

Efficient Sequence Detection of Multicarrier Transmissions over Doubly Dispersive Channels

Sung-Jun Hwang and Philip Schniter

Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210, USA

Received 2 June 2005; Revised 1 May 2006; Accepted 12 May 2006

We propose a high-spectral-efficiency multicarrier system for communication over the doubly dispersive (DD) channel which yields very low frame error rate (FER), with quadratic (in the frame length) receiver complexity. To accomplish this, we combine a non-(bi)orthogonal multicarrier modulation (MCM) scheme recently proposed by the authors with novel sequence detection (SD) and channel estimation (CE) algorithms. In particular, our MCM scheme allows us to accurately represent the DD channels otherwise complicated intercarrier interference (ICI) and intersymbol interference (ISI) response with a relatively small number of coefficients. The SD and CE algorithms then leverage this sparse ICI/ISI structure for low-complexity operation. Our SD algorithm combines a novel adaptive breadth-first search procedure with a new fast MMSE-GDFE preprocessor, while our CE algorithm uses a rank-reduced pilot-aided Wiener technique to estimate only the significant ICI/ISI coefficients.

Copyright © 2006 Hindawi Publishing Corporation. All rights reserved.

1. INTRODUCTION

In wireless data communication, the information signal undergoes multipath propagation which, due to variations among path lengths, induces a time-domain spreading effect on the information signal. Furthermore, relative motion between the transmitter, receiver, and scattering objects imparts each path with a unique Doppler shift, so that multipath propagation also induces a frequency-domain spreading effect on the information signal. We refer to such channels as “doubly dispersive” (DD).

Reliable high-spectral-efficiency communication over the DD channel is difficult. Consider that a sequence of N symbols transmitted over this channel will appear, to the receiver, as a complicated time-variant mixture corrupted by additive noise. The mixing may make it difficult to correctly infer the transmitted sequence, even when optimal maximum-likelihood (ML) sequence detection (SD) is used. Furthermore, the complexity of MLSD may be impractical. In general, communication over the DD channel is a compromise between spectral efficiency, frame error rate (FER), and implementation complexity. For example, by sacrificing spectral efficiency, one could transmit symbols separated far enough in time and/or frequency to avoid interference, thereby guaranteeing simple optimal reception. However, since low spectral efficiency cannot usually be tolerated, the properties of DD-induced interference play a fundamental role in communication performance and complexity.

We can identify two major approaches to the design of coherent communication schemes for the DD channel. In the so-called maximum-diversity linear precoding (MDLP) approach [1], linear modulation waveforms are designed to maximize the exploitable diversity at the channel output in an effort to minimize the FER achieved by MLSD in the high-SNR regime. MDLP makes liberal use of time-domain and frequency-domain guard intervals, which limits its spectral efficiency to about 0.5 QAM-symbols/s/Hz for the DD channels we consider, which have time-frequency spreading factors in the range 0.03–0.1. More significantly, such channels require long MDLP frames (e.g., $N \sim 1000$) for which MLSD is infeasible. Though suboptimal reduced-complexity decision feedback (DF) detectors have been proposed to alleviate this problem [2], they too remain computationally impractical for these highly dispersive channels.

In what we will refer to as the multicarrier modulation (MCM) approach [3], linear modulation waveforms are designed to yield a “simple” interference response—in order to ease the SD task—without explicitly considering the achievable FER performance. The vast majority of DD-channel communication schemes fit into this category, for example, cyclic-prefix (CP) orthogonal frequency-division multiplexing (OFDM) [4], zero-padded (ZP) OFDM [5], and Strohmer and Beaver’s “optimal” OFDM [6]. For example, CP-OFDM and ZP-OFDM were originally designed for time-dispersive—rather than doubly dispersive—channels, and are capable of totally suppressing intersymbol interference

(ISI). When used in DD channels, however, CP-OFDM and ZP-OFDM succumb to significant intercarrier interference (ICI) which greatly complicates SD. In response, more sophisticated MCM schemes have been proposed based on smooth ISI/ICI-minimizing pulses. Though these “pulse-shaped” MCM schemes succumb to less ICI than their ZP-OFDM and CP-OFDM counterparts, their ISI/ICI responses are, in general, still too complicated for practical MLSD.

Due to the impracticality of the MLSD in DD-channel MCM, several methods of reduced-complexity reception have been proposed. These schemes are typically based on the combination of *ISI/ICI truncation* with *suboptimal SD*. By *ISI/ICI truncation*, we mean that only the “significant” ICI/ISI coefficients are estimated at the receiver and used in SD. Examples of suboptimal SD include linear detection (e.g., [7–9]), DF detection (e.g., [10–12]), iterative/turbo detection (e.g., [13–15]), and approximate-ML detection (e.g., [16–19]). We conclude that the judicious design of a DD-channel communication system includes

- (1) MCM that near-perfectly suppresses all but a small number of ISI/ICI coefficients,
- (2) a near-ML SD algorithm which leverages the structure of significant-ISI/ICI for complexity reduction, and
- (3) a high-performance estimation of the significant-ISI/ICI coefficients.

In the present paper, we combine the non-(bi)orthogonal (NBO) MCM previously proposed by the authors in [14, 15] with near-ML sequential decoding (SqD) algorithms [20–22]—sometimes referred to as lattice decoders or tree search decoders—with rank-reduced pilot-aided Wiener channel estimation for high-spectral-efficiency, high-performance, and low-complexity multicarrier communication over the DD channel. By “near ML,” we mean FER performance equivalent to that attained by MLSD at a fraction-of-a-dB lower signal-to-noise ratio (SNR). We tolerate this small loss because, as we will see, it enables huge complexity savings relative to true MLSD. We choose the NBO-MCM scheme from [14, 15] because of its high spectral efficiency and excellent ISI/ICI suppression; these considerations will be discussed further in Section 2.1. We propose SqD based on a novel fast MMSE-GDFE preprocessor [23] and on a novel channel-adaptive T -algorithm [24], both of which are specifically tailored to the ISI/ICI structure induced by NBO-MCM over the DD channel. We discuss, in Section 2.3, the shortcomings of traditional SqDs on these channels. Numerical experiments are conducted to evaluate the efficacy of the NBO-MCM scheme, the proposed SqD, the channel estimator, and their combination, relative to other designs.

The paper is organized as follows. Section 2 reviews MCM and SqD and establishes our system model. Section 3 presents the low-complexity preprocessing techniques, the channel-adaptive T -algorithm, and the rank-reduced channel estimation algorithm. Numerical results are given in Section 4 and conclusions in Section 5.

We use $(\cdot)^T$ to denote the transpose, $(\cdot)^*$ the conjugate, and $(\cdot)^H$ the conjugate transpose. $\mathcal{D}(\mathbf{b})$ denotes the diagonal matrix created from vector \mathbf{b} , \mathbf{I}_L denotes the $L \times L$ iden-

tity matrix, and $[\mathbf{B}]_{m,n}$ denotes the element in the m th row and n th column of matrix \mathbf{B} , where row/column indices begin with zero. Similarly, $[\mathbf{b}]_m$ denotes the m th entry of vector \mathbf{b} . Expectation is denoted by $E\{\cdot\}$, the ℓ_2 norm by $\|\cdot\|$, the Kronecker delta by δ_l , and the modulo- N operation by $\langle \cdot \rangle_N$. Finally, \mathbb{R} denotes the real field, \mathbb{C} the complex field, and \mathbb{Z} the integers.

2. BACKGROUND

2.1. Multicarrier modulation

Equations (1)–(4) describe the baseband-equivalent operation of a QAM-based MCM system in a DD channel. The MCM transmitter uses time-frequency shifts of the pulse $a(t)$ to modulate the QAM data $\{s_{k,n}\}$ onto the transmitted waveform $s(t)$. In (1), T_s denotes the symbol spacing and F_s the subcarrier spacing. The channel, characterized by the time-varying impulse response $h(t, \tau)$ and the noise waveform $z(t)$, produces the received signal $x(t)$. The receiver then uses time-frequency shifts of the pulse $b(t)$ to generate the sub-channel outputs $\{x_{l,m}\}$. Equation (4) decomposes $x_{l,m}$ into its desired, ICI, ISI, and noise components, respectively, using the pulse-shaped channel coefficients $\{h_{l,m,k,n}\}$. Though it is straightforward to write $h_{l,m,k,n}$ in terms of $h(t, \tau)$, $a(t)$, and $b(t)$, we omit the expression here for brevity:

$$s(t) = \sum_{n=-\infty}^{\infty} \sum_{k=0}^{N-1} s_{k,n} a(t - nT_s) e^{j2\pi k F_s (t - nT_s)}, \quad (1)$$

$$x(t) = \int_0^{T_h} h(t, \tau) s(t - \tau) d\tau + z(t), \quad (2)$$

$$x_{l,m} = \int_{-\infty}^{\infty} x(t) b^*(t - mT_s) e^{-j2\pi l F_s t} dt \quad \text{for } 0 \leq l < N \quad (3)$$

$$= h_{l,m} s_{l,m} + \sum_{k \neq l} h_{l,m,k} s_{k,m} + \sum_{k=0}^{N-1} \sum_{n \neq m} h_{l,m,k,n} s_{k,n} + z_{l,m}. \quad (4)$$

In MCM systems based on offset-QAM [25], the real and imaginary components of each QAM symbol are transmitted with a relative time offset of $T_s/2$ seconds, requiring a reformulation of (1).

The pulses $a(t)$ and $b(t)$ are typically designed to suppress ISI and/or ICI, assuming knowledge of the channel statistics (e.g., maximum delay and Doppler spreads), but not of channel realizations, which change very quickly in the DD case. MCM designs can be categorized into orthogonal (e.g., [6, 26–28]), biorthogonal (e.g., [29, 30]), and non-(bi)orthogonal (e.g., [11, 13–16, 31]) designs. We give a brief overview of these three schemes below; see [25] for a comprehensive overview of orthogonal and biorthogonal MCM.

Orthogonal MCM sets $b(t) = a(t)$, and constrains $a(t)$ to be orthogonal to $a(t - nT_s) e^{j2\pi k F_s (t - nT_s)}$ for all nonzero $(n, k) \in \mathbb{Z}^2$. Orthogonal MCM has the intuitively satisfying properties that, in a nonspreading channel with flat

noise spectral density, ICI/ISI will vanish and the subchannel noise $\{z_{l,m}\}$ will be white. Because the Gaussian pulse $g_\sigma(t) := (2\sigma)^{0.25} e^{-\pi\sigma t^2}$ achieves the optimal time-frequency localization, several authors have proposed MCM based on orthogonalization of $g_\sigma(t)$ [6, 27]. For example, Strohmer and Beaver [6] specified an orthogonalization procedure that yields an “optimally time-frequency localized” $a(t)$, that is, $a(t)$ that is closest (in the L_2 sense) to $g_\sigma(t)$ among all possible orthogonal pulse shapes. Biorthogonal MCM allows $b(t)$ to be different than $a(t)$, as long as $b(t)$ remains orthogonal to $a(t - nT_s)e^{j2\pi kF_s(t-nT_s)}$ for all nonzero $(n, k) \in \mathbb{Z}^2$. In biorthogonal MCM, ICI/ISI vanishes in nonspreading channels though the noise samples $\{z_{l,m}\}$ may be correlated [29]. Due to more freedom in pulse design, biorthogonal MCM can suppress DD-channel-induced ICI/ISI better than orthogonal MCM (at the same spectral efficiency). Non-(bi)orthogonal (NBO) MCM goes one step further and removes the ICI/ISI-free constraint for nonspreading channels in the hope of better ICI/ISI suppression in DD channels.

In striving for near-ML performance, it is of critical importance to suppress residual ICI/ISI. In [19], for example, residual ICI was ignored so that the Viterbi algorithm [19, 32] could be applied in DD-channel CP-OFDM, with the result being a large gap between ICI/ISI-truncated Viterbi performance and true MLS. For efficient near-ML SD, we also find it essential that the subchannel noise $\{z_{l,m}\}$ is white, since the whitening of colored subchannel noise would effectively destroy the sparse ICI/ISI structure which we wish to exploit in complexity reduction. Finally, we desire an MCM scheme with high spectral efficiency, since we consider data rate to be of paramount importance.

We know of only one MCM technique which ensures white noise, high spectral efficiency, and near-perfectly suppressed residual ICI/ISI: the “max-SINR” transmission-pulse (MSTP)-MCM that we proposed in [14, 15]. In this NBO-MCM scheme, the transmission pulse $a(t)$ is designed to maximize a signal-to-interference-plus-noise ratio (SINR), where “signal” refers to the average energy contributed to $x_{l,m}$ from $s_{l,m}$, and where interference-plus-noise refers to the average energy contributed to $x_{l,m}$ from ISI, from ICI beyond a radius of D subcarriers, and from additive noise. The MSTP-MCM reception pulse $b(t)$ is rectangular, as in CP-OFDM, to facilitate white subchannel noise. For pulse design, we assume that the channel’s maximum delay and Doppler spreads are known,¹ though not the channel’s realization. For even highly spread channels, MSTP-MCM performs well at the Nyquist rate of 1 QAM-symbol/s/Hz, that is, that of CP-OFDM with zero-length CP. For more details on MSTP-MCM, see [14, 15]. Section 4 conducts a detailed comparison of MSTP-MCM, CP-OFDM, ZP-OFDM, and Strohmer and Beaver’s “optimal” orthogonal MCM.

