dB. It is seen that the proposed dual stage algorithm, at the end of stage 2, has settled in about 3 iterations.

Figure 8 compares the bit error rate (BER) versus SNR curves of the proposed dual stage algorithm with the least squares based channel estimation and coherent data detection algorithms LS CDD (FID) and LS CDD (PID, Optm. Wt.), the sparse channel recovery based algorithms reported in [3], [7] and the genie-aided data detection as a baseline. Note that LS CDD (FID) and LS CDD (PID, Optm. Wt.) perform coherent data detection at the FID and optimally combined PID outputs, respectively. The dual stage iterative algorithm proposed in this paper clearly outperforms the existing sparse channel recovery based algorithms and the least squares channel estimation based data detection algorithms, at all SNRs. Also, compared to the OMP based sparse channel recovery algorithm, the BER curve of the MVR based sparse channel recovery is closer to that of the genie aided data detection at the end of the stage 2.

Figure 9 shows the CRBs corresponding to the PID and the FID output, computed using (51) and (54), respectively. The CRB on the MSE of estimators that use only the FID output is higher than those that makes use of the output from the PID. Moreover, among the two sparse channel recovery algorithms, MVR based channel estimation achieves the CRB corresponding to PID observation at an SNR of about 30 dB, at the end of stage 2.

Figure 10 shows the BER performance when the Doppler scale is varied. While all the schemes perform nearly identical at zero Doppler spread, the performance gap of between the proposed scheme and other schemes widens as the Doppler scale increases, highlighting the significance of PID based channel estimation in high Doppler spread scenarios.

Next, we examine the effect of pilot density on the BER performance [25], [26]. Figure 11 depicts the BER performance as the number of pilots used is varied while

[6]Increasing M will in general improve the performance due to additional measurements being made available, but will also increase the computational complexity. Also, increasing M beyond a certain point will not yield significantly more effective measurements.

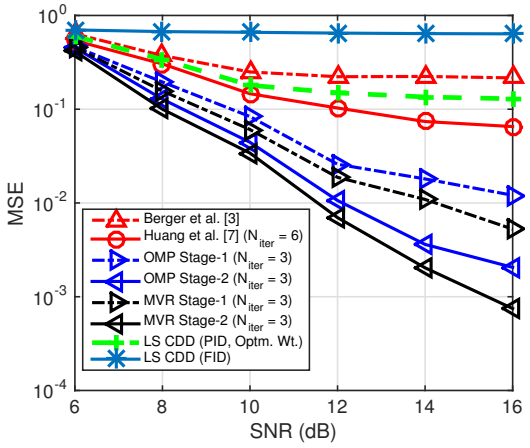

Figure 6. MSE in the channel matrix estimate of various sparse signal recovery algorithms. Doppler spread $b_{\max} = 10^{-3}$.

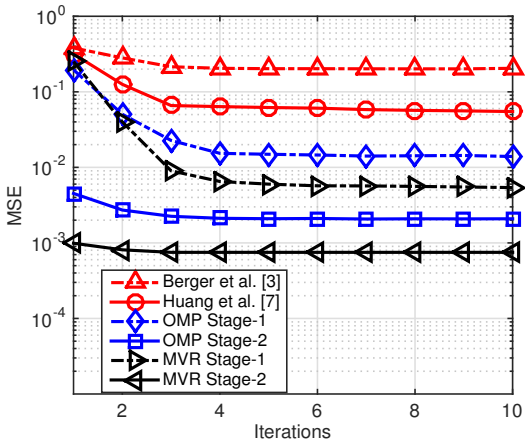

Figure 7. MSE versus number of iterations N_{iter} for various sparse channel recovery algorithms at SNR = 16 dB. Doppler spread $b_{\max} = 10^{-3}$.

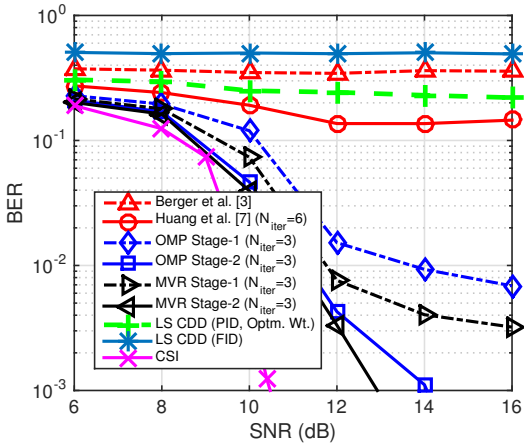

Figure 8. BER of various sparse channel recovery based algorithms and the genie-aided perfect CSI lower bound. Doppler spread $b_{\max} = 10^{-3}$.

