

Equalization for Multi-scale Multi-lag OFDM Channels

Zijian Tang*[†], Rob Remis[†], Tao Xu[†], Geert Leus[†] and Magnus Lundberg Nordenvaad[‡]

*TNO, P.O. Box 96864, 2509 JG the Hague, the Netherlands

Email: zijian.tang@tno.nl

[†]Delft University of Technology, Mekelweg 4, 2628 CD Delft, the Netherlands

Email: {z.tang, r.f.remis, t.xu, g.j.t.leus}@tudelft.nl

[‡]Swedish Defence Research Agency (FOI), SE- 164 90 Stockholm, Sweden

Email: maglun@foi.se

Abstract—We consider an orthogonal frequency-division multiplexing (OFDM) transmission scheme over wideband underwater acoustic channels, where the propagation paths can experience distinct Doppler effects (manifested in signal scales) and time of arrivals (manifested in lags). We capture such an effect in this paper with a multi-scale multi-lag (MSML) model, and show that the resulting frequency-domain MSML-OFDM channel is subject to inter-carrier interference (ICI), whose amount differs per subcarrier. The corresponding channel matrix can still be approximated as highly sparse, but lacks a specific structure that can optimally be exploited by those low-complexity equalizers proposed for narrowband channels. In this paper, we propose to use the conjugate gradient (CG) algorithm to equalize the channel iteratively. The suitability of the preconditioning technique, that often accompanies the CG to accelerate the convergence, is discussed for the MSML-OFDM channel. We show that in order for the preconditioner to function properly, optimal resampling is indispensable.

Index Terms—OFDM, wideband, multi-lag, multi-scale, Doppler, ICI, conjugate gradient, preconditioning, resampling

I. INTRODUCTION

Orthogonal frequency division multiplexing (OFDM) gains most of its popularity thanks to its ability to transform frequency-selective channels into subcarriers without mutual interference. This is especially attractive for channels with a long memory. For this reason, OFDM is also considered as a promising technique for underwater acoustic communications, where the channel is typically lengthy due to a slow sound speed. However, underwater acoustic channels are very sensitive to Doppler resulting from motion of the vessels as well as the waves, which imposes a huge challenge on the receiver design.

In this paper, we will model the underwater acoustic channel with a multi-scale multi-lag (MSML) model [1], [2]. In a nutshell, the MSML model assumes that the transmitted signal propagates via diverse paths, and impinges on the receiver at different time instances with different incident angles. The former gives rise to the multi-lag effect; the latter implies that the radial velocity experienced from diverse paths will be disparate. In a wideband system, which is often the case for underwater acoustic communications, the Doppler spread manifests itself as a dilation or compression of the transmitted

waveforms, and therefore we have to deal with a multi-scale effect. Corresponding to an MSML channel, the discrete channel matrix in the frequency domain will not be diagonal as in the frequency-selective channel case, but has non-zero entries everywhere. Moreover, it is possible that the major channel energy is not located on the main diagonal of the channel matrix, but shifted away depending on the channel scales as well as the sampling rate adopted at the receiver.

Equalizing such a channel can be costly. For instance, a least-squares (LS) equalizer requires an inversion of the channel matrix, which incurs a complexity that is cubic in the number of subcarriers. In a narrowband system, where a similar situation of time-varying OFDM channel occurs, the channel matrix is assumed to be banded to enable a low-complexity equalizer [3]–[5]. In a wideband system, a sparse approximation might be more suitable to simplify the channel matrix, and accordingly, we propose to use the conjugate gradient (CG) algorithm for equalization [6], [7], which can benefit maximally from the channel sparsity. Compared to the LS equalizer which inverts the channel matrix directly, CG is an iterative approach, where the result yielded by each iteration is constrained in the Krylov subspace. This implies that the performance of CG can be less sensitive to the condition of the channel matrix in contrast to the LS method. On the other hand, the convergence rate of CG is inversely proportional to the channel matrix condition number. This is especially of significance if a truncated CG is to be used in practice, which halts the algorithm after a limited number of iterations in order to further reduce the complexity. Therefore, it is still desired by CG that the channel matrix is well conditioned to ensure a fast convergence. To this end, preconditioning techniques can be invoked which enforce the eigenvalues to cluster around 1 [8]. At the same time, we desire that the design and implementation of the preconditioner should be simple enough such that the overall complexity stays low. A common practice is to restrict the preconditioner to be a diagonal matrix such as in [9], whose diagonal entries can be designed following the steps given in [10]. However, in the case where the major channel energy is located on the off-diagonals of the channel matrix, we can show that a diagonal preconditioner will render a negative effect on the channel

matrix spectrum by clustering the eigenvalues around 0 instead of 1. The condition number gets consequently enlarged, which reduces the convergence rate of CG instead of increasing it.

To ensure that the preconditioner works properly, we need to find a way to “push” the channel energy back to the main diagonal. Recall that in a single-scale multi-lag (SSML) channel case [11], [12], we can achieve a diagonal channel matrix by using resampling at the receiver complying with the channel scale, and apply a proper phase shift in the time domain. For an MSML channel, we can adopt the same method by optimizing the resampling rate such that the channel energy on the main diagonal will be maximized. Note that optimal resampling is also considered in [13], but uses different metrics for optimization.

The remainder of the paper is organized as follows. In Section II, the data model of the OFDM channel matrix is presented and an analysis of the inter-carrier interference (ICI) is given. CG is proposed in Section II for channel equalization where the focus is laid on the design of a preconditioner. We discuss optimal resampling in Section IV, and simulation results are given in Section V.

Notation: We use upper (lower) bold face letters to denote matrices (column vectors). $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$ represent conjugate, transpose and complex conjugate transpose (Hermitian), respectively. $[\mathbf{x}]_p$ indicates the p th element of the vector \mathbf{x} and $[\mathbf{X}]_{p,q}$ indicates the (p, q) th entry of the matrix \mathbf{X} . $\text{diag}\{\mathbf{x}\}$ is used to denote a diagonal matrix with \mathbf{x} on the diagonal; \odot represents the Hadamard product. \mathbf{I}_N stands for the $N \times N$ identity matrix and \mathbf{e}_k is reserved for the k th column of \mathbf{I}_N ; $\mathbf{1}_{M \times N}$ stands for an $M \times N$ all-one matrix. Finally, we use $\|\mathbf{X}\|_{\text{Fro}}$ and $\|\mathbf{X}\|_2$ to denote the Frobenius norm and the ℓ_2 -norm of the matrix \mathbf{X} , respectively.