¹ In CP-OFDM and ZP-OFDM, knowledge of delay spread is implicitly assumed in guard length selection. In nearly all orthogonal and (bi)orthogonal MCMs, knowledge of both delay and Doppler spread is implicitly assumed in pulse design.

2.2. System model

We consider an N -subcarrier QAM-based² MCM system operating in a noisy baseband-equivalent DD channel, as described by (1)–(4). A square QAM constellation of size, Q^2 with real and imaginary components chosen from the Q-ary PAM constellation $\mathcal{S} := \{-(Q-1)/2, -(Q-1)/2 + 1, \dots, (Q-1)/2\}$, is assumed. By splitting the complex-valued elements $\{x_{l,m}\}_{l=0}^{N-1}$, $\{s_{k,m}\}_{k=0}^{N-1}$, $\{z_{l,m}\}_{l=0}^{N-1}$, and $\{h_{l,m,k,n}\}_{l,k=0}^{N-1}$ from (4) into their real and imaginary components, we obtain the real-valued vector model (5), which will be more convenient for SqD implementation. In particular, the vector $\mathbf{x}_m \in \mathbb{R}^{2N}$ is constructed so that $[\mathbf{x}_m]_{2l} = \text{Re}(x_{l,m})$ and $[\mathbf{x}_m]_{2l+1} = \text{Im}(x_{l,m})$ for $0 \leq l < N$, while $\mathbf{s}_m \in \mathbb{R}^{2N}$, $\mathbf{z}_m \in \mathbb{R}^{2N}$, and $\mathbf{H}_{m,n} \in \mathbb{R}^{2N \times 2N}$ are constructed in a similar manner:

$$\mathbf{x}_m = \sum_{n=-\infty}^{\infty} \mathbf{H}_{m,n} \mathbf{s}_{m-n} + \mathbf{z}_m. \quad (5)$$

Note that the matrix sequence $\{\mathbf{H}_{m,n}\}_{n=-\infty}^{\infty}$ specifies the impulse response relating the transmitted multicarrier-symbol sequence $\{\mathbf{s}_n\}_{n=-\infty}^{\infty}$ to the time- m modulator output \mathbf{x}_m ; it is a function of the pulse shapes $\{a(t), b(t)\}$ and the channel realization $h(t, \tau)$. Thus, the matrix coefficients $\{\mathbf{H}_{m,n}\}_{n \neq 0}$ characterize the intersymbol interference (ISI) while the off-diagonal elements of $\mathbf{H}_{m,0}$ characterize the intercarrier interference (ICI).

While much of the theoretical MCM literature assumes continuous pulse shapes as in (1)–(3), practical MCM implementations use pulse sequences $\{a_k\}$ and $\{b_k\}$ to modulate a chip-waveform $p(t)$ with approximate time support $T_c = 1/NF_s$ and approximate frequency support NF_s [25], that is, $a(t) = \sum_k a_k p(t - kT_c)$ and $b(t) = \sum_k b_k p(t - kT_c)$. In this case, the significant entries in $\mathbf{H}_{m,0}$ lie within the “quasi-banded” support shown in Figure 1(a), where the “ICI radius” D depends on the pulse designs and channel spreading characteristics. Specifically, D is chosen so that $D = 2(\lceil f_d T_c N \rceil + C_{\min})$, where $f_d T_c$ denotes the maximum single-sided Doppler spread and C_{\min} is a small nonnegative integer that is chosen based on the pulse design.³ This phenomenon motivates the partition $\mathbf{H}_{m,0} = \mathbf{H}_m^D + \hat{\mathbf{H}}_m^D$, where \mathbf{H}_m^D extracts the coefficients of $\mathbf{H}_{m,0}$ inside the shaded region of Figure 1(a), and where $\hat{\mathbf{H}}_m^D$ extracts the coefficients outside the shaded region. More precisely, for $0 \leq D < N$,

$$[\mathbf{H}_m^D]_{k,l} := \begin{cases} [\mathbf{H}_{m,0}]_{k,l} & \text{for } k, l \text{ s.t. } -D \leq \langle k - l + N \rangle_{2N} - N \leq D, \\ 0 & \text{else.} \end{cases} \quad (6)$$

² Though the real-valued equation (5) is capable of modeling OQAM-MCM, we restrict the focus of this paper to QAM-MCM.

³ For MSTP-MCM, we find that $C_{\min} = 2$ yields the best FER performance; $C_{\min} = 1$ performs only slightly worse.

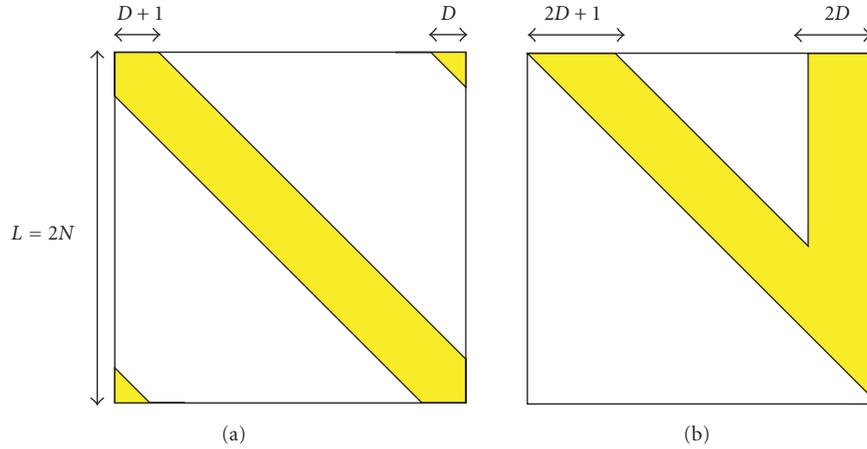


FIGURE 1: Channel matrices associated with MCM: (a) “quasibanded” channel matrix, (b) “V-shaped” channel matrix.

Using this partition, we rewrite (5) as

$$\mathbf{x}_m = \mathbf{H}_m^D \mathbf{s}_m + \underbrace{\tilde{\mathbf{H}}_m^D \mathbf{s}_m + \sum_{n \neq 0} \mathbf{H}_{m,n} \mathbf{s}_{m-n}}_{:= \mathbf{w}_m} + \mathbf{z}_m, \quad (7)$$

where $\mathbf{H}_m^D \mathbf{s}_m$ contains the signal and “significant ICI,” while \mathbf{w}_m contains the noise, ISI, and “insignificant ICI.” We will see that MSTP-MCM [14, 15] guarantees $E\{\mathbf{z}_m \mathbf{z}_m^T\} = \sigma_z^2 \mathbf{I}$ and suppresses both ISI and insignificant ICI to a level well below the noise floor, so that $E\{\mathbf{w}_m \mathbf{w}_m^T\} \approx \sigma_z^2 \mathbf{I}$, even with a highly dispersive channel over a broad range of SNR.

2.3. Sequential decoding

The MCM features noted at the end of Section 2.2 allow us to focus on a system model free of ISI and insignificant ICI. Suppressing the m and D notation, (7) becomes

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{w}, \quad (8)$$

where \mathbf{H} retains the quasibanded structure in Figure 1(a) and \mathbf{w} is white Gaussian noise. Since (8) involves $2N$ -dimensional real-valued vectors, we define $L := 2N$ for use in the sequel. By definition, the MLSD solution to (8) under known \mathbf{H} has the form

$$\hat{\mathbf{s}}_{\text{ML}} = \arg \min_{\mathbf{s} \in \mathcal{S}^L} \|\mathbf{x} - \mathbf{H}\mathbf{s}\|^2. \quad (9)$$

The brute-force approach to finding $\hat{\mathbf{s}}$ requires $\mathcal{O}(Q^L)$ operations, which is impractical for large L . If \mathbf{H} was banded with a band radius of D , then the Viterbi algorithm could be used to solve (9) with a complexity of $L(2D+1)Q^{(2D+1)}$ real multiply-accumulate (MAC) operations per frame [19]. Since \mathbf{H} is only quasibanded, a different approach is needed. For example, one could instead use a “tail-biting” MLSD which hypothesizes an initial state at an arbitrary location within the frame, runs the standard Viterbi algorithm from that state, and forces a termination back to that state. Exhaustively searching among the Q^{2D} possible hypotheses yields an

MLSD algorithm with a complexity of $L(2D+1)Q^{(4D+1)}$ real MACs per frame. However, these Viterbi algorithms, while much cheaper than brute force search, will still be impractical in many applications.

Closest lattice point search (CLPS) algorithms present an alternative to brute-force and Viterbi MLSD [33]. After converting the linear system (8) to upper triangular form, efficient CLPS algorithms based on sequential decoding (SqD) [20, 21] or sphere decoding (SpD) [34, 35] can be used to implement MLSD with an average complexity far below $\mathcal{O}(Q^L)$. Since SqD and SpD are closely related (see, e.g., [36]), we refer to them collectively as SqD. For the system (8) with general (i.e., nonbanded) channel matrix \mathbf{H} , for example, sphere decoding maintains an average complexity of approximately $\mathcal{O}(L^3)$ at high SNR, regardless of constellation size Q [36]. This remarkable fact encourages a more thorough investigation of SqD algorithms capable of leveraging the quasibanded structure of \mathbf{H} for further complexity reduction. In fact, we will show that quasibanded \mathbf{H} allows near-ML SqD with an average complexity close to $\mathcal{O}(L^2)$. SqD consists of a preprocessing step and a tree search step; both are discussed next.

2.3.1. SqD preprocessing

We refer to “SqD preprocessing” as that which converts the linear system (8) to upper triangular form. The traditional SqD preprocessing method uses the QR decomposition $\mathbf{H} = \mathbf{Q}\mathbf{R}$ to transform (8) into the equivalent system $\mathbf{x}' = \mathbf{Q}^T \mathbf{x} = \mathbf{R}\mathbf{s} + \mathbf{w}'$, where \mathbf{R} is upper triangular and \mathbf{w}' is statistically equivalent to \mathbf{w} . In this case, the detection problem (9) is equivalently restated as

$$\hat{\mathbf{s}}_{\text{ML}} = \arg \min_{\mathbf{s} \in \mathcal{S}^L} \|\mathbf{x}' - \mathbf{R}\mathbf{s}\|^2. \quad (10)$$

It is not unusual for the preprocessed channel matrix \mathbf{R} to be ill-conditioned. When this is the case, the complexity of near-ML SqD is known to grow significantly [22].

Minimum mean-squared error (MMSE) generalized decision feedback equalization (GDFE) preprocessing [23, 36]

was recently proposed as an alternative to the traditional QR preprocessing. It is motivated by the well-known fact that, under perfect decision feedback, the MMSE-GDFE [37] exhibits higher signal to interference-plus-noise ratio (SINR) than the zero-forcing DFE at the decision point. We now outline the main ideas behind the MMSE-GDFE preprocessing algorithm in [23]. Under the assumptions that \mathbf{s} and \mathbf{w} are zero-mean uncorrelated random vectors with covariance matrices $\sigma_s^2 \mathbf{I}_L$ and $\sigma_w^2 \mathbf{I}_L$, respectively, we define $\gamma := \sigma_s^2 / \sigma_w^2$ and the augmented channel matrix $\tilde{\mathbf{H}}$ in (11):

$$\tilde{\mathbf{H}} := \begin{pmatrix} \mathbf{H} \\ \frac{1}{\sqrt{\gamma}} \mathbf{I}_L \end{pmatrix} \quad (11)$$

$$= \tilde{\mathbf{Q}} \tilde{\mathbf{R}} = \begin{pmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \end{pmatrix} \tilde{\mathbf{R}}. \quad (12)$$

Equation (12) gives the QR decomposition of $\tilde{\mathbf{H}}$, where $\tilde{\mathbf{Q}}$ has orthonormal columns and $\tilde{\mathbf{R}}$ is upper triangular with positive diagonal entries. MMSE-GDFE preprocessing produces the transformed observation $\boldsymbol{\rho} := \mathbf{Q}_1^T \mathbf{x}$ which is used in the detection problem

$$\hat{\mathbf{s}}_{\text{PP}} = \arg \min_{\mathbf{s} \in \mathcal{S}^L} \|\boldsymbol{\rho} - \tilde{\mathbf{R}}\mathbf{s}\|^2. \quad (13)$$

Because $\mathbf{Q}_1 \in \mathbb{R}^{L \times L}$ is not guaranteed to be orthogonal, we cannot claim (for general⁴ constellations \mathcal{S}) that $\hat{\mathbf{s}}_{\text{PP}} = \hat{\mathbf{s}}_{\text{ML}}$. When \mathbf{H} is fully populated (i.e., not quasibanded) as in flat-fading multiantenna communication, Damen [23] demonstrated that, at moderate-to-high SNR, $\hat{\mathbf{s}}_{\text{PP}}$ is near-ML and can be found, via SqD, at an average search complexity of $O(L^3)$, regardless of constellation size Q . We note, for later use, that the error $\mathbf{n} := \boldsymbol{\rho} - \tilde{\mathbf{R}}\mathbf{s}$, while signal dependent and non-Gaussian, is white with covariance $\sigma_w^2 \mathbf{I}_L$ [39].