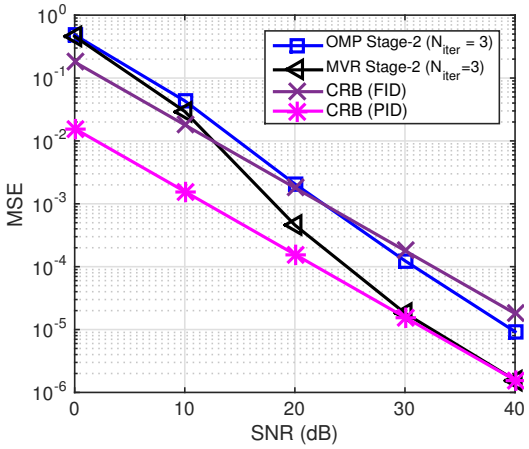

Figure 9. Comparison of the MSE in the channel matrix estimation of various sparse channel recovery based algorithms with the lower bound on the MSE for the FID and the PID observation models. Doppler spread $b_{\max} = 10^{-3}$.

keeping the SNR constant at 12 dB and $b_{\max} = 10^{-3}$. We consider a pilot arrangement of sub-arrays of pilots, where each sub-array consists of 32 pilots at a regular spacing of 4. Also, the sub-arrays are spaced uniformly. For example, when $|S_P| = 96$, we form 3 sub-arrays each comprising of 32 pilots, and the spacing between the first elements of two consecutive sub-arrays is 128 subcarriers. We see that the gap between the proposed and existing algorithms increases dramatically with pilot density, implying that the proposed algorithm can achieve a given performance at a significantly lower pilot density.

B. Simulations using Model II

Now we examine the performance when the channel is simulated according to the model proposed in [2]. The time varying frequency response of the channel is modeled as

$$\tilde{H}(f,t) = \bar{H}_0(f) \sum_{p=1}^{N_p} h_p \tilde{\gamma}_p(f,t) e^{-j2\pi f \tau_p}, \qquad (56)$$

where $\bar{H}_0(f)$ is the nominal frequency response of the direct path between the source and receiver that results in a frequency dependent propagation loss, h_p is the nominal channel coefficient of the p^{th} path that arrives at a nominal delay of τ_p, and the stochastic term $\tilde{\gamma}_p(f,t) = \gamma_p(f,t)e^{2\pi a_p(t) f t}$ is composed of the small-scale fading $\gamma_p(f,t)$ and the Doppler scale factor $a_p(t)$ corresponding to the p^{th} propagation path. The small-scale fading $\gamma_p(f,t)$ of the p^{th} path, arises from scattering at the rough sea surface and bottom that leads to a bunch of micro-paths whose amplitudes and delays are randomly distributed around that of the nominal ray path. The Doppler scale factor $a_p(t)$ is a composite effect of the vehicular motion, surface wave perturbations, and relative source-receiver drifts. The received signal for this channel is given by

$$\tilde{y}(t) = \int_0^T \tilde{c}(\tau,t)\tilde{x}(t-\tau)d\tau + \tilde{n}(t), \qquad (57)$$

where $\tilde{c}(\tau,t) = \mathcal{F}_f^{-1}\{\tilde{H}(f,t)\}$ is the time-varying channel impulse response and $\mathcal{F}_f^{-1}$ is the inverse Fourier transform.

Note that the frequency dependence of the stochastic term $\tilde{\gamma}_p(f,t)$ distinguishes the channel model in (56) from the

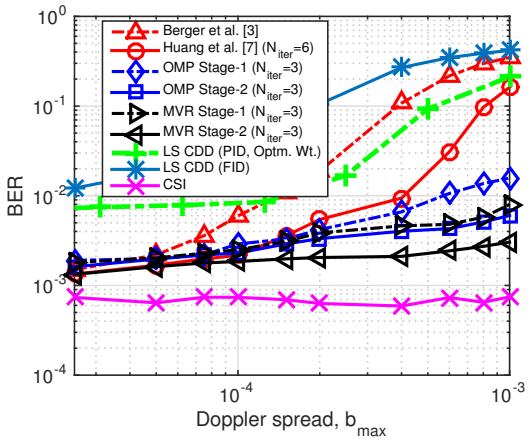

Figure 10. BER comparison of various channel estimation and data detection algorithms for different Doppler spreads and at SNR = 12 dB.