II. DATA MODEL

A. Input/Output Relationship

We consider an OFDM system with K subcarriers. At the transmitter, the baseband time-domain signal $s(t)$ can be expressed as

$$s(t) = \sum_p \sum_{k=0}^{K-1} b_{k,p} e^{j\frac{2\pi}{KT}k(t-T_{\text{CP}})} u[t - p(KT + T_{\text{CP}})], \quad (1)$$

where $b_{k,p}$ stands for the p th information symbol modulated onto the k th subcarrier; T is the reciprocal of the signal bandwidth and T_{CP} denotes the length of a cyclic prefix (CP). Hence, the OFDM symbol has a duration of $KT + T_{\text{CP}}$. $u(t)$ is a windowing function, which is, for the sake of simplicity, defined to be a rectangular window within $[0, KT + T_{\text{CP}}]$ and zero outside the range. Without loss of generality, in the sequel we will focus just on a single OFDM symbol case, e.g., $p = 0$, and therefore omit the index p . Accordingly, $s(t)$ can be more concisely expressed as

$$s(t) = \sum_{k=0}^{K-1} b_k e^{j\frac{2\pi}{KT}k(t-T_{\text{CP}})} u(t). \quad (2)$$

The baseband signal is first modulated onto the carrier frequency f_c resulting in

$$x(t) = s(t)e^{j2\pi f_c t}, \quad (3)$$

which is afterwards sent over the channel. The channel impulse response $h(t, \tau)$, comprising the joint effects of the transmit and receive filter as well as the actual propagation channel, is approximated in this paper by a discrete multi-path model:

$$h(t, \tau) \approx \sum_{l=0}^{L-1} h_l(t) \delta(\tau - \tau_l(t)), \quad (4)$$

where $\delta(t)$ is the Kronecker delta function; L represents the total number of paths we consider; $h_l(t)$ stands for the time-varying gain of the l th path, and $\tau_l(t)$ the time-varying delay, which can be expressed as

$$\tau_l(t) = \alpha_l(t)(\tau_l - t), \quad (5)$$

with τ_l being a constant corresponding to propagation delay, and $\alpha_l(t)$ denoting the time dilation/compression (scale) effect due to Doppler on the l th path. Suppose that the instant radial velocity of the transmitter corresponding to the l th path is $v_l(t)$, and the medium speed is c with $v_l(t) \ll c$. Then it follows that $\alpha_l(t) \approx 1 + \frac{2v_l(t)}{c}$. It is noteworthy that Dirichlet kernele the signal waveforms impinge on the receiver at different angles, $v_l \neq v_{l'}$ for $l \neq l'$, and therefore, the scaling effects corresponding to different paths are distinctive. This fact deviates from the assumption adopted in many papers that a single scale is universally present in all the paths [11], [12]. In this paper, we assume that both $h_l(t)$ and $\alpha_l(t)$ are constant during the transmission of the OFDM symbol. As a result, the received signal in the noiseless case, denoted as $r_c(t)$, can be expressed as

$$r_c(t) = \sum_{l=0}^{L-1} h_l \sqrt{\alpha_l} x[\alpha_l(t - \tau_l)]. \quad (6)$$

When demodulated to baseband, the CP is stripped off from the received signal $r_c(t)e^{-j2\pi f_c t}$ such that the resulting signal $r(t)$ can be expressed as

$$r(t) = \sum_{l=0}^{L-1} h_l \sqrt{\alpha_l} \sum_{k=0}^{K-1} b_k e^{j\frac{2\pi}{K}k\frac{\alpha_l(t-\tau_l)}{T}} e^{j2\pi f_c[\alpha_l(t-\tau_l)]} e^{-j2\pi f_c t}. \quad (7)$$

For a multi-scale channel, the sampling rate at the receiver is not straightforward to determine. For the moment, let us assume that the receiver adopts a sampling rate of $\frac{T}{\beta}$ with β being a positive number between $\min(\alpha_l)$ and $\max(\alpha_l)$, and define the signal $r_n^{(\beta)}$ as the sample obtained at the n th time interval equal to $r(n\frac{T}{\beta})$. Accordingly, we can write

$$\begin{aligned} r_n^{(\beta)} &= \sum_{l=0}^{L-1} h_l \sqrt{\frac{\alpha_l}{\beta}} \sum_{k=0}^{K-1} b_k e^{j\frac{2\pi}{K}\frac{\alpha_l}{\beta}kn} e^{j\frac{2\pi}{K}k\frac{\alpha_l\tau_l}{T}} e^{j2\pi f_c\frac{\alpha_l-1}{\beta}nT} e^{-j2\pi f_c\alpha_l\tau_l}, \\ &= \sum_{l=0}^{L-1} g_n^{(\alpha_l-1, \beta)} h_l^{(\beta)} \sum_{k=0}^{K-1} b_k e^{j\frac{2\pi}{K}\frac{\alpha_l}{\beta}kn} e^{-j\frac{2\pi}{K}k\rho_l}. \end{aligned} \quad (8)$$

In the above, the following definitions have been introduced

$$\begin{aligned} g_n^{(\alpha,\beta)} &:= e^{j2\pi f_c \frac{\alpha}{\beta} nT}, \\ \rho_l &:= \frac{\alpha_l \tau_l}{T}, \\ h_l^{(\beta)} &:= h_l \sqrt{\frac{\alpha_l}{\beta}} e^{-j2\pi f_c \rho_l T}. \end{aligned}$$

If we stack these time-domain samples $r_n^{(\beta)}$ for $n = 0, \dots, K-1$ into a vector $\mathbf{r}_T^{(\beta)} := [r_0^{(\beta)}, \dots, r_{K-1}^{(\beta)}]^T$, it will admit the following matrix/vector expression as

$$\mathbf{r}_T^{(\beta)} = \sum_{l=0}^{L-1} h_l^{(\beta)} \text{diag}\{\mathbf{g}^{(\alpha_l-1,\beta)}\} \mathbf{F}_{K,\frac{\alpha_l}{\beta}}^H \text{diag}\{\boldsymbol{\lambda}_l\} \mathbf{b}, \quad (9)$$

where \mathbf{b} and $\mathbf{g}^{(\alpha,\beta)}$ are similarly defined as $\mathbf{r}_T^{(\beta)}$; $\boldsymbol{\lambda}_l := \sqrt{K}[1, \dots, e^{j\frac{2\pi}{K}(K-1)\rho_l}]^T$, and $\mathbf{F}_{K,\alpha}$ stands for the fractional discrete Fourier transform (DFT) matrix

$$\mathbf{F}_{K,\alpha} := \frac{1}{\sqrt{K}} \begin{bmatrix} 1 & e^{-j\frac{2\pi}{K}\alpha 0 \cdot 1} & \dots & e^{-j\frac{2\pi}{K}\alpha 0 \cdot (K-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-j\frac{2\pi}{K}\alpha (K-1) \cdot 1} & \dots & e^{-j\frac{2\pi}{K}\alpha (K-1) \cdot (K-1)} \end{bmatrix} \quad (10)$$

Note that $\mathbf{F}_{K,\alpha}$ reduces to the normal DFT matrix \mathbf{F}_K when $\alpha = 1$.