It is important to realize that, when \mathbf{H} has the quasibanded structure in Figure 1(a), $\tilde{\mathbf{R}}$ will have the “V-shaped” structure in Figure 1(b). Since, as we will see, the V-shaped structure can have a profound effect on SqD behavior, it is worthwhile to consider the conditions under which this V-shaping arises. As suggested by Figure 1, we measure the degree of V-shaping by the ratio $(4D + 1)/2N$; as $(4D + 1)/2N$ decreases below 1, the V-shaping becomes more prominent. Recalling $D = 2(\lceil f_d T_c N \rceil + C_{\min})$ and assuming the typical choice $N = 4N_h$, where $N_h := T_h/T_c$ denotes the normalized delay spread, we find

$$\frac{4D + 1}{2N} = \frac{8\lceil 4f_d T_c N_h \rceil + 8C_{\min} + 1}{8N_h} = \frac{1.125 + C_{\min}}{N_h}, \quad (14)$$

where the second equality in (14) holds for all reasonable spreading factors, that is, for $0 < 2f_d T_h \leq 0.5$. When $C_{\min} = 2$ (as used in Section 4), $(4D + 1)/2N = 3.125/N_h$, and so $\tilde{\mathbf{R}}$ will be V-shaped for $N_h > 3$. In most applications of inter-

est, though, we have $N_h \gg 3$, in which case $\tilde{\mathbf{R}}$ is prominently V-shaped.

Additional SqD preprocessing might also be considered. For example, relaxing the constraint $\mathbf{s} \in \mathcal{S}^L$ in (13) to $\mathbf{s} \in \mathcal{Z}^L$ allows more freedom in the choice of lattice basis [22]. In our application, however, we are interested in preserving the quasibanded structure of \mathbf{H} , which limits the types of preprocessing that can be performed. These issues will be discussed further in Section 3.1.2.

2.3.2. Tree search

The preprocessed SD problems (10) and (13) both correspond to tree search over a tree with depth L , where every tree node has Q children. A brute-force approach to tree search would entail the examination of the Euclidean metrics (10) and (13) at each of the Q^L leaf nodes. We are interested in search algorithms which prune branches that are unlikely to contain the ML path, thus drastically reducing the search complexity. Unlike their ML counterparts, near-ML tree search algorithms can, in some cases, discard the ML path, and hence return a suboptimal sequence estimate. Thus, each near-ML algorithm achieves a particular tradeoff between performance and complexity.

Tree search algorithms can be categorized as breadth-first, depth-first, or best-first search algorithms [21, 22]. Breadth-first search algorithms include, for example, the M -algorithm [21], T -algorithm [24], statistical pruning algorithms [40], Wozencraft SqD [41], and Pohst sphere decoder [42]. Depth-first search algorithms include, for example, the Schnor-Euchner sphere decoder (SE-SpD) and its variants [34–36]. Best-first search algorithms include, for example, the stack and Fano algorithms [20, 22, 43]. Since the SqD literature is large and rapidly growing, an exhaustive comparison of existing SqD algorithms is difficult if not impossible. Instead, we focus on a few representative SqDs and discuss their strengths and weaknesses in the context of solving (13) for the DD-channel MCM application, that is, when $\tilde{\mathbf{R}}$ has the V-shaped structure in Figure 1(b), as opposed to the general case of (13) that results from, for example, flat-fading multiantenna channels and time-dispersive single-antenna channels—neither⁵ of which yield V-shaped $\tilde{\mathbf{R}}$. In fact, we find that the structure of $\tilde{\mathbf{R}}$ has a profound effect on SqD behavior.

We now briefly discuss depth-first, breadth-first, and best-first SqD algorithms to gain insight into their behavior in the DD-channel MCM application. But first, we have some notation. We associate every node on the “ i th level” of the tree ($i \geq 0$) with a realization of the partial path

$$\mathbf{s}^{(i)} := [s_i, s_{i+1}, \dots, s_{L-1}]^T \in \mathcal{S}^{L-i}. \quad (15)$$

⁴ It has been established that $\hat{\mathbf{s}}_{\text{ML}} = \mathbf{s} \Rightarrow \hat{\mathbf{s}}_{\text{PP}} = \mathbf{s}$ when the data is uncoded QPSK [38].

⁵ The ICI span of properly designed MCM (i.e., $2D+1$) will be much shorter than the ISI span of an equivalent single-carrier system (i.e., $2N_h$). Thus, while a time-domain channel matrix would be banded, it would have a much wider band than our quasibanded \mathbf{H} . Unless \mathbf{H} has a narrow band, $\tilde{\mathbf{R}}$ will not be V-shaped.

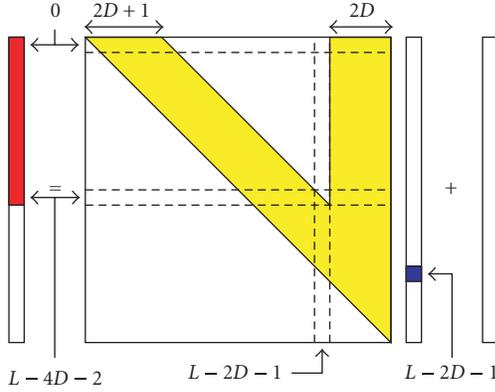


FIGURE 2: Illustration of $\boldsymbol{\rho} = \tilde{\mathbf{R}}\mathbf{s} + \mathbf{n}$ for V-shaped $\tilde{\mathbf{R}}$. The PAM symbol s_{L-2D-1} does not affect $\{\rho_0, \dots, \rho_{L-4D-2}\}$.

The root node corresponds to the L th level and the leaf nodes to the 0th level. The Euclidean partial-path metric associated with $\mathbf{s}^{(i)}$ is defined in (16) using $\tilde{r}_{k,l} := [\tilde{\mathbf{R}}]_{k,l}$:

$$\mathcal{M}(\mathbf{s}^{(i)}) := \sum_{k=i}^{L-1} \left| \rho_k - \sum_{l=k}^{L-1} \tilde{r}_{k,l} s_l \right|^2. \quad (16)$$

(i) Depth-first search

Depth-first search (DFS) algorithms proceed down the tree by following the minimum-cost branch at each level. The first full path obtained in this manner, corresponding to the classical DFE sequence estimate, is kept as a reference. The DFS algorithm then backs up one level at a time, reexamining the discarded branches at each level and pursuing any that have a chance at beating the reference. If a new best-sequence is found, it is used as the new reference and the process is repeated. DFS yields very low search complexity when the initial (i.e., DFE) sequence estimate is ML, since no other branches will be reexamined. For this reason, DFS complexity approaches DFE complexity at high SNR. At low SNR, however, DFS can waste a lot of effort on non-ML paths, leading to very costly searches.

When $\tilde{\mathbf{R}}$ is V-shaped, as in MCM-shaped DD channels, and the SNR is moderate to low, DFS will not be efficient in solving (13). To see why, consider Figure 2, which shows that s_{L-2D-1} does not affect $\{\rho_0, \dots, \rho_{L-4D-2}\}$. Consequently, an error in \hat{s}_{L-2D-1} will be invisible to the branch metrics at levels $i \in \{0, \dots, L-4D-2\}$. When such an error occurs, all DFS branch reexaminations at levels $i \in \{0, \dots, L-4D-2\}$ will be performed in vain. Similar situations occur with errors in \hat{s}_k for $k \in \{2D+1, \dots, L-2D-2\}$. Note that this behavior does not manifest for general upper-triangular $\tilde{\mathbf{R}}$. Thus, while DFS algorithms like the SE-SpD may be attractive in multiantenna or time-dispersive channels, they are not well suited to MCM-shaped DD channels. These notions will be confirmed numerically in Section 4.

(ii) Best-first search

Best-first search (BeFS) algorithms maintain a sorted list of the best partial paths (of possibly different lengths). At each iteration, BeFS extends the best partial path, replaces its list entry with that of its children, and re-sorts the list. BeFS terminates as soon as the best partial path reaches a leaf node, since, at that point, all other partial paths are destined to yield inferior full-path metrics. The Fano algorithm is a near-ML BeFS algorithm that uses the biased partial-path metric

$$\mathcal{M}_{\text{Fano}}(\mathbf{s}^{(i)}) := \sum_{k=i}^{L-1} \left| \rho_k - \sum_{l=k}^{L-1} \tilde{r}_{k,l} s_l \right|^2 - (L-i)b \quad \text{for } b > 0. \quad (17)$$

Larger b biases Fano in favor of longer paths, yielding quicker searches; for very large b , Fano behaves like DFS, greedily extending the best path at every level and returning the DFE sequence estimate. In practice, b is chosen to achieve a particular complexity/performance tradeoff.

A recent comprehensive comparison [22] suggested that a properly designed Fano algorithm achieves a better complexity/performance tradeoff than all other known SqD algorithms when $\tilde{\mathbf{R}}$ has a fully populated upper triangle. For V-shaped $\tilde{\mathbf{R}}$, however, BeFS algorithms (like Fano) can face difficulties. Recalling Figure 2, when the best partial path includes an error in \hat{s}_{L-2D-1} , the branch metrics at levels $i \in \{0, \dots, L-4D-2\}$ will be noninformative about this error, and thus BeFS algorithms can waste lots of time pursuing extensions of this “best” path in vain. Similar situations occur with errors in \hat{s}_k for $k \in \{2D+1, \dots, L-2D-2\}$. Furthermore, best-partial-path errors in any of these \hat{s}_k 's will be gradually deemphasized by the Fano bias term in (17) as these “best” partial paths are extended, making the Fano algorithm less likely to revisit the shorter stack elements without the error in \hat{s}_k . Consequently, Fano exhibits an exploding complexity at low SNR and an inferior complexity/performance tradeoff at high SNR when used with the $\tilde{\mathbf{R}}$ that results from MCM-shaped DD channels. These notions will be confirmed numerically in Section 4.

(iii) Breadth-first search

As we saw earlier, the complexity of DFS and BeFS explodes at low SNR because a huge amount of searching is needed to eliminate suboptimal paths, and the problem is exacerbated by V-shaped $\tilde{\mathbf{R}}$. Breadth-first search (BrFS) complexity, in contrast, is much less sensitive to SNR and the structure of $\tilde{\mathbf{R}}$, suggesting that it might be advantageous in our application. The M -algorithm, for example, has complexity that is *invariant* to both SNR and $\tilde{\mathbf{R}}$. The M -algorithm starts at the root node (i.e., level L) and chooses the M best child nodes at level $L-1$. The children of these level- $(L-1)$ nodes are then evaluated, and the M best are chosen. This process repeats at every level, extending M nodes per level, until finally the best leaf node is chosen as the sequence estimate.

At high SNR, however, the M -algorithm is much more expensive than DFS and BeFS because it is not aggressive enough in branch pruning. Hence, a better complexity/performance tradeoff might be achieved by a BrFS algorithm that varies the number of nodes considered at each level. For example the T -algorithm only extends paths from nodes whose Euclidean metrics lie in the interval $[\mathcal{M}(\mathbf{s}_*^{(i)}), \mathcal{M}(\mathbf{s}_*^{(i)}) + T]$, where $\mathcal{M}(\mathbf{s}_*^{(i)})$ denotes the minimum Euclidean metric among all considered nodes, and where T is a threshold parameter that is chosen to achieve a particular complexity/performance tradeoff. Several approaches to the design of T have been proposed. For example, [24] took an experimental approach, while [44, 45] used SNR and code structure. In Section 3.2 we propose an *adaptive* T -algorithm which uses the elements in $\tilde{\mathbf{R}}$, as well as SNR, to optimize T at each level. We will see that this adaptive T -algorithm results in a superior complexity/performance tradeoff for MCM-shaped DD channels.

3. PROPOSED MCM SEQUENCE DETECTION

In the proposed MCM receiver, a fast SqD preprocessing is applied to the subchannel outputs $\{\mathbf{x}_m\}$ prior to SqD via the adaptive T -algorithm. The channel coefficients used in SqD are estimated via pilot symbols. Below, we describe each receiver component in detail.

3.1. SqD preprocessing

In this section we describe low-complexity SqD preprocessing which leverages the quasibanded structure in \mathbf{H} . For simplicity, we assume system model (8) rather than its notationally elaborate equivalent (5). In Section 3.1.1 we describe a low-complexity implementation of MMSE-GDFE preprocessing, while in Section 3.1.2 we describe a simple ordering scheme which preserves the quasibanded structure in \mathbf{H} .

3.1.1. Fast MMSE-GDFE preprocessing

The MMSE-GDFE preprocessing originally proposed in [23] involves QR decomposition with complexity $\mathcal{O}(L^3)$. In this section, we propose an $\mathcal{O}(D^2L)$ implementation of MMSE-GDFE preprocessing that leverages the quasibanded structure of \mathbf{H} found in our application. We note connections to the fast MMSE-DFE in [11], which was formulated for a banded (as opposed to quasibanded) matrix \mathbf{H} that occurs when the edge subcarriers are inactive.