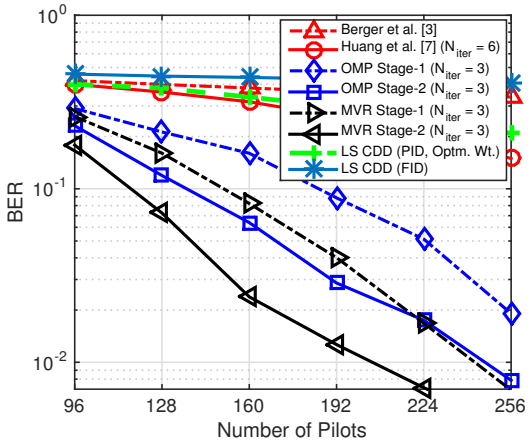

Figure 11. BER performance of system II with number of pilots used. SNR = 12 dB, Doppler spread $b_{\max} = 10^{-3}$.

channel model implicit in (3). The model in (3) is widely used in the underwater communication literature for designing algorithms and benchmarking performance. In this work, additionally, we test the robustness of our proposed algorithm by evaluating its performance under the received signal given by (57) instead of (3).

Figure 12 shows a sample realization of the effective channel impulse response $\tilde{c}_r(\tau, t) = \tilde{c}(\tau, \frac{t}{1+\tilde{a}})$ after resampling at the receiver. The channel is generated using the acoustic channel simulator code available in [27]. Table III shows the environmental parameters and source-receiver geometry used in this simulation. The transmitter and receiver are in a shallow water environment overlying a soft bottom. Small-scale surface variations, and relative drifts between the source and receiver, cause the channel taps to randomly fluctuate about their slowly varying mean. Specifically, the Doppler scales for this channel vary between $[-8.75 \times 10^{-4},\ 10.0 \times 10^{-4}]$. Figure 13 shows the Doppler spectrum of the simulated channel averaged across the channel taps. A sample plot of an instantaneous channel response across frequency, obtained at subcarrier spacing, is shown in Figure 14. Significant variations are noticed between adjacent subcarriers for a typical instance.

Table III
UNDERWATER CHANNEL SIMULATION PARAMETERS.

Ocean depth (m)	100
Transmitter depth (m)	90
Receiver depth (m)	50
Channel distance (m)	1000
Spreading factor	1.7
Sound speed in water, c_w (m/s)	1500
Sound speed in bottom, c_b (m/s)	1200
Surface variance, σ_s^2 (m^2)	1.125
Bottom variance, σ_b^2 (m^2)	0.5
3 dB width of the PSD of intra-path delays, $B_{\delta,p}$ (Hz)	0.05
Number of intra-paths, S_p	20
Mean of intra-path amplitudes, μ_p	0.3
Variance of intra-path amplitudes, ν_p	10^{-4}
Transmitter drifting speed, v_{td} (m/s)	0.3
Transmitter drifting angle, θ_{td} (rad)	$\mathcal{U}(0, 2\pi)$
Receiver drifting speed, v_{rd} (m/s)	0.1
Receiver drifting angle, θ_{rd} (rad)	$\mathcal{U}(0, 2\pi)$
Transmitter vehicular speed, v_{tv} (m/s)	$\mathcal{N}(0, 1)$
Transmitter vehicular angle, θ_{tv} (rad)	$\mathcal{U}(0, 2\pi)$
Receiver vehicular speed, v_{rv} (m/s)	-3
Receiver vehicular angle, θ_{rv} (rad)	$\mathcal{U}(0, 2\pi)$
Surface variation amplitude, A_w (m)	0.9
Surface variation frequency, f_w (mHz)	0.6

The channel *parameters* (path delays, Doppler scales, and amplitudes) vary smoothly between the successive OFDM symbols, for the acoustic channel simulator. For such channels, pilots need to be inserted only in the first OFDM block where the stage 1 runs in the pilot assisted mode followed by stage 2 that uses both pilot and data subcarrier measurements.[7] In the subsequent blocks, where no pilots are available, the algorithm switches to a decision directed mode where the pilots are now replaced with tentative estimates of data symbols. The tentative symbol estimates are formed using channel parameters found at the end of stage 2 in the previous block. An iteration of stage 1 and stage 2 tracks the channel and updates the symbols for the current block. More generally, the schemes in [28], [29] exploit the channel coherence between the OFDM blocks to reduce the pilot overhead. They can be applied to improve the initial symbol estimates at the beginning of a new block in our algorithmic framework also. However, here we implement the simple approach just described.

Figure 15 shows the BER of the different algorithms for this channel. Pilots are employed only in the first OFDM block. While there are only about five significant ray paths in the channel impulse response, each ray path is in turn a bundle of several micro-paths. Therefore, we set the OMP based sparse channel estimator to recover a higher number of paths. We set the number of nonzero entries to be recovered to $\hat{N}_p = 46$ and use the same dictionary as before with $b_{\max} = 10^{-3}$. The proposed algorithm maintains a strong relative performance even in this difficult environment.

[7]A pilot overhead of 30%, as in [3] and [7], is required for the previous channel model because the channel *parameters* were drawn independently in successive OFDM symbols.