Normally speaking, the channel equalization of an OFDM system is implemented in the frequency domain. To this end, the received signal $\mathbf{r}_T^{(\beta)}$ is first transformed into the frequency domain by means of DFT yielding

$$\mathbf{r}_F^{(\beta)} := \mathbf{F}_K \mathbf{r}_T^{(\beta)} = \mathcal{H}_F^{(\beta)} \mathbf{b}, \quad (11)$$

where

$$\mathcal{H}_F^{(\beta)} := \sum_{l=0}^{L-1} h_l^{(\beta)} \underbrace{\mathbf{F}_K \text{diag}\{\mathbf{g}^{(\alpha_l-1,\beta)}\} \mathbf{F}_{K,\frac{\alpha_l}{\beta}}^H}_{\mathcal{H}_{F,l}^{(\beta)}} \text{diag}\{\boldsymbol{\lambda}_l\}. \quad (12)$$

In the above, the symbol $\mathcal{H}_{F,l}^{(\beta)}$ is introduced to denote the frequency-domain channel matrix corresponding to the l th path resulting from a sampling rate of β at the receiver.

B. Structure of $\mathcal{H}_F^{(\beta)}$

To study the structure of $\mathcal{H}_F^{(\beta)}$, it is instrumental to first focus on the structure of $\mathcal{H}_{F,l}^{(\beta)}$ whose (m, k) th entry can be expressed as

$$\begin{aligned} [\mathcal{H}_{F,l}^{(\beta)}]_{m,k} &= \sum_{n=0}^{K-1} e^{-j\frac{2\pi}{K}mn} e^{j2\pi f_c \frac{\alpha_l-1}{\beta} nT} e^{j\frac{2\pi}{K} \frac{\alpha_l}{\beta} nk}, \\ &= K e^{j\frac{\pi}{K}(K-1)(m - \frac{\alpha_l}{\beta}k - \frac{\alpha_l-1}{\beta}D)} \times \\ &\quad \frac{\text{sinc}(\pi(m - \frac{\alpha_l}{\beta}k - \frac{\alpha_l-1}{\beta}D))}{\text{sinc}(\pi(m - \frac{\alpha_l}{\beta}k - \frac{\alpha_l-1}{\beta}D)/K)}, \end{aligned} \quad (13)$$

where in the second equality $D := f_c TK$. Obviously, in case of time-invariant channels with $\alpha_l = \beta = 1$, we have $[\mathcal{H}_{F,l}^{(\beta)}]_{m,k} = \delta(m-k)$, which suggests that $\mathcal{H}_{F,l}^{(\beta)}$ is a diagonal matrix. For time-varying channels with $\alpha_l \neq 1$, $[\mathcal{H}_{F,l}^{(\beta)}]_{m,k}$

becomes a Dirichlet kernel. Most notably, the mainlobe of this Dirichlet kernel function is not necessarily located on the main diagonal of $\mathcal{H}_{F,l}^{(\beta)}$, but could shift away to off-diagonal entries. To realize this, we can derive from (13) that $|\mathcal{H}_{F,l}^{(\beta)}]_{m,k}| = \max|\mathcal{H}_{F,l}^{(\beta)}]_{m,k}|$ if

$$\text{mod}[m - \frac{\alpha_l}{\beta}k - \frac{\alpha_l-1}{\beta}D]_K = 0. \quad (14)$$

This suggests that a majority of the energy transmitted through the k th subcarrier could be received on a different subcarrier, whose difference depends on k . We will refer to this phenomenon, which is unique to wideband time-varying channels, as subcarrier offset. Because we can only deal with integer subcarrier indexes, to study the effect of subcarrier offset in practice, let us introduce the following symbol

$$\Delta_{k,l}^{(\beta)} = \lceil \text{mod}[\frac{\alpha_l}{\beta}k + \frac{\alpha_l-1}{\beta}D]_K \rceil - k, \quad (15)$$

where $\lceil \cdot \rceil$ denotes the ceiling operation. (15) suggests that the ICI depends on two factors: 1) it has a fixed part determined by $\frac{\alpha_l-1}{\beta}D$, which is more often known as the normalized Doppler shift for a narrowband time-varying channel; and 2) a variable part $\frac{\alpha_l}{\beta}k$ linear to the subcarrier index k . Compared to narrowband systems where the ICI results primarily from the normalized Doppler shift, and is thus on average similar for all the subcarriers, the variable part in wideband systems causes the ICI to be more severe with an increase of the subcarrier index k .

Because of the Dirichlet kernel, (13) also implies that the energy leaks into subcarriers adjacent to $k + \Delta_{k,l}^{(\beta)}$. On the other side, the height and width of the main-lobe or the side-lobes in (13) are constant and hence independent of α_l , β and D . To be more specific, the mainlobe has a peak amplitude equal to K , and the width between the first zero-crossings around the peak is always equal to two.

Based on the afore-mentioned observations, we summarize the following properties regarding the structure of $\mathcal{H}_{F,l}^{(\beta)}$.

Property 1: $\mathcal{H}_{F,l}^{(\beta)}$ is a sparse matrix with most of its significant entries clustered around an axis. This axis lies above the main diagonal if $\frac{\alpha_l}{\beta}$ is larger than one, or below the main diagonal otherwise, and it deviates farther away from the main diagonal with an increase of the subcarrier index.

Now that the actual frequency-domain channel matrix $\mathcal{H}_F^{(\beta)}$ is a weighted superposition of the matrices $\mathcal{H}_{F,l}^{(\beta)}$ from individual paths, the following properties concerning the structure of $\mathcal{H}_F^{(\beta)}$ are in order.

Property 2: It is still reasonable to approximate $\mathcal{H}_F^{(\beta)}$ as a banded matrix. However, it differs from the ‘‘normal’’ banded matrix in that 1) the direction of the band is not in parallel to the main diagonal; 2) the bandwidth is not constant but increases with the subcarrier index, implying that the band has a trapezoidal shape.

In Fig. 1, we give a numerical example of $\mathcal{H}_F^{(\beta)}$ with system parameters $K = 128$, $D = 256$, and channel parameters given in Table I; the sampling rate is chosen to be identical to that at the transmitter, i.e., $\beta = 1$.