Recall the augmented channel matrix $\tilde{\mathbf{H}}$ in (11) and its QR decomposition (12). Note that, while \mathbf{H} is quasibanded with $2D+1$ active diagonals (as defined by (6) and illustrated in Figure 1(a)), $\tilde{\mathbf{H}}$ is not quasibanded. However, the matrix $\tilde{\mathbf{H}}^T \tilde{\mathbf{H}}$, which can be computed in $(4D^2 + 4D + 2)L$ MACs, is quasibanded with $4D+1$ active diagonals. Now, since $\tilde{\mathbf{Q}}$ is an orthogonal matrix, we know $\tilde{\mathbf{H}}^T \tilde{\mathbf{H}} = \tilde{\mathbf{R}}^T \tilde{\mathbf{R}}$. Hence, $\tilde{\mathbf{R}}$ can be obtained via Cholesky factorization [46] of $\tilde{\mathbf{H}}^T \tilde{\mathbf{H}}$ in $\mathcal{O}(D^2L)$ operations. Algorithm 1 details the fast Cholesky factorization $\mathbf{A} = \mathbf{G}\mathbf{G}^T$, where $\mathbf{A} := \tilde{\mathbf{H}}^T \tilde{\mathbf{H}}$ and where $\mathbf{G} := \tilde{\mathbf{R}}^T$ is the

Say $\mathbf{A} = \mathbf{G}\mathbf{G}^T$, where \mathbf{G} is lower triangular and $\mathbf{A} \in \mathbb{R}^{L \times L}$ is quasibanded with $\pm 2D$ diagonals.

```

for  $j = 0 : L - 4D - 1$ 
   $v_{j:L-1} = [\mathbf{A}]_{j:L-1,j}$ 
   $m_1 = \max\{0, j - 2D - 1\}$ 
   $m_2 = j + 2D - 1$ 
  for  $i = m_1 : j - 1$ 
     $v_{j:m_2} = v_{j:m_2} - [\mathbf{G}]_{j,i} [\mathbf{G}]_{j:m_2,j}$ 
     $v_{L-2D-1:L-1} = v_{L-2D-1:L-1}$ 
     $\quad - [\mathbf{G}]_{j,i} [\mathbf{G}]_{L-2D-1:L-1,j}$ 
  end
   $[\mathbf{G}]_{j:m_2,j} = v_{j:m_2} / \sqrt{v_j}$ 
   $[\mathbf{G}]_{L-2D-1:L-1,j} = v_{L-2D-1:L-1} / \sqrt{v_j}$ 
end
for  $j = L - 4D : L - 2D - 1$ 
   $v_{j:L-1} = [\mathbf{A}]_{j:L-1,j}$ 
   $m_1 = \max\{0, j - 2D - 1\}$ 
  for  $i = m_1 : j - 1$ 
     $v_{j:L-1} = v_{j:L-1} - [\mathbf{G}]_{j,i} [\mathbf{G}]_{j:L-1,j}$ 
  end
   $[\mathbf{G}]_{j:L-1,j} = v_{j:L-1} / \sqrt{v_j}$ 
end
for  $j = L - 2D : L - 1$ 
   $v_{j:L-1} = [\mathbf{A}]_{j:L-1,j}$ 
  for  $i = 0 : j - 1$ 
     $v_{j:L-1} = v_{j:L-1} - [\mathbf{G}]_{j,i} [\mathbf{G}]_{j:L-1,j}$ 
  end
   $[\mathbf{G}]_{j:L-1,j} = v_{j:L-1} / \sqrt{v_j}$ 
end

```

ALGORITHM 1: Fast cholesky factorization of quasibanded \mathbf{A} .

lower triangular Cholesky factor. This fast computation of $\tilde{\mathbf{R}}$ can be shown to consume $(10D^2 + 11D + 2)L - (1/3)(74D^3 + 133D^2 + 44D + 3)$ MAC operations.⁶

Next, we consider the implementation of the preprocessing operation $\boldsymbol{\rho} = \mathbf{Q}_1^T \mathbf{x}$. Multiplication of this equality by $\tilde{\mathbf{R}}^T$ yields

$$\tilde{\mathbf{R}}^T \boldsymbol{\rho} = \tilde{\mathbf{R}}^T \mathbf{Q}_1^T \mathbf{x} = \mathbf{H}^T \mathbf{x} := \mathbf{b}. \quad (18)$$

Due to quasibanded \mathbf{H} , the vector \mathbf{b} can be computed in $(2D+1)L$ MAC operations. From \mathbf{b} we can solve (18) for $\boldsymbol{\rho}$ using forward substitution in $\mathcal{O}(DL)$ additional operations, because $\tilde{\mathbf{R}}^T$ has the sparse “V-shaped” structure in Figure 1(b). In total, this consumes $(6D+2)L - 6D^2 - 3D$ MAC operations (see footnote 5). Combining forward substitution with fast Cholesky decomposition, our fast MMSE-GDFE preprocessing requires $(14D^2 + 21D + 6)L - (76/3)D^3 - 53D^2 - (53/3)D - 1$ real MAC operations.

⁶ Contact the authors for details.

3.1.2. Circular ordering

In [36], Damen et al. outline three stages of SqD preprocessing: lattice reduction, column ordering, and MMSE-GDFE preprocessing. In our application, the lattice reduction and column ordering would destroy the quasibanded structure of \mathbf{H} , in which case the subsequent MMSE-GDFE preprocessing would require a complexity of $O(L^3)$. Since, in practice, $L = 2N$ can be quite large (e.g., in the hundreds or thousands), such a complexity would be impractical. For these reasons, we restrict ourselves to preprocessing operations which preserve the quasibanded structure of \mathbf{H} .

One admissible preprocessing operation is an n -place circular shift in column order of \mathbf{H} . Using the left circular shift matrix \mathbf{J} , the shifting operation transforms (8) into the equivalent system (19) with channel matrix $\mathbf{H}\mathbf{J}^{-n}$:

$$\mathbf{x} = (\mathbf{H}\mathbf{J}^{-n})\mathbf{J}^n\mathbf{s} + \mathbf{w}, \quad (19)$$

$$\mathbf{J} := \begin{pmatrix} \mathbf{0}_{L-1} & \mathbf{I}_{L-1} \\ 1 & \mathbf{0}_{L-1}^T \end{pmatrix}. \quad (20)$$

Though $\mathbf{H}\mathbf{J}^{-n}$ is not quasibanded in the sense of (6), the matrix $\tilde{\mathbf{H}}^T\tilde{\mathbf{H}} = \tilde{\mathbf{R}}^T\tilde{\mathbf{R}}$ is allowing the fast MMSE-GDFE processing from Section 3.1.1. Among the unique shifts $n \in \{0, \dots, L-1\}$, we choose the one which maximizes the norm of the rightmost column of $\mathbf{H}\mathbf{J}^{-n}$, that is, the norm of the rightmost column of $\tilde{\mathbf{R}}$. Thus, the PAM symbol contributing the most energy to \mathbf{x} is placed at the root of the tree. The complexity of this circular ordering stage is dominated by the evaluation of column norms, requiring $\mathcal{O}(DN)$ operations. We have observed, numerically, that this ‘‘circular ordering’’ scheme yields a modest improvement in terms of the performance/complexity tradeoff.

3.2. Channel-adaptive T -algorithm

In this section we propose a channel-adaptive version of the T -algorithm in which the threshold parameter T_i is adjusted at the i th level in the tree according to the channel realization and noise variance. Recall that the T -algorithm is a breadth-first search algorithm which, at the i th level, discards all partial paths $\mathbf{s}^{(i)}$ whose metric $\mathcal{M}(\mathbf{s}^{(i)})$ exceeds that of the best partial path $\mathbf{s}_*^{(i)} := \arg \min_{\mathbf{s}^{(i)}} \mathcal{M}(\mathbf{s}^{(i)})$ by an amount $\geq T_i$. (See Figure 3.) Thus, the T -algorithm will make a frame error if the *true* partial path $\mathbf{s}_T^{(i)}$ is discarded at any level $i \in \{L-1, L-2, \dots, 0\}$.

In our adaptive T -algorithm, we set the threshold T_i so that the true path is discarded with probability ϵ_o when the true path is not the best partial path:

$$\Pr \left\{ \mathcal{M}(\mathbf{s}_T^{(i)}) > \mathcal{M}(\mathbf{s}_*^{(i)}) + T_i \mid \mathcal{M}(\mathbf{s}_T^{(i)}) > \mathcal{M}(\mathbf{s}_*^{(i)}) \right\} < \epsilon_o. \quad (21)$$

Note that this is different from simply setting T_i so that the true path is discarded with probability ϵ_o . In the latter case, T_i will increase—thereby increasing search complexity—at low SNR. Intuition, however, tells us that it is not worthwhile to

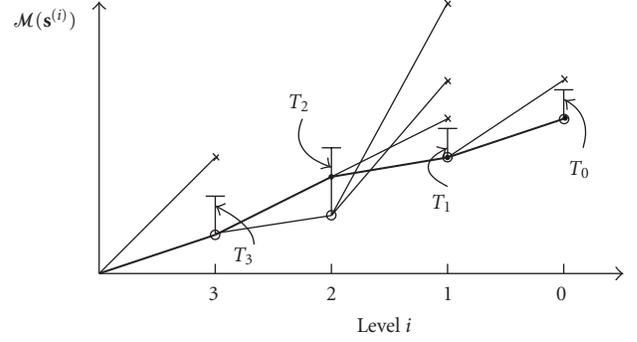


FIGURE 3: Illustration of path evolution in the T -algorithm when $Q = 2$ and $L = 4$. The circled points denote the minimum path metrics, the crossed points denote the discarded path metrics, and the bold line denotes the true path. Note that, in this example, $\mathcal{M}(\mathbf{s}_*^{(2)}) < \mathcal{M}(\mathbf{s}_T^{(2)})$.

search extensively at low SNR because, even if found, the ML path is more likely to be in error.

With $\mu^{(i)} := \mathcal{M}(\mathbf{s}_T^{(i)}) - \mathcal{M}(\mathbf{s}_*^{(i)})$, we can rewrite (21) as

$$\Pr \{ \mu^{(i)} > T_i \mid \mu^{(i)} > 0 \} < \epsilon_o. \quad (22)$$

We now analyze the random variable $\mu^{(i)}$. To do this, we define $\boldsymbol{\rho}^{(i)} := [\rho_i, \rho_{i+1}, \dots, \rho_{L-1}]^T$ and construct $\tilde{\mathbf{R}}^{(i)} \in \mathbb{R}^{(L-i) \times (L-i)}$ from the last $L-i$ rows and columns of $\tilde{\mathbf{R}}$, that is, $[\tilde{\mathbf{R}}^{(i)}]_{j,k} = [\tilde{\mathbf{R}}]_{j+i, k+i}$. This way, (16) can be written as $\mathcal{M}(\mathbf{s}^{(i)}) = \|\boldsymbol{\rho}^{(i)} - \tilde{\mathbf{R}}^{(i)}\mathbf{s}^{(i)}\|^2$. Using the error vector $\mathbf{e}^{(i)} := \mathbf{s}_*^{(i)} - \mathbf{s}_T^{(i)}$ and the interference vector $\mathbf{n}^{(i)} := \boldsymbol{\rho}^{(i)} - \tilde{\mathbf{R}}^{(i)}\mathbf{s}_T^{(i)}$, we find

$$\begin{aligned} \mu^{(i)} &= \|\boldsymbol{\rho}^{(i)} - \tilde{\mathbf{R}}^{(i)}\mathbf{s}_T^{(i)}\|^2 - \|\boldsymbol{\rho}^{(i)} - \tilde{\mathbf{R}}^{(i)}\mathbf{s}_*^{(i)}\|^2 \\ &= \|\mathbf{n}^{(i)}\|^2 - \|\mathbf{n}^{(i)} - \tilde{\mathbf{R}}^{(i)}\mathbf{e}^{(i)}\|^2 \\ &= 2\mathbf{n}^{(i)T}\tilde{\mathbf{R}}^{(i)}\mathbf{e}^{(i)} - \|\tilde{\mathbf{R}}^{(i)}\mathbf{e}^{(i)}\|^2. \end{aligned} \quad (23)$$

Since the statistics of $\mathbf{e}^{(i)}$ are difficult to characterize, we approximate $\mathbf{e}^{(i)}$ by the simple error event most likely to occur at the i th level, that is, an error vector of the form $\mathbf{e}^{(i)} = [0, \dots, 0, \pm 1, 0, \dots, 0]^T$. The partial metric $\mathcal{M}(\mathbf{s}^{(i)}) = \|\boldsymbol{\rho}^{(i)} - \tilde{\mathbf{R}}^{(i)}\mathbf{s}^{(i)}\|^2$ suggests that this error will occur at the index of the ‘‘weakest’’ column of $\tilde{\mathbf{R}}^{(i)}$. Thus we assume $[\mathbf{e}^{(i)}]_l = \pm \delta_{l-i}$ for

$$l_i := \arg \min_l \|\tilde{\mathbf{r}}_l^{(i)}\|, \quad (24)$$

where $\tilde{\mathbf{r}}_l^{(i)} \in \mathbb{R}^{L-i}$ denotes the l th column of $\tilde{\mathbf{R}}^{(i)}$. In this case,

$$\mu^{(i)} = \pm 2\mathbf{n}^{(i)T}\tilde{\mathbf{r}}_{l_i}^{(i)} - \|\tilde{\mathbf{r}}_{l_i}^{(i)}\|^2. \quad (25)$$

Recall from our discussion in Section 2.3 that the interference vector \mathbf{n} is zero-mean, white, and Gaussian in the case of ZF-GDFE preprocessing; and zero-mean, white, and

non-Gaussian in the case of MMSE-GDFE preprocessing. In the latter case, the non-Gaussianity of \mathbf{n} is due to a contribution from not-yet-detected PAM symbols, which we treat as random since their values are unknown when designing T_i . To proceed further, we approximate \mathbf{n} as Gaussian with covariance $\sigma_z^2 \mathbf{I}_L$. With these assumptions,

$$\mu^{(i)} \sim \mathcal{N}\left(-\|\tilde{\mathbf{r}}_i^{(i)}\|^2, 4\|\tilde{\mathbf{r}}_i^{(i)}\|^2 \sigma_z^2\right). \quad (26)$$

Using the statistical description (26), we can solve for T_i in (22) given a particular ϵ_o . From Bayes rule we find

$$\Pr\{\mu^{(i)} > T_i \mid \mu^{(i)} > 0\} = \begin{cases} \frac{\Pr\{\mu^{(i)} > T_i\}}{\Pr\{\mu^{(i)} > 0\}}, & T_i \geq 0, \\ 1, & \text{else,} \end{cases} \quad (27)$$

from which it is straightforward to show that

$$T_i = 2\sigma_z \|\tilde{\mathbf{r}}_i^{(i)}\| \mathcal{Q}^{-1}\left(\epsilon_o \mathcal{Q}\left(\frac{\|\tilde{\mathbf{r}}_i^{(i)}\|}{2\sigma_z}\right)\right) - \|\tilde{\mathbf{r}}_i^{(i)}\|^2 \quad (28)$$

using the tabulated function $\mathcal{Q}(x) := (1/\sqrt{2\pi}) \int_x^\infty e^{-x^2/2} dx$. From (28) we can see that the desired error probability ϵ_o is “weighted” by an SNR-dependent quantity; as SNR increases, so does the $\mathcal{Q}^{-1}(\cdot)$ term.