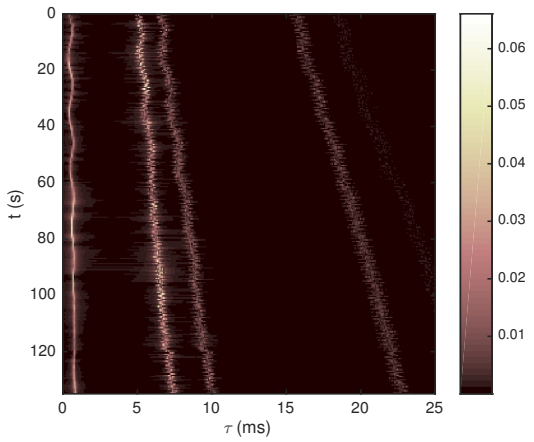

Figure 12. Acoustic channel impulse response based on model in [2].

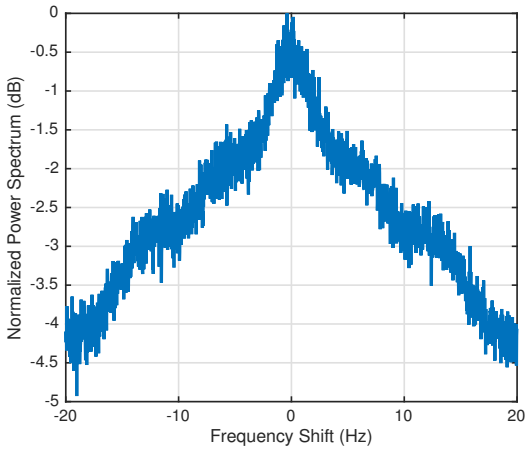

Figure 13. Doppler spectrum of the simulated acoustic channel.

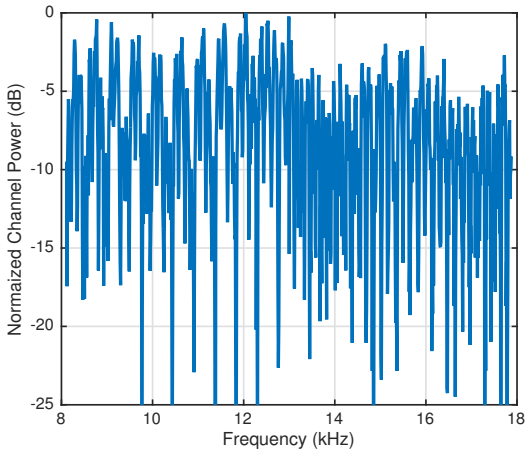

Figure 14. Instantaneous frequency response of the simulated channel.

C. *Performance on* WATERMARK *Data*

WATERMARK is a recently proposed benchmark for comparing the performance of physical layer algorithms for underwater acoustic communications [22], [30], [31]. WATERMARK comes packaged with real world records of time-varying channel impulse response measurements for different environments and source-receiver geometries. The measured responses include the acoustic propagation effects and the system hardware effects such as clock frequency offset. The replay facility allows for transmitting any communication waveform in the frequency band of measured responses. These features make WATERMARK a realistic and reproducible performance assessment tool.

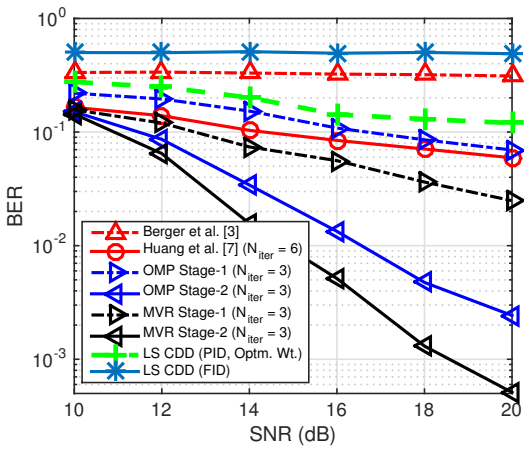

Figure 15. BER of various channel estimation and data detection algorithms on simulated acoustic channel.

We show the performance on two datasets, NOF1 and NCS1, corresponding to a low Doppler spread and high Doppler spread channel in the Norwegian shallow water and continental shelf, respectively (Table 2.1, [22]). The frequency band of both datasets is 10-18 kHz. Figure 16 shows an instantaneous channel response across frequency, obtained at subcarrier spacing, typical of the NOF1 and NCS1 channels. The Doppler spectrum of NOF1 has a sharp peak around zero frequency, with sidelobes less than -20 dB relative to the peak. On the other hand, the Doppler spectrum of NCS1 is significantly spread out within $[-15, 15]$ Hz (see Figure 3 in [30]). Doppler frequency spread of $\delta f = \pm 15$ Hz in NCS1 corresponds to a Doppler scale of $b_{\max} \approx \frac{|\delta f|}{f_c} = 1.1 \times 10^{-3}$, where $f_c = 14$ kHz is the band center frequency. While NOF1 is considered to be a benign channel, the NCS1 channel is more challenging due to its smaller coherence time.