	α_l	h_l	$\tau_l(m.s)$
$l = 0$	0.9866	0.7806	97.09
$l = 1$	0.9863	0.2869	27.50
$l = 2$	0.9866	0.5247	39.53
$l = 3$	0.9865	-0.1815	35.03

TABLE I
CHANNEL PARAMETERS USED IN FIG. 1.

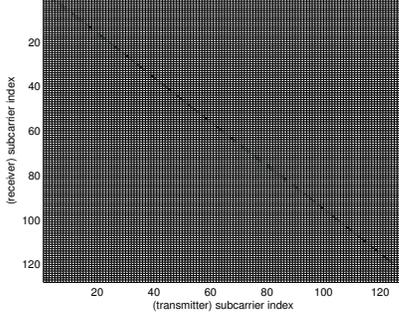


Fig. 1. An example of the frequency-domain channel matrix $\mathcal{H}_F^{(\beta)}$.

III. DESIGN OF CHANNEL EQUALIZERS

A. Low-Complexity Equalization

As evident from the previous section, due to the multiple-scale channel property, it is theoretically not possible to attain a diagonal frequency-domain channel matrix by means of resampling as claimed in e.g., [12]. This implies that for channel equalization, a direct inversion of $\mathcal{H}_F^{(\beta)}$ will inflict a complexity of $\mathcal{O}(K^3)$, which is obviously not desired for a practical system.

To facilitate a cheaper equalizer, we leverage Property 2, and approximate $\mathcal{H}_F^{(\beta)}$ as a truly sparse matrix $\tilde{\mathcal{H}}_F^{(\beta)}$. The approximation can be realized in two ways:

Method 1: Suppose \mathbf{Q}_B denotes a circulant banded matrix with bandwidth $2B + 1$, i.e., \mathbf{Q}_B has ones on its main diagonal, B super-diagonals and B sub-diagonals, and zeros on the remaining entries. Then we approximate $\mathcal{H}_F^{(\beta)}$ by $\tilde{\mathcal{H}}_F^{(\beta)} = \mathbf{Q}_B \odot \mathcal{H}_F^{(\beta)}$ where \odot stands for the Hadamard product, and B is the smallest integer for which

$$\frac{\|\tilde{\mathcal{H}}_F^{(\beta)} - \mathcal{H}_F^{(\beta)}\|_{\text{Fro}}^2}{\|\mathcal{H}_F^{(\beta)}\|_{\text{Fro}}^2} \leq \eta, \quad (16)$$

where η is a predefined percentage.

A second approach can be described as follows.

Method 2: For each row (or column) of $\mathcal{H}_F^{(\beta)}$, let the entries be sorted in an ascending order in terms of their power. Then we obtain a new matrix $\tilde{\mathcal{H}}_F^{(\beta)}$ by replacing the first $\eta\%$ of the entries with zero, where η is a predefined threshold.

Method 1 is adopted by many works [3]–[5], [14] for narrowband time-varying OFDM systems, for which the corresponding channel matrix is characterized with a relatively

small ICI that is at the same time located in the vicinity of the main diagonal. In that case, Method 1 renders a very tight approximation. However, for a multi-scale multi-lag channel matrix which is characterized by Property 2, Method 1 will retain unnecessarily many small entries by enforcing a strictly banded approximation on $\mathcal{H}_F^{(\beta)}$. As a comparison, Method 2 focuses on the significant entries per row(column), and hence is able to yield a much more sparse $\tilde{\mathcal{H}}_F^{(\beta)}$.

The banded structure of $\tilde{\mathcal{H}}_F^{(\beta)}$ resulting from Method 1 makes a lot of direct inversion methods applicable that have a complexity of $\mathcal{O}(KB^2)$, whereas $\tilde{\mathcal{H}}_F^{(\beta)}$ resulting from Method 2 often lacks a specific structure. For this reason, an alternative is to invert $\tilde{\mathcal{H}}_F^{(\beta)}$ iteratively using e.g., the conjugate gradient (CG) algorithm, which does not rely on a specific structure of the channel matrix, and therefore a sparse channel matrix approximation using Method 2 can also be applied. Moreover, the data estimates yielded by CG are always constrained in the Krylov subspace, making its performance less susceptible to the spectral distribution of $\mathcal{H}_F^{(\beta)}$. On the other hand, a truncated CG, which halts the algorithm after a limited number of iterations, is often desired in practice to reduce the complexity. This requires though a fast convergence of CG, whose convergence rate depends on the condition number of $\mathcal{H}_F^{(\beta)}$ as argued in e.g., [6], [7]. We will address this issue in the next subsection.

B. To precondition or Not?

It is well-known that the convergence of the CG can be accelerated by applying preconditioning on $\mathcal{H}_F^{(\beta)}$ [8], [10]. With \mathbf{P} denoting such a preconditioner, the I/O relationship given in (11) in the noiseless case can be rewritten as

$$\begin{aligned} \mathbf{r}_F &= \mathcal{H}_F^{(\beta)} \mathbf{b}, \\ &= \mathcal{H}_F^{(\beta)} \mathbf{P} \mathbf{P}^{-1} \mathbf{b}, \end{aligned} \quad (17)$$

from which an estimate of $\hat{\mathbf{x}} = \mathbf{P}^{-1} \mathbf{b}$ is first obtained by applying CG on the preconditioned matrix $\mathcal{H}_F^{(\beta)} \mathbf{P}$. Afterwards, $\hat{\mathbf{b}} = \mathbf{P} \hat{\mathbf{x}}$ is computed to obtain the final data estimates.

One of the approaches to design \mathbf{P} is to make the matrix after preconditioning as close as possible to the identity matrix in an ℓ_2 -norm sense. This approach is preferred for a hardware implementation by allowing for a parallel architecture, but a cost function expressed in an ℓ_2 -norm is often difficult to solve. As a relaxation, the Frobenius norm is used, which clusters most of the eigenvalues $\mathcal{H}_F^{(\beta)} \mathbf{P}$ around 1 with the exception of a few outliers [10]. Further, observing that the design of \mathbf{P} itself as well as the operation of $\mathcal{H}_F^{(\beta)} \mathbf{P}$ also inflicts an additional complexity, a common approach is to impose a sparse structure on \mathbf{P} , e.g., diagonal with $\mathbf{P} = \text{diag}\{[p_0, \dots, p_{K-1}]^T\}$ [10]. The above arguments can be mathematically formulated by the following optimization problem function

$$\arg \min_{p_0, \dots, p_{K-1}} \|\mathcal{H}_F^{(\beta)} \text{diag}\{[p_0, \dots, p_{K-1}]^T\} - \mathbf{I}_K\|_{\text{Fro}}^2, \quad (18)$$

from which we can solve each p_k separately as

$$\begin{aligned} p_{k,\text{opt}} &= \arg \min_{p_k} \|\mathcal{H}_F^{(\beta)} \mathbf{e}_k p_k - \mathbf{e}_k\|_{\text{Fro}}^2, \\ &= \frac{[\mathcal{H}_F^{(\beta)}]_{k,k}^*}{\|\mathcal{H}_F^{(\beta)} \mathbf{e}_k\|_{\text{Fro}}^2}, \end{aligned} \quad (19)$$

where \mathbf{e}_k stands for an all-zero vector except for its n th entry which equals 1.