3.3. Channel estimation

Here we propose a rank-reduced pilot-aided Wiener channel estimation scheme. We discuss the pilot pattern first and the estimation scheme later.

We choose a pilot pattern where one out of every $P \geq 2$ multicarrier symbols is used as a pilot. These pilot symbols are then used to estimate the channel coefficients of the $P - 1$ multicarrier data symbols in-between. Pilot patterns of this form are relatively common, having been used in several other works (e.g., [10, 47]). We choose this pattern over one where each multicarrier symbol contains a mixture of pilot and data sub-carriers for the following reason. Assuming a significant ICI radius equal to D , the pilot and data sub-carriers would interfere unless a frequency-domain guard with radius $2D$ was placed around each pilot tone. Since Nyquist sampling considerations imply the need for at least N_h pilot tones, prevention of pilot/data interference would require that at least $(4D + 1)N_h$ sub-carriers are spared from data transmission. For many applications of interest (e.g., the setup in Section 4), however, $(4D + 1)N_h > N$, making this scheme impractical. Since the design of optimal pilot symbols appears to be a challenging problem, we used values obtained from a semiexhaustive search.

We now define some quantities that follow from our pilot pattern. Say that, for all indices m corresponding to pilot

symbols, we have $\mathbf{s}_m = \mathbf{p}$. For these m , (7) implies that

$$\begin{aligned} \mathbf{x}_m &= \mathbf{P}\mathbf{h}_m + \mathbf{w}_m, \\ \mathbf{h}_m &:= \left[\text{diag}_{-D}(\mathbf{H}_m^D)^T, \dots, \text{diag}_D(\mathbf{H}_m^D)^T \right]^T \in \mathbb{R}^{(2D+1)L}, \\ \mathbf{P} &:= \left(\mathbf{J}^D \mathcal{D}(\mathbf{p}) \quad \dots \quad \mathbf{J}^{-D} \mathcal{D}(\mathbf{p}) \right), \end{aligned} \quad (29)$$

where $\mathcal{D}(\cdot)$ transforms a vector argument into a diagonal matrix, and where $\text{diag}_k(\cdot)$ extracts the k th sub-diagonal of its matrix argument, that is, $\text{diag}_k(\mathbf{H}) := [[\mathbf{H}]_{k,0}, [\mathbf{H}]_{k+1,1}, \dots, [\mathbf{H}]_{k+L-1,L-1}]^T$ with modulo- L indexing assumed. Recall that \mathbf{J} was defined in (20). Our goal is to estimate the local-ICI coefficients $\vec{h}_m := [\mathbf{h}_{m+1}^T, \dots, \mathbf{h}_{m+P-1}^T]^T$ from the pilot observations $\vec{x}_m := [\mathbf{x}_m^T, \mathbf{x}_{m+P}^T]^T$. Say that $\mathbf{h}_m = \mathbf{C}\mathbf{g}_m$, where $\mathbf{g}_m \in \mathbb{C}^{N_b N_h}$ contains all complex-baseband time-domain impulse response coefficients that affect the m th observation, and where \mathbf{C} is a function of the MCM pulse shapes $\{a_n\}$ and $\{b_n\}$.

The linear MMSE estimate of \vec{h}_m from \vec{x}_m is [48]

$$\hat{\vec{h}}_m = \mathbf{R}_{hx} \mathbf{R}_{xx}^{-1} \vec{x}_m, \quad (30)$$

where $\mathbf{R}_{hx} := \mathbb{E}\{\vec{h}_m \vec{x}_m^T\}$ and $\mathbf{R}_{xx} := \mathbb{E}\{\vec{x}_m \vec{x}_m^T\}$. We can write

$$\begin{aligned} \mathbf{R}_{hx} &= \begin{pmatrix} \mathbf{R}_{hx}^1 & \mathbf{R}_{hx}^{1-P} \\ \mathbf{R}_{hx}^2 & \mathbf{R}_{hx}^{2-P} \\ \vdots & \vdots \\ \mathbf{R}_{hx}^{P-1} & \mathbf{R}_{hx}^{-1} \end{pmatrix}, \\ \mathbf{R}_{xx} &= \begin{pmatrix} \mathbf{R}_{xx}^0 & \mathbf{R}_{xx}^{-P} \\ \mathbf{R}_{xx}^P & \mathbf{R}_{xx}^0 \end{pmatrix}, \end{aligned} \quad (31)$$

with

$$\begin{aligned} \mathbf{R}_{hx}^q &:= \mathbf{C} \mathbb{E}\{\mathbf{g}_m \mathbf{g}_{m-q}^H\} \mathbf{C}^H \mathbf{P}^T, \\ \mathbf{R}_{xx}^q &:= \mathbf{P} \mathbf{C} \mathbb{E}\{\mathbf{g}_m \mathbf{g}_{m-q}^H\} \mathbf{C}^H \mathbf{P}^T + \delta_q \sigma_z^2 \mathbf{I}_{2L}. \end{aligned} \quad (32)$$

Note that $\mathbb{E}\{\mathbf{g}_m \mathbf{g}_{m-q}^H\}$ is easily calculated from the time-domain channel autocorrelation function.

Because each of the $2N_h$ real-valued channel taps changes slowly over the pilot/data/pilot interval (i.e., $N_b + PN_s$ channel uses), it contributes only $K = 1 + [2f_d T_c (N_b + PN_s)]$ nonnegligible singular values to $\mathbf{R}_{hx} \mathbf{R}_{xx}^{-1}$. Thus, as in [10], optimal rank reduction [48] can be used to significantly reduce the complexity of channel estimation with little performance degradation. The optimal rank- $2N_h K$ estimate of \vec{h}_m is constructed as follows [48]. From the SVD $\mathbf{R}_{hx} \mathbf{R}_{xx}^{-1} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H$, we build \mathbf{U}_K and \mathbf{V}_K from the first $2N_h K$ columns of \mathbf{U} and \mathbf{V} , respectively, and we build $\mathbf{\Sigma}_K$ from the first $2N_h K$ rows and columns of $\mathbf{\Sigma}$. We find that $\mathbf{R}_{hx} \mathbf{R}_{xx}^{-1} \approx \mathbf{U}_K \mathbf{F}_K^H$ for $\mathbf{U}_K \in \mathbb{R}^{(P-1)(2D+1)L \times 2N_h K}$ and $\mathbf{F}_K := \mathbf{V}_K \mathbf{\Sigma}_K \in \mathbb{C}^{2L \times 2N_h K}$. Note that \mathbf{U}_K can be interpreted as the MMSE-optimal order- $2N_h K$ basis expansion for \vec{h}_m and \mathbf{F}_K^H can be interpreted as

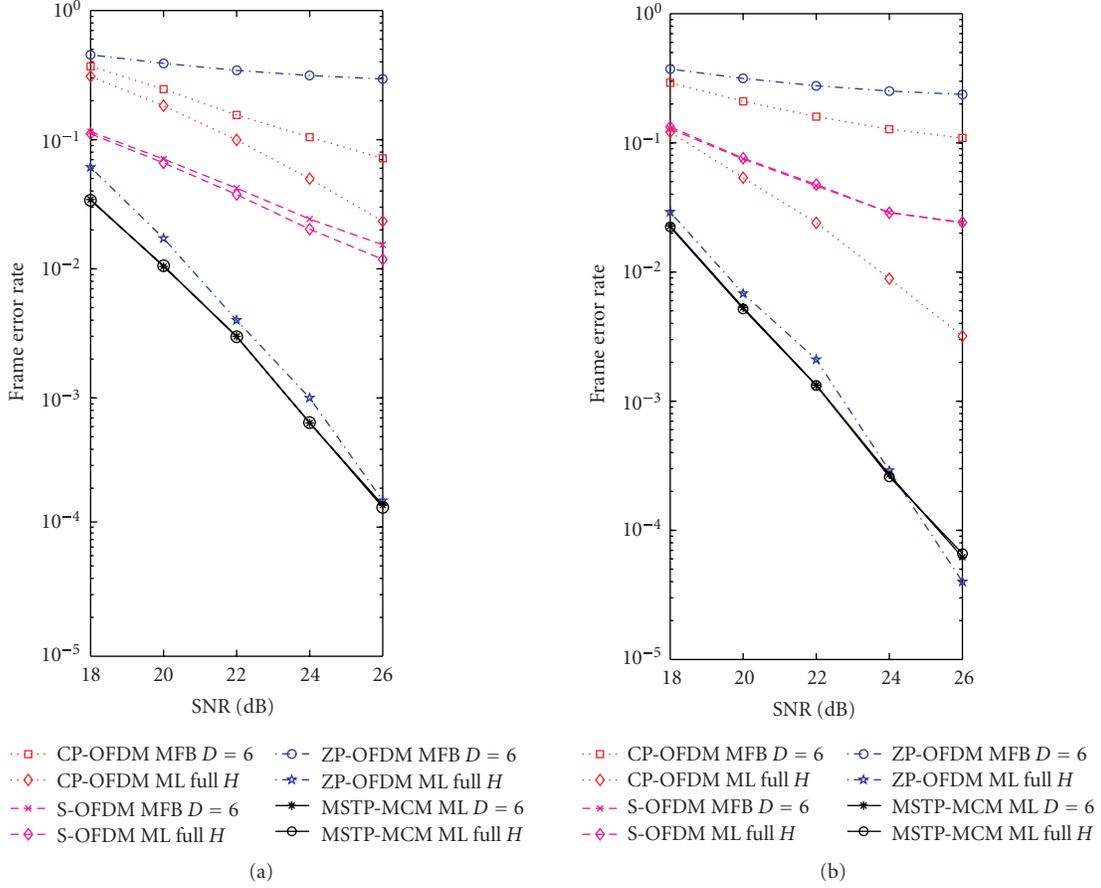


FIGURE 4: ML and MFB performance of several MCM schemes using global ICI (*full H*) or local ICI ($D = 6$) at (a) $f_d T_c = 0.001$; (b) $f_d T_c = 0.003$.

the linear MMSE estimator of the corresponding basis coefficients $\vec{\lambda}_m$. The resulting rank-reduced estimation procedure

$$\begin{aligned} \hat{\vec{\lambda}}_m &= \mathbf{F}_K^H \vec{x}_m, \\ \hat{\vec{g}}_m &= \mathbf{U}_K \hat{\vec{\lambda}}_m \end{aligned} \quad (33)$$

requires only $2N_h K [2L + (P-1)(2D+1)L]$ complex MACs per $P-1$ frames. In Section 4 we demonstrate that, with $K = 2$, the complexity of this channel estimation method is on par with that of preprocessed SqD. Experiments have confirmed that the rank-reduced performance is nearly indistinguishable from the full-rank performance [49].

4. NUMERICAL RESULTS

4.1. Setup

Our experiments employed the ICI/ISI-corrupted MCM system specified in complex-valued form by (4) and in real-valued form by (7). Uncoded QPSK symbols $\{s_{k,m}\}_{k=0}^{N-1}$ (i.e., $Q = 2$) were communicated over $N = 64$ MCM subcarriers (i.e., $L = 128$), and the demodulator outputs \mathbf{x}_m were used to detect the QPSK sequence \mathbf{s}_m . For SD, we focused on the case where only the “significant” ICI coefficients \mathbf{H}_m^D were known,

in which case ISI and residual ICI were treated as unknown interference.

Several methods of SD were examined: MLSD, near-ML SqD, and MMSE-DFE. In each case, we first apply circular ordering and fast MMSE-GDFE preprocessing to arrive at the detection problem (13), since, in the case of uncoded QPSK, solutions to (13) are known to be ML [38]. For MLSD, we solve (13) via SE-SpD, while for near-ML SqD, we obtain an approximate solution to (13) via suboptimal tree search. For MMSE-DFE, we decode the bits $\{s_{k,m}\}_{k=0}^{L-1}$ in the order $s_{L-1,m}, s_{L-2,m}, \dots, s_{0,m}$ by first making a hard decision on each bit and then subtracting its (estimated) contribution from \mathbf{x}_m [37].

We assumed a wide-sense stationary uncorrelated scattering (WSSUS) Rayleigh fading channel [50] whose realizations were generated using Jakes method. The channel had a uniform delay-profile with normalized⁷ delay spread $N_h = T_h/T_c = 16$ and a normalized single-sided Doppler spread $f_d T_c \in \{0.001, 0.003\}$. These parameters correspond to, for example, a system with subcarrier spacing $F_s = 20$ kHz, carrier frequency $f_c = 10$ GHz, delay spread $T_h = 12.25 \mu\text{s}$, and

⁷ These quantities are normalized to the “channel-use interval” or “chip interval,” $T_c = 1/NF_s$.