The CP-OFDM system parameters used in this study are as follows. Center frequency and bandwidth are 14 kHz and 8 kHz, respectively, to match the frequency band of measured channel responses in NOF1 and NCS1. A guard interval of 32 ms is used, considering the power delay profiles for NOF1 and NCS1, beyond which the channel response is attenuated significantly. The symbol interval for the 1024 subcarrier system is 128 ms and the subcarrier spacing is 7.8125 Hz. As with the acoustic channel simulation model, pilots are used only in the first OFDM block for training and the algorithm switches to a decision directed mode thereafter. Symbols are drawn from the QPSK constellation to enable a performance comparison with the algorithm in [9].

Figure 17 shows the BER performance on NOF1 channel. The algorithm `DCDD (PID)` performs differentially coherent data detection, as in [9], after combining the PID outputs. The combining weights are computed using channel estimate obtained through sparse channel recovery. The algorithms in [3], [7] and the proposed algorithm perform almost equally

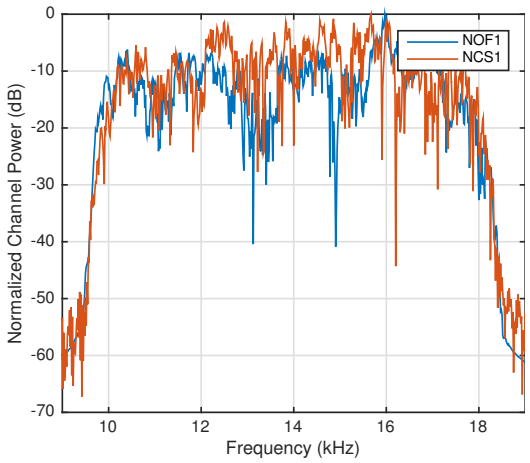

Figure 16. Instantaneous frequency response typical of the WATERMARK channels NOF1 and NCS1.

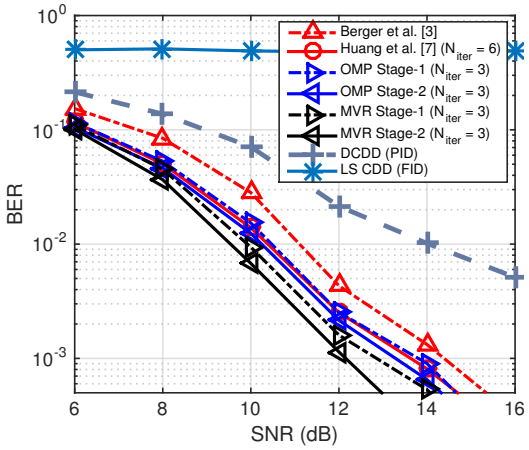

Figure 17. BER comparison of various channel estimation and data detection algorithms in the low Doppler spread WATERMARK channel NOF1.

well on this low Doppler spread channel. Figure 18 shows the BER performance on NCS1 channel. In this high Doppler spread channel, the proposed algorithm clearly outperforms the other algorithms.

VI. CONCLUSIONS

In this paper, we considered sparse channel estimation and data detection in a time-varying underwater acoustic channel for a CP-OFDM system. We used the measurements from the PID to track the time-variations within the OFDM symbol duration. We proposed a two-stage algorithm, where, in stage 1, we use the pilot-only measurements to estimate the channel and also detect the unknown data symbols. The MSE in the channel matrix estimation is reduced by iterating between the channel estimation and data detection for a fixed number of times. Thereafter, in stage 2, we use all the observations including data subcarriers to enhance the performance. We proposed a sparse channel recovery algorithm based on the minimum variance principle that bootstraps from the initial estimate provided by the OMP and refines this estimate.