One problem of the preconditioner designed by (18) is that the eigenvalues will in many cases cluster around 0 instead of 1, with the consequence that the condition number of the preconditioned channel matrix can considerably increase. To realize this, let us use ϵ_1 to denote the lowest upper-bound of the residual from (19) such that

$$\|\mathcal{H}_F^{(\beta)} \mathbf{e}_k p_k - \mathbf{e}_k\|_{\text{Fro}}^2 < \epsilon_1, \quad (20)$$

holds for $k = 0, \dots, K-1$. In [10], it is shown that the eigenvalues of $\mathcal{H}_F^{(\beta)} \text{diag}\{[p_{0,\text{opt}}, \dots, p_{K-1,\text{opt}}]^T\}$, denoted by λ_k , should satisfy

$$\sum_{k=0}^{K-1} |1 - \lambda_k|^2 \leq K \epsilon_1^2. \quad (21)$$

Based on the above, it is shown in [10] that all λ_k 's lie inside a disk of $\sqrt{K} \epsilon_1$ centered around 1. On the other hand, following similar steps as in [10], let us use ϵ_0 to denote the lowest upper-bound of the residual $\|\mathcal{H}_F^{(\beta)} \mathbf{e}_k p_k\|_{\text{Fro}}^2$ such that

$$\|\mathcal{H}_F^{(\beta)} \mathbf{e}_k p_{k,\text{opt}}\|_{\text{Fro}}^2 < \epsilon_0 \quad (22)$$

holds for $k = 0, \dots, K-1$. Further, let $\mathbf{U}\mathbf{W}\mathbf{U}^H$ be a Schur decomposition of $\mathcal{H}_F^{(\beta)} \text{diag}\{[p_{0,\text{opt}}, \dots, p_{K-1,\text{opt}}]^T\}$ such that $\mathbf{U}\mathbf{U}^H = \mathbf{I}$, and the diagonal of \mathbf{W} equals $[\lambda_0, \dots, \lambda_{K-1}]^T$. Then

$$\begin{aligned} \sum_{k=0}^{K-1} |\lambda_k|^2 &= \|\text{diag}\{\mathbf{W}\}\|_2^2 \leq \|\mathbf{W}\|_{\text{Fro}}^2 \\ &= \|\mathcal{H}_F^{(\beta)} \text{diag}\{[p_{0,\text{opt}}, \dots, p_{K-1,\text{opt}}]^T\}\|_{\text{Fro}}^2 \leq K \epsilon_0^2, \end{aligned}$$

which implies that all λ_k 's at the same time lie inside a disk of radius $\sqrt{K} \epsilon_0$ centered around 0. Obviously, if $\epsilon_0 < \epsilon_1$, then minimizing $\|\mathcal{H}_F^{(\beta)} \mathbf{P} - \mathbf{I}_K\|_{\text{Fro}}^2$ will at the same time minimize the Frobenius norm of $\mathcal{H}_F^{(\beta)} \mathbf{P}$ itself, making the eigenvalues more clustered around 0 rather than 1. With $p_{k,\text{opt}}$ obtained from (19), we can show that

$$\epsilon_1 = \max_k \frac{|\sum_{m=k}^{K-1} [\mathcal{H}_F^{(\beta)}]_{m,k}|^2 - |[\mathcal{H}_F^{(\beta)}]_{k,k}|^2}{\sum_{m=0}^{K-1} |[\mathcal{H}_F^{(\beta)}]_{m,k}|^2}, \quad (23)$$

and

$$\epsilon_0 = \max_k \frac{|[\mathcal{H}_F^{(\beta)}]_{k,k}|^2}{\sum_{m=0}^{K-1} |[\mathcal{H}_F^{(\beta)}]_{m,k}|^2}. \quad (24)$$

Obviously, if

$$|[\mathcal{H}_F^{(\beta)}]_{k,k}|^2 < \sum_{m=0, m \neq k}^{K-1} |[\mathcal{H}_F^{(\beta)}]_{m,k}|^2, \quad (25)$$

holds for $k = 0, \dots, K-1$, then a diagonal preconditioner will cluster the eigenvalues in a “wrong” area. Note that (25) arises when the sum of the off-diagonal power in each column is higher than the power on the diagonal. Such a situation could occur in multi-scale channels since significant channel power could be located on off-diagonal entries as we argued in the previous section where an example of such a channel matrix is given in Fig. 1. In the left plot of Fig. 2, the eigenvalues of this channel matrix with and without preconditioning are displayed on a complex plane, where one can see that preconditioning indeed clusters the eigenvalues around 0. In the right plot of Fig. 2, we show the convergence of CG which is evaluated in terms of the mean squared error (MSE) computed as

$$\text{MSE} = \frac{\|\hat{\mathbf{b}}^{[n]} - \mathbf{b}\|^2}{\|\mathbf{b}\|^2}, \quad (26)$$

with $\hat{\mathbf{b}}^{[n]}$ standing for the data estimation of \mathbf{b} yielded during the n th iteration.

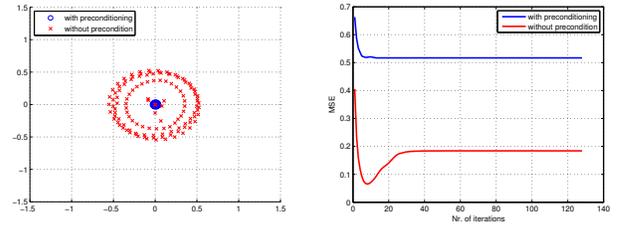


Fig. 2. Left plot: eigenvalues with and without preconditioning; right plot: convergence performance with and without preconditioning.

To enforce the eigenvalues to cluster in the “right” area, a more complex structured preconditioner is indispensable, which is, however, not desired due to complexity and implementation considerations. In light of this, we understand that CG without preconditioning can sometimes yield a better performance than with preconditioning, as opposed to what is claimed in [10].