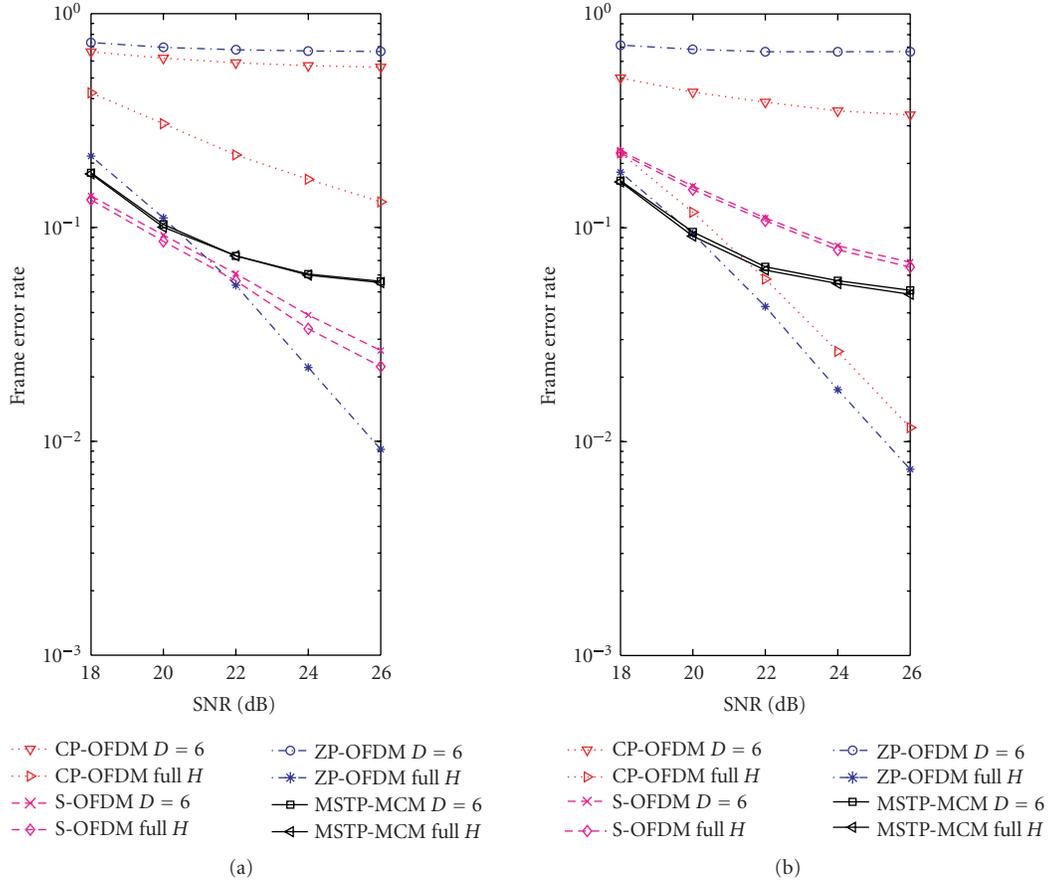


FIGURE 5: MMSE-DFE performance of several MCM schemes using global ICI (*full H*) or local ICI ($D = 6$) at (a) $f_d T_c = 0.001$; (b) $f_d T_c = 0.003$.

effective⁸ velocities of 138 km/h and 414 km/h, respectively. We defined SNR as the ratio of signal energy to noise energy in (pulse-shaped and sampled) receiver inputs.

Four FFT-based MCM schemes were considered: CP-OFDM [4], ZP-OFDM [5], Strohmer and Beaver’s “optimal” OFDM (S-OFDM) [6], and MSTP-MCM [14, 15]. Each of these schemes was allowed the same transmitted energy per information bit. With the exception of ZP-OFDM, all guarantee white subchannel noise samples (i.e., $E\{\mathbf{z}_m \mathbf{z}_m^T\} = \sigma_z^2 \mathbf{I}$). For CP-OFDM and ZP-OFDM, we employed a length- $N_g = 16$ guard to avoid ISI, yielding a spectral efficiency of 0.8 QPSK-symbols/s/Hz. For S-OFDM, $N = 64$ QPSK symbols were transmitted every 80 channel uses, also yielding a spectral efficiency of 0.8 QPSK-symbols/s/Hz. For MSTP-MCM, N QPSK symbols were transmitted every N channel uses, yielding a spectral efficiency of 1 QPSK-symbol/s/Hz. The dilation factor σ of the Gaussian pulse $g^\sigma(t)$ orthogonalized for S-OFDM (see [6, page 1114]) was numerically

optimized to minimize the total power of ICI plus ISI. The MSTP-MCM transmitter pulse $\{a_n\}$ was length $N_a = (3/2)N$ and the MSTP-MCM receiver pulse $\{b_n\}$ was length- N rectangular and preceded by $(3/2)N_h$ zeros.

4.2. FER performance

Figure 4 examines the frame⁹ error rate (FER) performance of the four MCM schemes with MLSD. When MLSD was too costly, the matched filter bound (MFB) was used as an approximation. When the MLSD has perfect *global*-ICI knowledge (i.e., knowledge of $\{\mathbf{H}_{m,0}\}$ in (5)), MSTP-MCM and ZP-OFDM performed similarly, and significantly outperformed S-OFDM and CP-OFDM. S-OFDM performed poorly due to a high level of ISI. Better S-OFDM performance was observed when the dilation factor was chosen to decrease ISI and increase ICI (which is incorporated in MLSD), but, since that was inconsistent with the S-OFDM design methodology in [6], we do not present those results here. CP-OFDM suffers from high FER because it wastes energy on a CP that is discarded by the receiver, and because CP-OFDM does

⁸ Effective velocity v can result from mobile velocity $v/3$ in, for example, a “triple-Doppler” situation, when the downlink signal bounces off of a reflector traveling directly towards (away from) the base station at velocity $v/3$ and is received by a mobile traveling away from (towards) the base station at velocity $v/3$.

⁹ We use the terms “frame” and “multicarrier symbol” interchangeably in this paper.

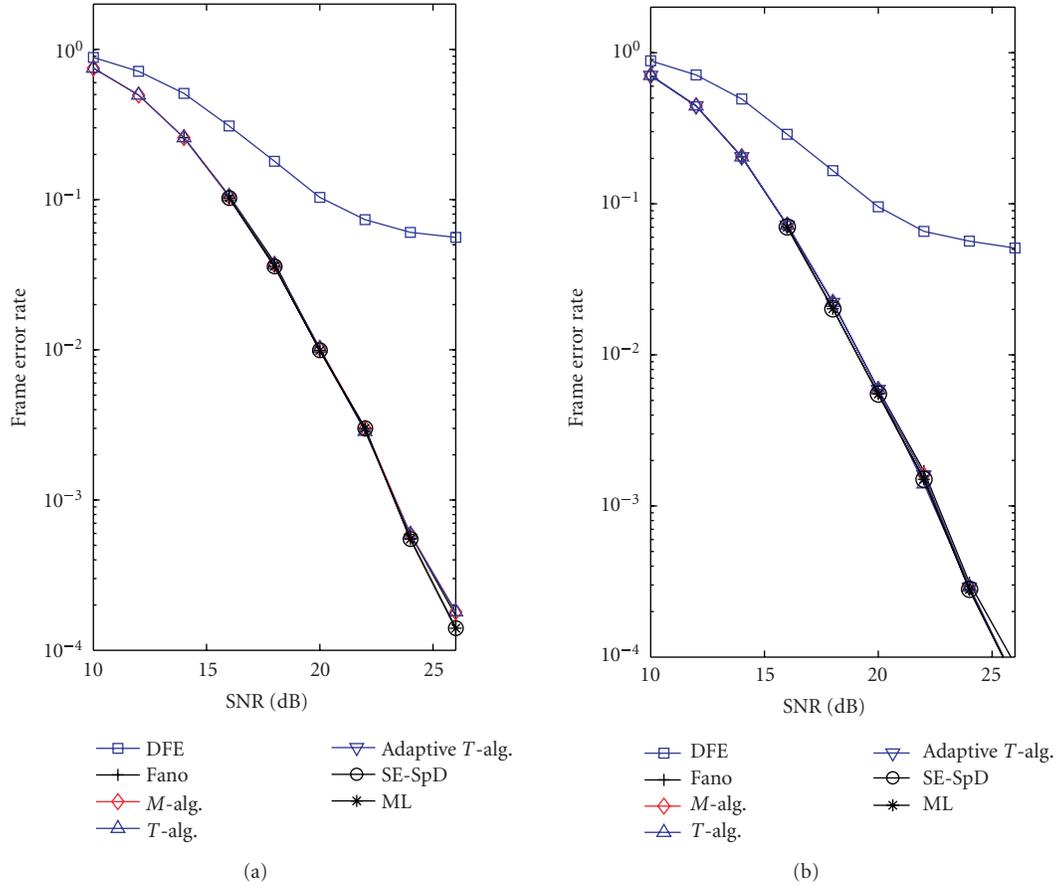


FIGURE 6: Performance of several SqDs on doubly dispersed MSTP-MCM with perfect knowledge of local ICI (i.e., $D = 6$) at (a) $f_d T_c = 0.001$; (b) $f_d T_c = 0.003$.

not make for easy extraction of delay diversity with uncoded transmissions. When using MLSD with global ICI knowledge, all MCMs schemes other than S-OFDM benefit from additional Doppler diversity at higher $f_d T_c$. S-OFDM, in contrast, reacts to the $f_d T_c$ increase by dilating the pulse in such a way that both ICI and ISI increase, but the increase in ISI hurts more than the increase in ICI helps. In comparing the schemes, it is important to remember that MSTP-MCM operates at 25% higher spectral efficiency than the other schemes.

When the receiver has only *local*-ICI knowledge up to ± 3 subcarriers (i.e., knowledge of \mathbf{H}_m^D for $D = 6$), Figure 4 shows that the MLSD performance of MSTP-MCM is indistinguishable from that with global-ICI knowledge. This confirms that MSTP-MCM suppresses nonlocal ICI well below the noise floor over the SNR range of interest. In contrast, the MLSD performance of ZP-OFDM and CP-OFDM collapse when only the local ICI is known; while S-OFDM avoids this collapse, it does so at the expense of high ISI power. Note that [19], which applies Viterbi SD to CP-OFDM under local-ICI knowledge, is lower-bounded by the *CP-OFDM MFB* $D = 6$ trace, and hence performs far worse than the proposed MSTP-MCM scheme.

Figure 5 examines FER performance of the four MCM schemes under MMSE-DFE detection. It is interesting to

note that, when the MMSE-DFE detector is given perfect global-ICI knowledge, the MSTP-MCM FER floors at high SNR. This is consistent with [11], which showed similar MMSE-DFE performance for max-SINR *reception*-pulse (MSRP) MCM. We conjecture that max-SINR pulse-shaped schemes are more prone to DFE error because the channel energy is not as well concentrated in the main diagonal of $\tilde{\mathbf{R}}$ as it is for CP-OFDM, ZP-OFDM, and S-OFDM. When the MMSE-DFE has only *local*-ICI knowledge up to ± 3 subcarriers, the FER performances of ZP-OFDM and CP-OFDM collapse, while the performance of MSTP-OFDM remains the same as that with global-ICI knowledge. As before, S-OFDM avoids this collapse, though at the cost of high ISI power. Once again, this confirms that MSTP-MCM suppresses nonlocal ICI well below the noise floor over the SNR range of interest. Note, however, that in all cases, the MMSE-DFE performances are significantly worse than their MLSD counterparts in Figure 4. Fortunately, there is no need to use MMSE-DFE detection on MSTP-MCM because (as we will see) SqD can be used to achieve near-ML performance with MMSE-DFE-like complexity.