We showed that using the output from the PID in the stage 1 indeed provides a good initial estimate of the channel matrix in a high Doppler spread scenario and is therefore key to the improved data detection performance of the proposed algorithm. We showed that the PID provides a larger number of effective measurements than the FID, and, hence, a lower MSE in the channel matrix estimate is achievable when using measurements from the PID. Our simulation results confirmed that the proposed two-stage algorithm significantly reduces the BER in time-varying channels. For sparse signal recovery, we considered the OMP and MVR algorithm. MVR provided better estimates of the channel than OMP. MVR exhibits a performance similar to SBL in just one pass through the algorithm, while the latter takes several iterations to converge. Results for the experimental channel data in WATERMARK reaffirm the strong performance of the proposed scheme in harsh channel conditions. In this work, we considered a grid based recovery of the Doppler and delay parameters; future work can consider gridless compressed sensing recovery methods and their performance. Extending the proposed approach

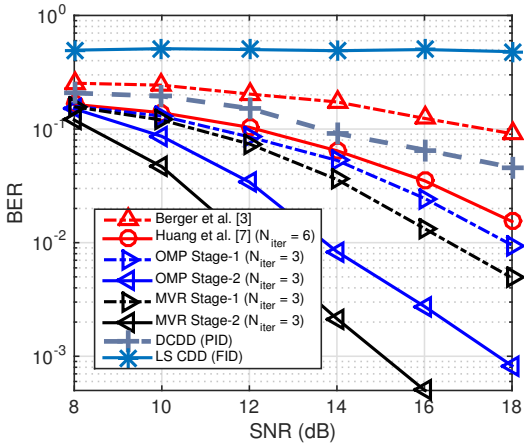

Figure 18. BER comparison of various channel estimation and data detection algorithms in the high Doppler spread ($b_{\max} \approx 10^{-3}$) WATERMARK channel NCS1.

to the case where the receiver is equipped with an array of hydrophones is also an interesting direction for future work.

APPENDIX

Proof of Lemma 1: The Khatri-Rao product $\mathbf{B} \otimes \mathbf{C} \in \mathbb{C}^{r \times pq}$ houses a submatrix (and another submatrix) whose rows are scaled versions of the rows of $\mathbf{B} \in \mathbb{C}^{r \times p}$ (respectively $\mathbf{C} \in \mathbb{C}^{r \times q}$) and hence its rank, and the k-rank, must be at least that of $\mathbf{B} \in \mathbb{C}^{r \times p}$ ($\mathbf{C} \in \mathbb{C}^{r \times q}$). The results follow.

Proof of Lemma 2: Let $\mathcal{R}$ denote a set of $r_k = k\text{-rank}\,(\mathbf{B} \otimes \mathbf{C}) + 1$ indices of the rows of $\mathbf{B} \otimes \mathbf{C}$ that are linearly dependent. It follows from the definition of k-rank that there exists a vector $\mathbf{d} \in \mathbb{C}^{r_k}$ with all entries nonzero such that $\mathbf{d}^T (\mathbf{B}_{\mathcal{R}} \otimes \mathbf{C}_{\mathcal{R}}) = \mathbf{0}$, where $\mathbf{B}_{\mathcal{R}} = \mathbf{I}_{\mathcal{R}}\mathbf{B}$ and $\mathbf{C}_{\mathcal{R}} = \mathbf{I}_{\mathcal{R}}\mathbf{C}$. Therefore, we have $\mathbf{B}_{\mathcal{R}}^T \mathbf{D} \mathbf{C}_{\mathcal{R}} = \mathbf{0} \in \mathbb{C}^{p \times q}$ where $\mathbf{D} = \text{diag}\,(\mathbf{d})$. $\mathbf{D}$ is non-singular since all entries of $\mathbf{d}$ are nonzero and hence, by the Sylvester inequality, $\mathbf{0} = \text{rank}\left(\mathbf{B}_{\mathcal{R}}^T \mathbf{D} \mathbf{C}_{\mathcal{R}}\right) \geq \text{rank}\,(\mathbf{B}_{\mathcal{R}}) + \text{rank}\,(\mathbf{C}_{\mathcal{R}}) - r_k \implies r_k \geq \text{rank}\,(\mathbf{B}_{\mathcal{R}}) + \text{rank}\,(\mathbf{C}_{\mathcal{R}})$. The rows of $\mathbf{B}_{\mathcal{R}}$ and $\mathbf{C}_{\mathcal{R}}$ are dependent by construction, and hence appending

more rows to these matrices cannot increase their k-rank: rank $(\mathbf{B}_{\mathcal{R}}) \geq k$-rank $(\mathbf{B}_{\mathcal{R}}) \geq k$-rank $(\mathbf{B})$ and rank $(\mathbf{C}_{\mathcal{R}}) \geq k$-rank $(\mathbf{C}_{\mathcal{R}}) \geq k$-rank $(\mathbf{C})$. Thus, we find $r_k = k$-rank $(\mathbf{B} \otimes \mathbf{C}) + 1 \geq k$-rank $(\mathbf{B}) + k$-rank $(\mathbf{C})$, from which the result follows.