IV. RESAMPLING AND DOPPLER COMPENSATION

The derivations in Section II leave the sampling rate β open for choice. Recall that resampling ($\beta \neq 1$) is a standard step taken in many underwater communication works to compensate for the Doppler effect, when a single-scale channel model is assumed. The objective of resampling is then to make the scaled channel after resampling as close as possible to a time-invariant channel, with the residual Doppler effect being modeled as a carrier frequency offset (CFO). The advantage of resampling is that after CFO correction, the frequency-domain channel matrix is approximately diagonal and a simple one-step equalizer, typical to OFDM, is still viable. In contrast to a single-scale channel assumption, resampling a multi-scale channel is less straightforward as argued in [13].

In this paper, we will study resampling from a slightly different point of view. Note that resampling itself does not inflict any information loss as long as the Nyquist criterion is

satisfied. Focusing on the significant entries of the frequency-domain channel matrix, whether or not this matrix is diagonal is in principle not crucial for the equalizers discussed in this paper. However, it is still beneficial to find a resampling factor β , such that we can attain a new channel matrix, denoted by $\mathcal{H}_{F, \text{opt}}$, for which the channel energy is as much as possible concentrated on the main diagonal for the following two reasons: 1) the analysis from the previous section learns that in order for the diagonal preconditioner to be effective, the diagonal power needs to be larger than the sum of the power on the off-diagonal elements [c.f. (25)]; 2) it is well-known from the Geršgorin Theorem [15] that the eigenvalues of $\mathcal{H}_{F, \text{opt}}$ lie in the union of the Geršgorin discs, which are circles in the complex plane centered at $[[\mathcal{H}_{F, \text{opt}}]_{k,k}]$ for $k = 0, \dots, K-1$, with a radius of $\sum_{m=0, m \neq k}^{K-1} |[\mathcal{H}_{F, \text{opt}}]_{m,k}|$. Optimal resampling implies that these Geršgorin discs will be maximally compressed and located farthest away from 0. Note that a lower condition number can already enhance the convergence rate of CG without preconditioning.

Because it is difficult to maximize the diagonal power of $\mathcal{H}_{F, \text{opt}}$ for each diagonal entry individually, as a relaxation, we desire that the function $f(\mathcal{H}_{F, \text{opt}})$ defined as

$$f(\mathcal{H}_{F, \text{opt}}) = \frac{\sum_{k=0}^{K-1} |[\mathcal{H}_{F, \text{opt}}]_{k,k}|^2}{\sum_{k=0}^{K-1} \sum_{m=0, m \neq k}^{K-1} |[\mathcal{H}_{F, \text{opt}}]_{m,k}|^2}, \quad (27)$$

should yield a maximum. To this end, let us first introduce the time-domain channel matrix resulting from a resampling rate of β as [c.f. (9)]

$$\mathcal{H}_T^{(\beta)} := \sum_{l=0}^{L-1} h_l^{(\beta)} \text{diag}\{\mathbf{g}^{(\alpha_l-1, \beta)}\} \mathbf{F}_{K, \frac{\alpha_l}{\beta}}^H \text{diag}\{\boldsymbol{\lambda}_l\}, \quad (28)$$

and consider the matrix $\mathcal{H}_{F, \text{opt}} = \mathbf{F}_K \text{diag}\{\mathbf{g}^{(\gamma-1, \beta)H}\} \mathcal{H}_T^{(\beta)}$ with β and γ denoting the parameters that is obtained by solving the following optimization problem

$$\{\beta_{\text{opt}}, \gamma_{\text{opt}}\} = \arg \max_{\beta, \gamma} f(\mathbf{F}_K \text{diag}\{\mathbf{g}^{(\gamma-1, \beta)H}\} \mathcal{H}_T^{(\beta)}). \quad (29)$$

The necessity of $\text{diag}\{\mathbf{g}^{(\gamma-1, \beta)H}\}$ is due to the fact that resampling alone is not able to maximize the diagonal energy. Even in the simplest case of a single-scale model where $\alpha_0 = \dots = \alpha_{L-1} = \alpha$, after resampling, the most energy will be located on one of the sub-diagonals (or super-diagonals) of the resulting channel matrix, which is commonly known as the CFO. It is $\text{diag}\{\mathbf{g}^{(\gamma, \beta)}\}$ that compensates for this CFO and shifts all the energy to the main diagonal. Obviously, in this single-scale case, both the optimal values of β and γ are both equal to α .

Because the Frobenius norm of $\mathcal{H}_T^{(\beta)}$, which equals $\sum_{k=0}^{K-1} \sum_{m=0}^{K-1} |[\mathcal{H}_{F, \text{opt}}]_{m,k}|$ is almost constant for different

β 's, we are allowed to simplify (29) as

$$\begin{aligned} & \arg \max_{\beta, \gamma} \sum_{k=0}^{K-1} |[\mathbf{F}_K \text{diag}\{\mathbf{g}^{(\gamma, \beta)}\} \mathcal{H}_T^{(\beta)}]_{k,k}|^2 \\ \Leftrightarrow & \arg \max_{\beta, \gamma} \frac{1}{\beta} \sum_{k=0}^{K-1} \left| \sum_{l=0}^{L-1} c_l \frac{e^{j\frac{2\pi}{K}(k-\frac{\alpha_l}{\beta}k-\frac{\alpha_l+\gamma}{\beta}D)K-1}}{e^{j\frac{2\pi}{K}(k-\frac{\alpha_l}{\beta}k-\frac{\alpha_l+\gamma}{\beta}D)-1}} \right|^2, \quad (30) \end{aligned}$$

with $c_l := \sqrt{\alpha_l} h_l e^{-j2\pi(f_c T - \frac{k}{K})\rho_l}$. A closed-form solution to the above cost function is challenging. Note that for underwater acoustic channels with a realistic vessel velocity, the typical range of scale is between 0.98 and 1.02. For this reason, we can assume that the value of β and γ should lie within this range. Furthermore, if we agree on a solution resolution of 10^{-3} , then (30) can be solved by virtue of a 2-D exhaustive search for all possible values of β and γ .

As an example, we show in Fig. 3 the result of evaluating (30) upon $\beta = \{0.98, 0.981, \dots, 1.02\}$ and $\gamma = \{0.98, 0.981, \dots, 1.02\}$, for the channel whose parameters are given in Table I. It shows that with $\beta_{\text{opt}} = 0.985$ and $\gamma_{\text{opt}} = 0.998$, we have $f(\mathcal{H}_{F, \text{opt}}) \approx 2.17$. The intensity of $\mathcal{H}_{F, \text{opt}}$ is plotted in Fig. 4. In comparison, if no resampling is adopted, and the receiver uses the same sampling rate as the transmitter $\beta = 1$ (and $\gamma = 1$), then the corresponding channel matrix $\mathcal{H}_F^{(1)}$, whose intensity is plotted in Fig. 1, will have $f(\mathcal{H}_F^{(1)}) \approx 0.002$. Note also that the condition number of $\mathcal{H}_{F, \text{opt}}$ is around 8.84, while the condition number of $\mathcal{H}_F^{(1)}$ is around 1960! In the left plot of Fig. 5, we show that after optimal resampling, the eigenvalues resulting from preconditioning are indeed clustered around 1, which results in a faster convergence as shown in the right plot of Fig. 5.