Figure 6 shows the FER performance of various SqD algorithms operating on DD-channel MSTP-MCM with perfect knowledge of ± 3 subcarriers of local ICI (i.e., $D = 6$). For the *M*-algorithm, we set $M = 8$, and, for the

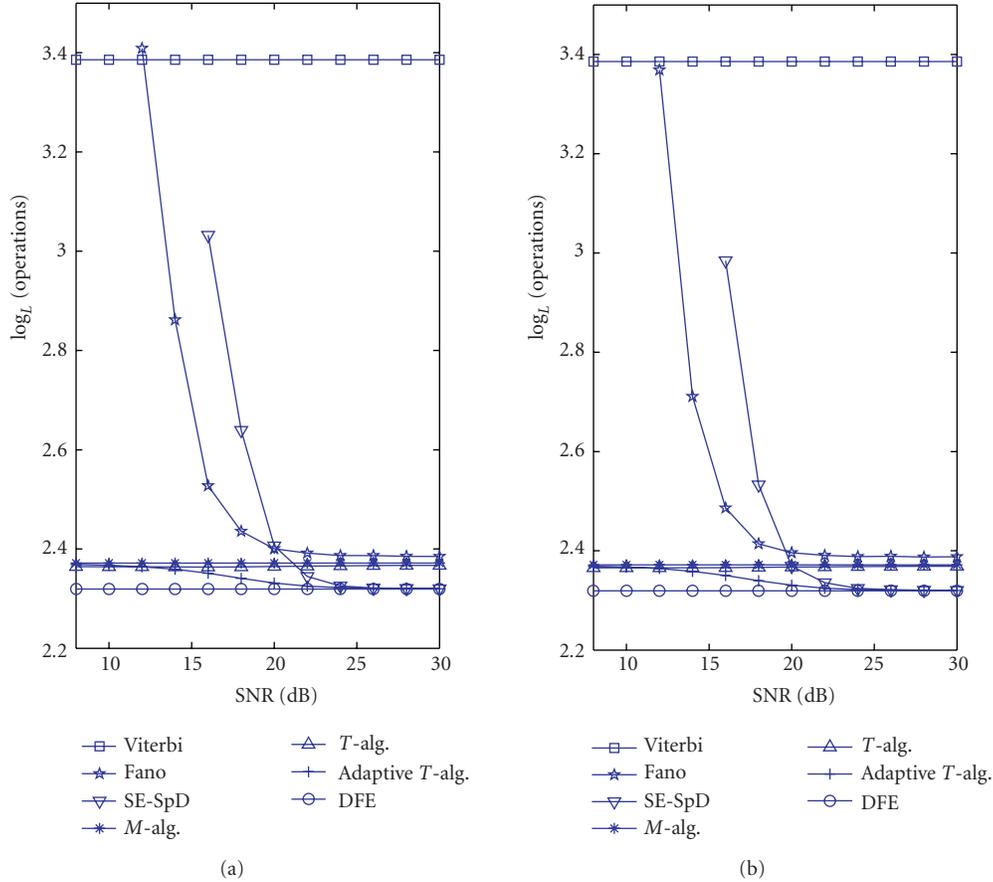


FIGURE 7: Number of real MAC operations per frame for doubly dispersed MSTP-MCM reception at (a) $f_d T_c = 0.001$; (b) $f_d T_c = 0.003$. Viterbi complexity above assumes inactive edge subcarriers (as in [19]); for active subcarriers, the tail-biting version would require $L^{5.10}$ MACs/frame.

T -algorithm, we set $T = 0.4$ when $f_d T_c = 0.001$, and $T = 0.5$ when $f_d T_c = 0.003$. For both algorithms, these were the computationally cheapest parameter settings which guaranteed near-ML performance over the SNR range of interest. For the adaptive T -algorithm, we set $\epsilon_o = 10^{-5}$ and limited the maximum list size to 8. For the Fano algorithm of [22], we set¹⁰ the bias $b = \sigma_z^2/2$ and the “step size” $= \sigma_z^2$. For SE-SpD, we initialized the squared sphere radius at 1.2 times the average squared Euclidean distance between the observation and the ML point. From Figure 6 we see that, with the exception of MMSE-DFE, all SqD algorithms give near-ML performance. In fact, the ML and SE-SpD traces are identical since MMSE-GDFE preprocessed SE-SpD yields ML performance with uncoded QPSK [38]. The MMSE-DFE error floor is consistent with that observed in [11] for MSRP-MCM.

4.3. Complexity

Figure 7 compares the average complexity of the SqD algorithms and the Viterbi algorithm used in [19] when operat-

ing on MSTP-MCM¹¹ with perfect knowledge of ± 3 subcarriers of local ICI (i.e., $D = 6$). Here “complexity” is measured in real MAC operations per decoded frame and is plotted on a log (base- L) scale, as in other near-ML SqD studies (e.g., [22, 36]). For the SqD algorithms, we plot the average number of operations required to achieve the FER performance of Figure 6, including that required for MSTP-MCM demodulation, circular ordering, and fast MMSE-GDFE preprocessing.

For the reasons discussed in Section 2.3, both the SE-SpD and Fano algorithms exhibit reasonable complexity at high SNR but explosive complexity at low SNR. As expected, the M -algorithm has the same complexity at all SNRs. Interestingly, the T -algorithm has almost the same complexity as the M -algorithm. Remarkably, the adaptive T -algorithm yields DFE-like complexity at high SNR and complexity that is not higher than that of the M - and T -algorithms at low SNR. As discussed in Section 2.3, the adaptive T -algorithm’s excellent low-SNR complexity is a consequence of its BrFS nature,

¹⁰ These settings were personally suggested by the authors of [22].

¹¹ Since Viterbi complexity is invariant to the channel realization, the values reported in Figure 7 apply equally well to CP-OFDM transmission, as was used in [19].

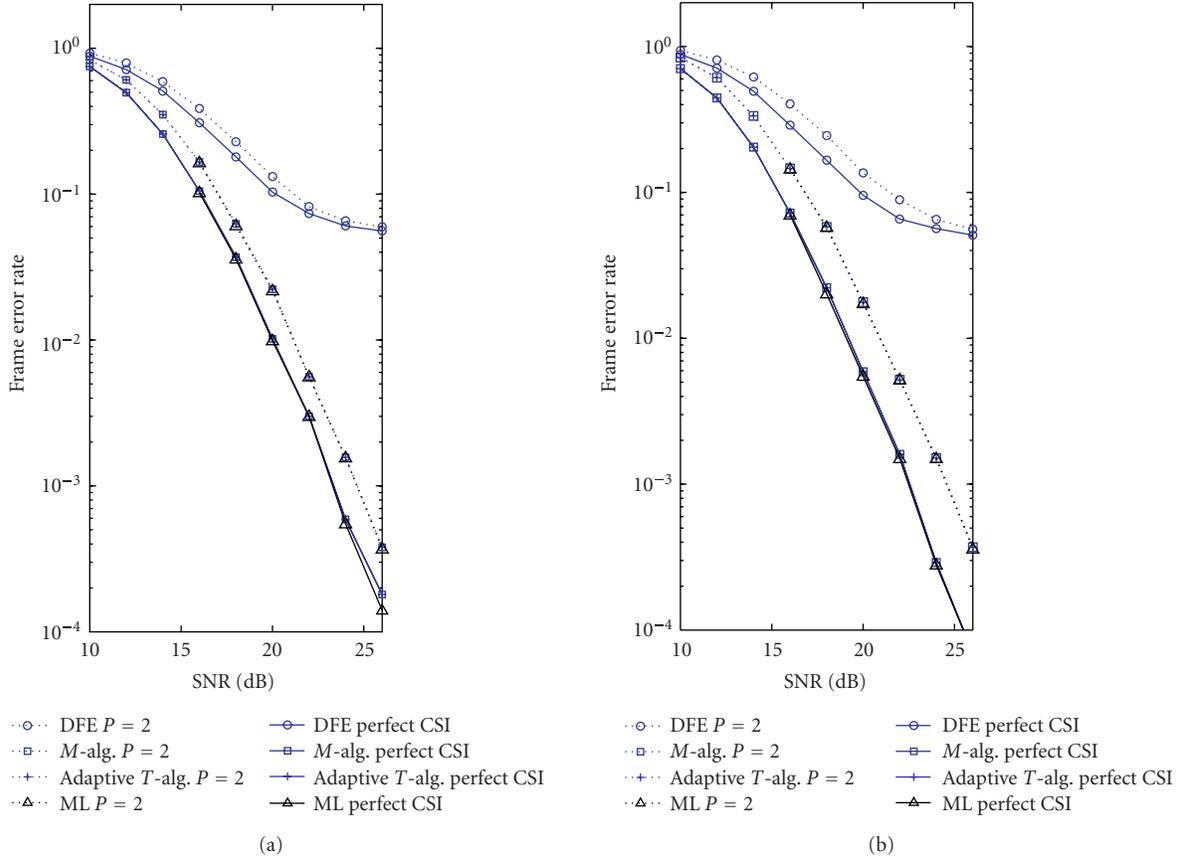


FIGURE 8: Performance of several SqDs operating on doubly dispersed MSTP-MCM with rank-reduced pilot-aided estimates of local-ICI (i.e., $D = 6$) at (a) $f_d T_c = 0.001$; (b) $f_d T_c = 0.003$, using pilot spacing $P = 2$.

while its excellent high-SNR complexity is a consequence of the fact that it uses channel knowledge to intelligently guide the search process.

The $D = 6$ Viterbi complexity is much larger than that of the near-ML BrFS algorithms and the MMSE-DFE. Furthermore, the Viterbi complexity plotted in Figure 7 (i.e., $L(2D + 1)Q^{(4D+1)} = L^{3.39}$) is valid only when D edge subcarriers are inactive. For our case, where all subcarriers are active, the “tail-biting” Viterbi approach proposed in Section 2.3 would be an admissible MLSD with complexity $L(2D + 1)Q^{(4D+1)} = L^{5.10}$ (which is well outside the range of Figure 7). In conclusion, Figure 7 shows that, by sacrificing a fraction of a dB in performance relative to MLSD, SqD can be implemented with near-MMSE-DFE complexity, even when all subcarriers are active.

4.4. Effect of imperfect channel estimates

In Figures 8, 9 we investigate the effect of channel estimation error on SqD performance in DD-channel MSTP-MCM. For this, we use the rank- $2N_h K$ pilot-aided Wiener estimator proposed in Section 3.3 with $K = 2$. With pilot-spacing $P = 2$, Figure 8 shows that imperfect channel estimates yield an SNR loss of about 1 dB at $f_d T_c = 0.001$ and about 2 dB at $f_d T_c = 0.003$ for all the SqD algorithms examined, that is, MMSE-DFE, the M -algorithm, the adaptive

T -algorithm, and MLSD. For our parameter choices, the channel estimation complexity reported in Section 3.3 translates into about $L^{2.4}$ real MACs per frame, which is comparable to the complexity of preprocessed SqD. With pilot-spacing $P = 3$, Figure 9 shows that imperfect channel estimates yield a significantly higher SNR loss, especially at $f_d T_c = 0.003$, making this choice impractical. A more thorough investigation of reduced-rank pilot-aided Wiener channel estimation appears in [49].

5. CONCLUSION

This paper proposed a means of high-spectral-efficiency MCM with practically realizable near-ML SqD suitable for communication over the DD channel. Our solution consisted of three components: (1) MCM that guarantees a small number of significant ICI/ISI coefficients while maintaining high spectral efficiency and white subchannel noise samples; (2) near-ML SD which leverages the MCM-shaped DD-channel ICI/ISI structure for low complexity over a wide SNR range; and (3) rank-reduced pilot-aided Wiener estimation of significant ICI/ISI coefficients. Because traditional MCM schemes (i.e., CP-OFDM, ZP-OFDM, and S-OFDM) were shown (numerically) to yield too many nonnegligible ISI/ICI coefficients, we utilized the MSTP-MCM scheme previously proposed by the authors in [14, 15]. MSTP-MCM

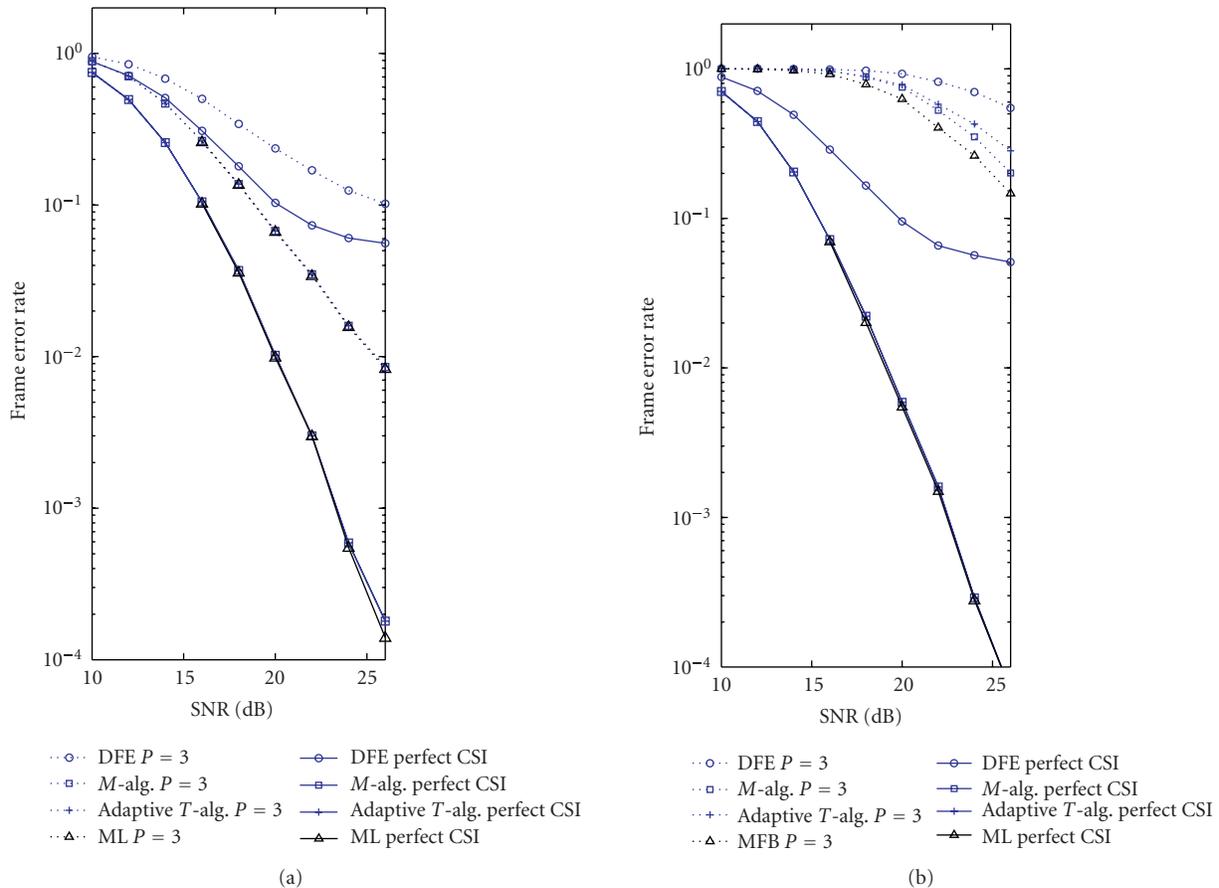


FIGURE 9: Performance of several SqDs operating on doubly dispersed MSTP-MCM with rank-reduced pilot-aided estimates of local-ICI (i.e., $D = 6$) at (a) $f_d T_c = 0.001$; (b) $f_d T_c = 0.003$, using pilot spacing $P = 3$.

combines an SINR-maximizing transmission pulse with a rectangular reception pulse, permitting ICI/ISI truncation with negligible loss in performance. Because traditional ML and near-ML SDs (e.g., Viterbi, Fano, and sphere decoders) were shown (numerically) to incur high search complexity when used with the MCM-shaped DD channel, a novel SD algorithm was proposed. The new SD combines a fast MMSE-GDFE preprocessor with a channel-adaptive T -algorithm, of which both components were specifically tuned to the MCM-shaped DD-channel's ICI/ISI structure. The new SD demonstrated FER performance indistinguishable from that of MLSD with a complexity that is approximately quadratic in the frame length, that is, similar to that of fast MMSE-DFE. Finally, the use of rank-reduced pilot-aided channel estimates was observed to cause only a small loss in FER performance (relative to perfect channel knowledge) without significantly increasing the overall receiver complexity.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation CAREER Grant CCR-0237037.