Proof of Theorem 1: From (42) and (43), we get

$$\begin{aligned} M|S_P| &\geq \text{rank}\left(\tilde{\mathbf{\Gamma}}_{S_P}\right) \\ &\geq k\text{-rank}\left(\tilde{\mathbf{\Gamma}}_{S_P}\right) \\ &\geq \min\left(N_b + N_\tau - 1, M|S_P|\right) \\ &= M|S_P|, \end{aligned} \tag{58}$$

where the last inequality follows from Lemma 2, and hence

$$\text{rank}\left(\tilde{\mathbf{\Gamma}}_{S_P}\right) = M|S_P|. \tag{59}$$

Therefore, the row spaces of $\mathbf{I}_{S_P}\mathbf{\Gamma}^{(m)}, m = 1, \ldots, M$, form a virtually disjoint partition of the $M|S_P|$-dimensional row space of $\tilde{\mathbf{\Gamma}}_{S_P}$. Due to equation (25) and the condition in (44), the row space of $\mathbf{A}_{S_P}^{(m)}$ is equal to that of $\mathbf{I}_{S_P}\mathbf{\Gamma}^{(m)}$ for $m = 1, \ldots, M$. Hence, the row spaces of $\mathbf{A}_{S_P}^{(m)}, m = 1, \ldots, M$, also form a virtually disjoint partition of the $M|S_P|$-dimensional row space of $\tilde{\mathbf{\Gamma}}_{S_P}$. The result in (45) follows.

Proof of Theorem 2: First, we note that the channel matrix $\mathbf{H}$ can be related to the entries of the channel vector $\mathbf{x}$, defined in (12), by making use of (9). This allows us to express the channel matrix $\mathbf{H}$ as a function of the channel vector $\mathbf{x}$. The MSE of any unbiased estimator of $\mathbf{H}$ that makes use of the output from the PID, given by (13), cannot be better than the Cramér Rao bound (CRB) [32] of the subclass of unbiased estimators which know the true dictionary matrices $\mathbf{A}^{(m)}, m = 1, 2, \ldots, M$, i.e., where a genie provides the estimator with the knowledge of the data symbols. The CRB, given the true dictionary matrices in (13), is the same as the MSE of an oracle estimator that knows the support $\mathcal{S}$ of the channel vector $\mathbf{x}$ [33]. The result follows.

REFERENCES

[1] K. P. Arunkumar, C. R. Murthy, and V. Elango, "Joint sparse channel estimation and data detection for underwater acoustic channels using partial interval demodulation," in *2016 IEEE 17th International Workshop on Sig. Proc. Advances in Wireless Comm. (SPAWC)*, July 2016.

[2] P. Qarabaqi and M. Stojanovic, "Statistical characterization and computationally efficient modeling of a class of underwater acoustic communication channels," *IEEE J. Ocean. Eng.*, vol. 38, no. 4, pp. 701–717, Oct. 2012.

[3] C. R. Berger, S. Zhou, J. C. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 2941–2965, May 2011.

[4] Y. Huang, L. Wan, S. Zhou, Z. H. Wang, and J.-Z. Huang, "Comparison of sparse recovery algorithms for channel estimation in underwater acoustic OFDM with data-driven sparsity learning," *Elsevier Physical Communication*, vol. 13, no. 3, pp. 156–167, Dec. 2014.

[5] F. Qu, X. Nie, and W. Xu, "A two-stage approach for the estimation of doubly spread acoustic channels," *IEEE J. Ocean. Eng.*, vol. 40, no. 1, pp. 131–143, Jan. 2015.

[6] W. Li and J. Preisig, "Estimation of rapidly time-varying sparse channels," *IEEE J. Ocean. Eng.*, vol. 32, no. 4, pp. 927–939, Oct. 2007.

[7] J. Z. Huang, S. Zhou, J. Huang, C. R. Berger, and P. Willett, "Progressive inter-carrier interference equalization for OFDM transmission over time-varying underwater acoustic channels," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 8, pp. 1524–1536, Dec. 2011.

[8] S. Yerramalli, M. Stojanovic, and U. Mitra, "Partial FFT demodulation: A detection method for highly Doppler distorted OFDM systems," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5906–5918, Nov. 2012.

[9] Y. M. Aval and M. Stojanovic, "Differentially coherent multichannel detection of acoustic OFDM signals," *IEEE J. Ocean. Eng.*, vol. 40, no. 2, pp. 251–268, Apr. 2015.

[10] C. R. Berger, J. P. Gomes, and J. M. F. Moura, "Sea-trial results for cyclic-prefix OFDM with long symbol duration," in *Proc. MTS/IEEE OCEANS Conf.*, June 2011.

[11] B. Li, S. Zhou, M. Stojanovic, L. Freitag, and P. Willett, "Multicarrier communication over underwater acoustic channels with nonuniform Doppler shifts," *IEEE J. Ocean. Eng.*, vol. 33, no. 2, pp. 1638–1649, Apr. 2008.