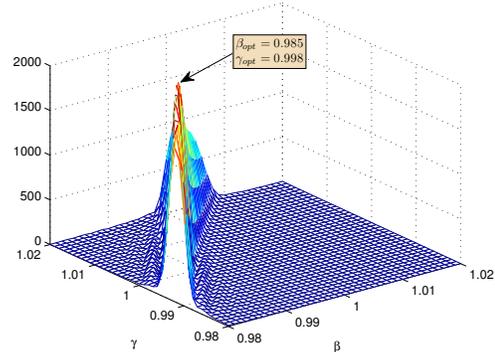


Fig. 3. Evaluation of (30) for a range of β 's and γ 's.

As mentioned earlier, optimal resampling for multi-scale multi-lag channels is also discussed in [13], but a slightly different criterion is applied in [13]. For the discrete data model in particular, [13] aims to find a virtual single-scale multi-lag channel that is closest (in the ℓ_2 -norm sense) to the original discrete channel, which is obtained using the same sampling rate at the receiver as at the transmitter. The virtual single-scale multi-lag channel is assumed in [13] to have the

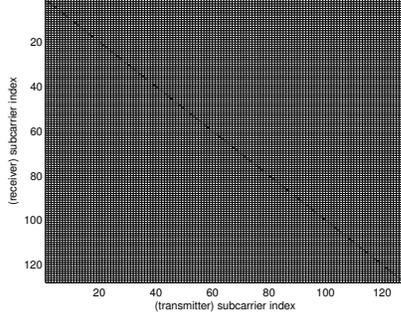


Fig. 4. Optimally resampled $\mathcal{H}_{F,\text{opt}}$.

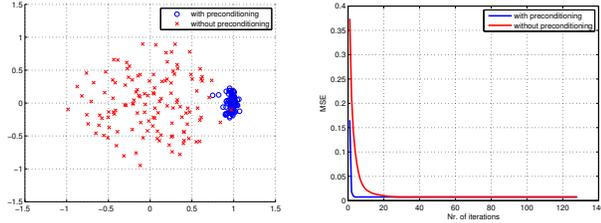


Fig. 5. Left plot: eigenvalues with and without preconditioning; right plot: convergence performance with and without preconditioning. Channel matrix is obtained after optimal resampling.

same path gain and delay as the original channel, but a single scale β . As a result, the optimization cost function, after some algebra, can be equivalently formulated as

$$\arg \min_{\beta} \left\| \mathcal{H}_{\text{T}}^{(1)} - \sum_{l=0}^{L-1} h_l^{(1)} \text{diag}\{\mathbf{g}^{(\beta,1)}\} \mathbf{F}_{K,\beta}^H \text{diag}\{\lambda_l\} \right\|_2^2. \quad (31)$$

In Fig. 6, we show the evaluation of (31) upon $\beta = \{0.98, 0.981, \dots, 1.02\}$, where it can be seen that due to a different design criterion, the corresponding optimal resampling rate is $\tilde{\beta}_{\text{opt}} = 0.987$; If the receiver adopts such a sampling rate, the resulting channel matrix $\mathcal{H}_{\text{F}}^{(\tilde{\beta}_{\text{opt}})}$ will have $f(\mathcal{H}_{\text{F}}^{(\tilde{\beta}_{\text{opt}})}) = 2.10$, which is thus slightly smaller than that of $\mathcal{H}_{\text{F},\text{opt}}$. This is not surprising since $\mathcal{H}_{\text{F}}^{(\tilde{\beta}_{\text{opt}})}$ can be expressed as $\mathbf{F}_K \text{diag}\{\tilde{\gamma}_{\text{opt}}, \tilde{\beta}_{\text{opt}}\} \mathcal{H}_{\text{T}}^{(\tilde{\beta}_{\text{opt}})}$, with $\tilde{\gamma}_{\text{opt}} = 1$, which is thus suboptimal in terms of $f(\cdot)$.

V. SIMULATIONS AND DISCUSSION

We compare the performance of CG with and without preconditioning under the impact of optimal resampling with bit error rate (BER) versus signal-to-noise ratio (SNR) as the performance metric. Here, SNR is defined as the ratio between the received signal strength and the noise. To reduce the complexity, we use a sparse approximation of the real channel matrix for all the equalizers, which is obtained following Method 2 where $\eta = 90\%$ of the entries in each row is discarded. In addition, we halt CG after 5 iterations.

OFDM with $K = 128$ subcarriers is considered. Onto each subcarrier, a QPSK symbols is modulated. The OFDM

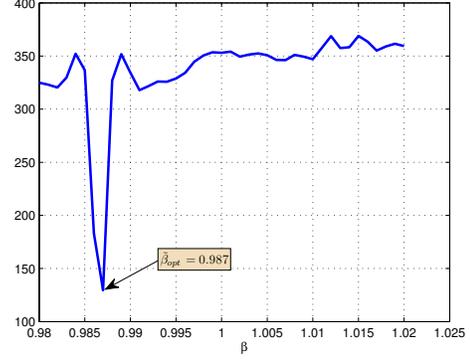


Fig. 6. Evaluation of (30) for a range of β 's.

	Case 1	Case 2	Case 3
ν [knots]	1	20	20
\mathcal{B}	8×10^{-4}	8×10^{-4}	1×10^{-2}

TABLE II
CHOICE OF PATH SCALES

symbols are transmitted at a center frequency of $f_c = 10$ kHz with a bandwidth of 5 kHz. The MSML channel is assumed to have $L = 4$ paths with a power profile of $[0, -3, -5, -7]$ dB. The path delay is chosen as a random variable that has a uniform distribution within the range $[10, 150]$ ms. Likewise, the path scale is chosen as a random variable that has a uniform distribution within the range $[\mathcal{M} - \mathcal{B}/2, \mathcal{M} + \mathcal{B}/2]$, where \mathcal{M} is related to the vessel velocity ν as $\mathcal{M} = 1 + \frac{2\nu}{c}$ with $c = 1500\text{m/s}$ standing for the sound speed in water, and \mathcal{B} gives the size of the range. We differentiate between three multi-scale cases as summarized in Table II.