REFERENCES

- [1] X. Ma and G. B. Giannakis, "Maximum-diversity transmissions over doubly selective wireless channels," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1832–1840, 2003.
- [2] Z. Tang and G. Leus, "A receiver architecture for maximum diversity transmissions over doubly-selective channels," in *Proceedings of IEEE 6th Workshop on Signal Processing Advances in Wireless Communications (SPAWC '05)*, pp. 171–175, New York, NY, USA, June 2005.
- [3] J. A. C. Bingham, "Multicarrier modulation for data transmission: an idea whose time has come," *IEEE Communications Magazine*, vol. 28, no. 5, pp. 5–14, 1990.
- [4] L. J. Cimini Jr., "Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing," *IEEE Transactions on Communications*, vol. 33, no. 7, pp. 665–675, 1985.
- [5] B. Muquet, Z. Wang, G. B. Giannakis, M. De Courville, and P. Duhamel, "Cyclic prefixing or zero padding for wireless multi-carrier transmissions?" *IEEE Transactions on Communications*, vol. 50, no. 12, pp. 2136–2148, 2002.
- [6] T. Strohmer and S. Beaver, "Optimal OFDM design for time-frequency dispersive channels," *IEEE Transactions on Communications*, vol. 51, no. 7, pp. 1111–1122, 2003.

- [7] W. G. Jeon, K. H. Chang, and Y. S. Cho, "An equalization technique for orthogonal frequency-division multiplexing systems in time-variant multipath channels," *IEEE Transactions on Communications*, vol. 47, no. 1, pp. 27–32, 1999.
- [8] A. Stamoulis, S. N. Diggavi, and N. Al-Dhahir, "Intercarrier interference in MIMO OFDM," *IEEE Transactions on Signal Processing*, vol. 50, no. 10, pp. 2451–2464, 2002.
- [9] I. Barhumi, G. Leus, and M. Moonen, "Time-domain and frequency-domain per-tone equalization for OFDM over doubly selective channels," *Signal Processing*, vol. 84, no. 11, pp. 2055–2066, 2004.
- [10] Y.-S. Choi, P. J. Voltz, and F. A. Cassara, "On channel estimation and detection for multicarrier signals in fast and selective Rayleigh fading channels," *IEEE Transactions on Communications*, vol. 49, no. 8, pp. 1375–1387, 2001.
- [11] L. Rugini, P. Banelli, and G. Leus, "Block DFE and windowing for Doppler-affected OFDM systems," in *Proceedings of IEEE 6th Workshop on Signal Processing Advances in Wireless Communications (SPAWC '05)*, pp. 470–474, New York, NY, USA, June 2005.
- [12] X. Cai and G. B. Giannakis, "Bounding performance and suppressing intercarrier interference in wireless mobile OFDM," *IEEE Transactions on Communications*, vol. 51, no. 12, pp. 2047–2056, 2003.
- [13] P. Schniter, "Low-complexity equalization of OFDM in doubly selective channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 4, pp. 1002–1011, 2004.
- [14] S. Das and P. Schniter, "A new pulse shaped frequency division multiplexing technique for doubly dispersive channels," in *Proceedings of Conference Record of the 38th Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 657–661, Pacific Grove, Calif, USA, November 2004.
- [15] S. Das and P. Schniter, "Max-SINR ISI/ICI-shaping multicarrier modulation for the doubly dispersive channel," submitted to *IEEE Transactions on Signal Processing*.
- [16] K. Matheus and K.-D. Kammeyer, "Optimal design of a multicarrier systems with soft impulse shaping including equalization in time or frequency direction," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '97)*, vol. 1, pp. 310–314, Phoenix, Ariz, USA, November 1997.
- [17] K. Matheus, K. Knoche, M. Feuersänger, and K.-D. Kammeyer, "Two-dimensional (recursive) channel equalization for multicarrier systems with soft impulse shaping (MCSIS)," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '98)*, vol. 2, pp. 956–961, Sydney, NSW, Australia, November 1998.
- [18] K. Sigloch, M. R. Andrews, P. P. Mitra, and D. J. Thomson, "Communicating over nonstationary nonflat wireless channels," *IEEE Transactions on Signal Processing*, vol. 53, no. 6, pp. 2216–2227, 2005.
- [19] S. Ohno, "Maximum likelihood inter-carrier interference suppression for wireless OFDM with null subcarriers," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 849–852, Philadelphia, Pa, USA, March 2005.
- [20] A. J. Viterbi and J. Omura, *Principles of Digital Communication and Decoding*, McGraw-Hill, New York, NY, USA, 1979.
- [21] J. B. Anderson and S. Mohan, "Sequential coding algorithms: a survey and cost analysis," *IEEE Transactions on Communications*, vol. 32, no. 2, pp. 169–176, 1984.
- [22] A. Murugan, H. El Gamal, M. O. Damen, and G. Caire, "A unified framework for tree search decoding: rediscovering the sequential decoder," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 933–953, March 2006.
- [23] M. O. Damen, H. El Gamal, and G. Caire, "MMSE-GDFE lattice decoding for under-determined linear channels," in *Proceedings of 38th Annual Conference on Information Sciences and Systems (CISS '04)*, Princeton, NJ, USA, March 2004.
- [24] S. J. Simmons, "Breadth-first trellis decoding with adaptive effort," *IEEE Transactions on Communications*, vol. 38, no. 1, pp. 3–12, 1990.
- [25] H. Bölcskei, "Orthogonal frequency division multiplexing based on offset QAM," in *Advances in Gabor Theory*, H. G. Feichtinger and T. Strohmer, Eds., pp. 321–352, Birkhäuser, Boston, Mass, USA, 2002.
- [26] R. W. Chang, "Synthesis of band-limited orthogonal signals for multichannel data transmission," *Bell System Technical Journal*, vol. 45, pp. 1775–1796, 1966.
- [27] B. Le Floch, M. Alard, and C. Berrou, "Coded orthogonal frequency division multiplex," *Proceedings of the IEEE*, vol. 83, no. 6, pp. 982–996, 1995.
- [28] R. Haas and J.-C. Belfiore, "A time-frequency well-localized pulse for multiple carrier transmission," *Wireless Personal Communications*, vol. 5, no. 1, pp. 1–18, 1997.
- [29] W. Kozek and A. F. Molisch, "Nonorthogonal pulseshapes for multicarrier communications in doubly dispersive channels," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1579–1589, 1998.
- [30] H. Bölcskei, P. Duhamel, and R. Hleiss, "Design of pulse-shaping OFDM/OQAM systems for high data-rate transmission over wireless channels," in *Proceedings of IEEE International Conference on Communications (ICC '99)*, vol. 1, pp. 559–564, Vancouver, British Columbia, Canada, June 1999.
- [31] P. Schniter, "A new approach to multicarrier pulse design for doubly-dispersive channels," in *Proceedings of Annual Allerton Conference on Communication, Control, and Computing*, pp. 1012–1021, Monticello, Ill, USA, October 2003.
- [32] G. D. Forney Jr., "The Viterbi algorithm," *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, 1973.
- [33] W. H. Mow, "Maximum likelihood sequence estimation from the lattice viewpoint," *IEEE Transactions on Information Theory*, vol. 40, no. 5, pp. 1591–1600, 1994.
- [34] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1639–1642, 1999.
- [35] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Transactions on Information Theory*, vol. 48, no. 8, pp. 2201–2214, 2002.
- [36] M. O. Damen, H. El Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2389–2402, 2003.
- [37] N. Al-Dhahir and A. H. Sayed, "The finite-length multi-input multi-output MMSE-DFE," *IEEE Transactions on Signal Processing*, vol. 48, no. 10, pp. 2921–2936, 2000.
- [38] S.-J. Hwang and P. Schniter, "On the optimality of MMSE-GDFE pre-processed sphere decoding," in *Proceedings of Annual Allerton Conference on Communication, Control, and Computing*, Monticello, Ill, USA, October 2005.
- [39] H. El Gamal, G. Caire, and M. O. Damen, "Lattice coding and decoding achieve the optimal diversity-multiplexing tradeoff of MIMO channels," *IEEE Transactions on Information Theory*, vol. 50, no. 6, pp. 968–985, 2004.
- [40] R. Gowaikar and B. Hassibi, "Efficient statistical pruning for maximum likelihood decoding," in *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '03)*, vol. 5, pp. 49–52, Hong Kong, 2003.

- [41] J. M. Wozencraft and B. Reiffen, *Sequential Decoding*, MIT Press, Cambridge, Mass, USA; John Wiley & Sons, New York, NY, USA, 1961.
- [42] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice," *Mathematics of Computation*, vol. 44, no. 170, pp. 463–471, 1985.
- [43] R. M. Fano, "A heuristic discussion of probabilistic decoding," *IEEE Transactions on Information Theory*, vol. 9, no. 2, pp. 64–74, 1963.
- [44] H. Zamiri-Jafarian and S. Pasupathy, "Adaptive T-algorithm in mlsd/mlsde receivers for fading channels," in *Proceedings of IEEE International Conference on Communications (ICC '99)*, vol. 1, pp. 539–543, Vancouver, British Columbia, Canada, June 1999.
- [45] M. Kokkonen and K. Kalliojarvi, "Soft-decision decoding of binary linear block codes using reduced breadth-first search algorithms," *IEEE Transactions on Communications*, vol. 48, no. 6, pp. 905–907, 2000.
- [46] G. H. Golub and C. F. Van Loan, *Matrix Computations*, John Hopkins University Press, Baltimore, Md, USA, 3rd edition, 1996.
- [47] T. Zemen and C. F. Mecklenbräuker, "Time-variant channel estimation using discrete prolate spheroidal sequences," *IEEE Transactions on Signal Processing*, vol. 53, pp. 3597–3607, 2005.
- [48] L. L. Scharf, *Statistical Signal Processing*, Addison-Wesley, Reading, Mass, USA, 1991.
- [49] P. Schniter, "On doubly dispersive channel estimation for pilot-aided pulse-shaped multi-carrier modulation," in *Proceedings of Conference on Information Sciences and Systems (CISS '06)*, Princeton, NJ, USA, March 2006.
- [50] J. G. Proakis, *Digital Communications*, McGraw-Hill, New York, NY, USA, 4th edition, 2001.

received a Schlumberger Fellowship and an Intel Foundation Fellowship. He was awarded the 1999 Prize Paper Award from the IEEE Energy Development and Power Generation Committee for work relating to his M.S. thesis. In 2003, he received the National Science Foundation CAREER Award, and in 2005, the OSU College of Engineering Lumley Research Award. His areas of research include signal processing and communication theory.

Sung-Jun Hwang received the B.S. and M.S. degrees in electronic engineering from Yonsei University, Seoul, Korea, in 1993 and 1995, respectively. He worked as an Engineer for Samsung Electronics Co. Ltd., in Seoul, Korea, from 1995 to 2000, and Samsung Thales Co. Ltd., in Kiheung, Korea, from 2000 to 2003. Since 2004 he has been working toward the Ph.D. degree in electrical and computer engineering at the Ohio State University, Columbus, Ohio. His research interest is estimation and detection algorithms for multicarrier and MIMO communication systems.



Philip Schniter received the B.S. and M.S. degrees in electrical and computer engineering from the University of Illinois at Urbana Champaign in 1992 and 1993, respectively. From 1993 to 1996 he was employed by Tektronix Inc. in Beaverton, Ore as a Systems Engineer. In 2000, he received the Ph.D. degree in electrical engineering from Cornell University in Ithaca, NY. Subsequently, he joined the Department of Electrical and Computer Engineering at the Ohio State University in Columbus, Ohio, where he is now an Associate Professor. He currently serves as an Associate Editor for the IEEE Signal Processing Letters and sits on the IEEE Signal Processing for Communications Technical Committee. While pursuing his Ph.D. degree, He