[12] S. Yerramalli and U. Mitra, "Optimal resampling of OFDM signals for multiscale–multilag underwater acoustic channels," *IEEE J. Ocean. Eng.*, vol. 36, no. 1, pp. 126–138, Jan. 2011.

[13] C. Liu, Y. V. Zakharov, and T. Chen, "Doubly selective underwater acoustic channel model for a moving transmitter/receiver," *IEEE Trans. Veh. Technol.*, vol. 61, no. 13, pp. 126–138, Mar. 2012.

[14] S. M. Kay, "Modern spectrum estimation: Theory and application." Prentice Hall Signal Processing, 1999.

[15] D. Vu, M. Xue, X. Tan, and J. Li, "A Bayesian approach to SAR imaging," *Digit. Signal Process.*, vol. 23, no. 3, pp. 852–858, May 2013.

[16] R. Hunger, "Floating point operations in matrix-vector calculus," Technical Report, Version 3, https://mediatum.ub.tum.de/, Associate Institute for Signal Processing, Munich Univ. of Technol., Germany, 2007.

[17] B. L. Sturm and M. G. Christensen, "Comparison of orthogonal matching pursuit implementations," *20th European Signal Processing Conference (EUSIPCO 2012)*, pp. 220–224, Aug. 2012.

[18] P. R. Halmos, "Finite dimensional vector spaces." Springer, 1958.

[19] H. Wan, R.-R. Chen, J. W. Choi, A. Singer, J. Preisig, and B. Farhang-Boroujeny, "Markov chain Monte Carlo detection for frequency-selective channels using list channel estimates," *IEEE J. Sel. Topics in Signal Process.*, vol. 8, no. 5, pp. 1537–1547, Dec. 2011.

[20] Q. Lu, Y. Huang, Z.-H. Wang, and S. Zhou, "Characterization and receiver design for underwater acoustic channels with large doppler spread," in *Proc. of IEEE/MTS OCEANS conference*, Oct. 2015.

[21] S. Zhou and Z. Wang, "OFDM for underwater acoustic communications." Wiley, 2014.

[22] P. A. van Walree, R. Otnes, and T. Jenserud, "The Watermark manual and user's guide version 1.0," in *Norwegian Defense Research Establishment (FFI)*, Nov. 2016.

[23] R. Otnes and T. H. Eggen, "Underwater acoustic communications: Long-term test of turbo equalization in shallow water," *IEEE J. Ocean. Eng.*, vol. 33, no. 3, pp. 321–334, July 2008.

[24] M. Elad, "Sparse and redundant representations: From theory to applications in signal and image processing." Springer, 2010.

[25] C. R. Berger, J. P. Gomes, and J. M. F. Moura, "Study of pilot designs for cyclic-prefix OFDM on time-varying and sparse underwater acoustic channels," in *Proc. MTS/IEEE OCEANS Conf.*, June 2011.

[26] C. R. Berger, J. Huang, and J. M. F. Moura, "Study of pilot overhead for iterative OFDM receivers on time-varying and sparse underwater acoustic channels," in *Proc. MTS/IEEE OCEANS Conf.*, Sept. 2011.

[27] M. Stojanovic. Acoustic channel simulator. [Online]. Available: http://millitsa.coe.neu.edu/?q=research/simulator

[28] J.-Z. Huang, S. Zhou, and Z.-H. Wang, "Robust initialization with reduced pilot overhead for progressive underwater acoustic OFDM receivers," in *Proc. of MILCOM Conf.*, Nov. 2011, pp. 406–411.

[29] Z.-H. Wang, S. Zhou, J. Preisig, K. R. Pattipati, and P. Willett, "Clustered adaptation for estimation of time-varying underwater acoustic channels," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 3079–3091, June 2012.

[30] P. A. van Walree, F. Socheleau, R. Otnes, and T. Jenserud, "The watermark benchmark for underwater acoustic modulation schemes," *IEEE J. Ocean. Eng.*, vol. 42, no. 4, pp. 1007–1018, Oct. 2017.

[31] R. Otnes, P. A. van Walree, and T. Jenserud, "Validation of replay-based underwater acoustic communication channel simulation," *IEEE J. Ocean. Eng.*, vol. 38, no. 4, pp. 689–700, Oct. 2013.

[32] H. L. Van Trees, K. L. Bell, and Z. Tian, "Detection, estimation and modulation theory: Part I." John Wiley and Sons, 2013.

[33] Z. Ben-Haim and Y. C. Eldar, "The Cramér-Rao bound for estimating a sparse parameter vector," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3591–3603, June 2010.