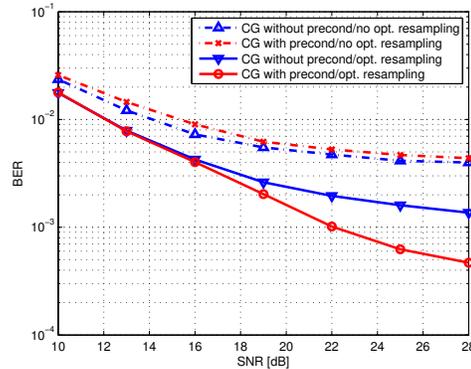


Fig. 7. BER performance of Case 1

The BER performance resulting from Case 1 is plotted in Fig. 7 where one can observe that in case of no optimal resampling, CG with preconditioning performs even worse than without preconditioning. When optimal resampling is applied,

which maximizes the channel energy on the main diagonal, the channel matrix condition is expected to be reduced. As a result, all the equalizers yield a better performance, and CG with preconditioning outperforms CG without preconditioning.

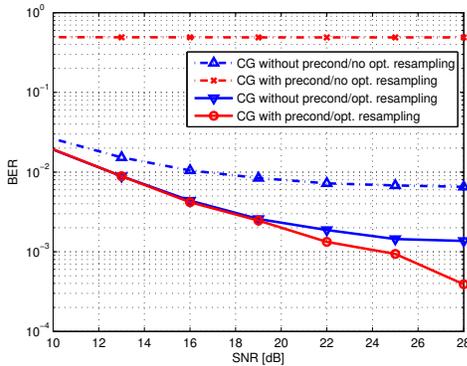


Fig. 8. BER performance of Case 2

The BER performance resulting from Case 2 is plotted in Fig. 8. In this case, because of a higher vessel velocity, the channel energy will be “pushed” farther away from the main diagonal, thus the condition number gets even worse. As a result, only CG without preconditioning can work properly. This situation is greatly improved when optimal resampling is employed.

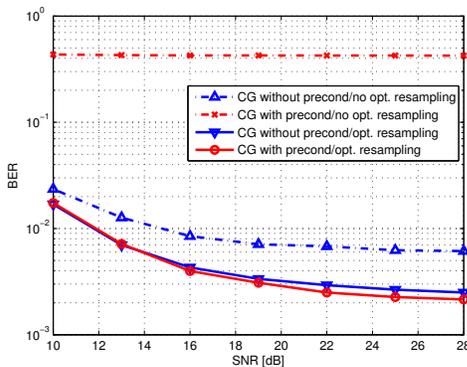


Fig. 9. BER performance of Case 3

The BER performance resulting from Case 3 is plotted in Fig. 8. When no optimal resampling is applied, a similar situation as for Case 2 can be observed as a result of a high vessel velocity. What is different in this case is that the scale discrepancy of each path, represented by the parameter β is much larger than in Case 2, which implies that it will be more difficult to enforce the channel energy to concentrate on the main diagonal. As a result, we observe that the performance lead of CG with preconditioning after optimal resampling becomes less pronounced in this case.

VI. CONCLUSION

In this paper, we gave a brief ICI analysis on the OFDM channel matrix resulting from a multi-scale multi-lag (MSML) channel model. It has been shown that with a realistic Doppler scenario, it is still reasonable to approximate the channel matrix as highly sparse such that a low-complexity equalization is applicable. We have considered the conjugate gradient (CG) algorithm, and argued that CG with a diagonal preconditioner, which is commonly used to accelerate the convergence speed of CG, does not necessarily lead to a performance improvement with respect to CG without preconditioning. This is due to the fact that the major channel energy can deviate from the main diagonal of the channel matrix in the MSML model, and the resulting diagonal preconditioner will enlarge the channel matrix condition making CG converge more slowly. To alleviate this problem, we have proposed to use optimal resampling to ensure that the channel energy on the main diagonal of the channel matrix is maximized.

ACKNOWLEDGMENT

This work is supported in part by NWO-STW under the VICI program (project 10382).

REFERENCES

- [1] Y. Jiang and A. Papandreou-Suppappola, “Discrete time-scale characterization of wideband time-varying systems,” *IEEE Transactions on Signal Processing*, vol. 54, p. 13641375, Apr. 2006.
- [2] U. Mitra and G. Leus, “Equalizers for multi-scale/multi-lag wireless channels,” in *Globecom*, (Miami, Florida), 2010.
- [3] P. Schniter, “Low-complexity equalization of OFDM in doubly-selective channels,” *IEEE Transactions on Signal Processing*, vol. 52, pp. 1002–1011, Apr. 2004.
- [4] L. Rugini, P. Banelli, and G. Leus, “Simple equalization of time-varying channels for OFDM,” *IEEE Communications Letters*, vol. 9, pp. 619–621, July 2005.
- [5] Z. Tang and G. Leus, “A novel receiver architecture for single-carrier transmission over time-varying channels,” *IEEE Journal on Selected Areas in Communications*, vol. 26, pp. 366–377, Feb. 2008.
- [6] G. H. Golub and C. F. van Loan, *Matrix Computations*. The Johns Hopkins University Press, 1989.
- [7] Y. Saad, *Iterative methods for sparse linear systems*, 2nd edition. SIAM, 2003.
- [8] M. Benzi, “Preconditioning techniques for large linear systems: A survey,” *Journal of Computational Physics*, vol. 182, pp. 418–477, Nov. 2002.
- [9] T. Hrycak, S. Das, G. Matz, and H. G. Feichtinger, “Low complexity equalization for doubly selective channels modeled by a basis expansion,” *IEEE Transactions on Signal Processing*, vol. 58, pp. 5706–5719, Nov. 2010.
- [10] M. J. Grote and T. Huckle, “Parallel preconditioning with sparse approximate inverses,” *SIAM Journal on Scientific Computing*, vol. 18, pp. 838–853, May 2007.
- [11] B. S. Sharif, J. Neasham, O. R. Hinton, and A. E. Adams, “A computationally efficient Doppler compensation system for underwater acoustic communications,” *IEEE Journal of Oceanic Engineering*, vol. 25, pp. 52–61, Jan. 2000.
- [12] B. Li, S. Zhou, M. Stojanovic, L. Freitag, and P. Willett, “Multicarrier communication over underwater acoustic channels with nonuniform Doppler shifts,” vol. 51, pp. 198–209, June 2003.
- [13] S. Yerramalli and U. Mitra, “On optimal resampling for ofdm signaling in doubly-selective underwater acoustic channels,” *IEEE Journal of Oceanic Engineering*, vol. 36, pp. 126–138, Jan. 2011.
- [14] L. Rugini, P. Banelli, and G. Leus, “Low-complexity banded equalizers for OFDM systems in Doppler spread channels,” *EURASIP Journal on Applied Signal Processing*, pp. Article ID 67404, 13 pages, 2006.
- [15] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1999.