Low-Complexity Equalization of OFDM in Doubly Selective Channels

Philip Schniter, *Member, IEEE*

***Abstract*—Orthogonal frequency division multiplexing (OFDM) systems may experience significant inter-carrier interference (ICI) when used in time- and frequency-selective, or doubly selective, channels. In such cases, the classical symbol estimation schemes, e.g., minimum mean-squared error (MMSE) and zero-forcing (ZF) estimation, require matrix inversion that is prohibitively complex for large symbol lengths. An analysis of the ICI generation mechanism leads us to propose a novel two-stage equalizer whose complexity (apart from the FFT) is linear in the OFDM symbol length. The first stage applies optimal linear preprocessing to restrict ICI support, and the second stage uses iterative MMSE estimation to estimate finite-alphabet frequency-domain symbols. Simulation results indicate that our equalizer has significant performance and complexity advantages over the classical linear MMSE estimator in doubly selective channels.**

***Index Terms*—Doubly dispersive channel, doubly selective channel, equalization, intercarrier interference, OFDM.**

I. Introduction

ORTHOGONAL frequency division multiplexing (OFDM) [1], [2] has emerged as one of the most practical techniques for data communication over frequency-selective fading channels. In OFDM, the computationally-efficient fast Fourier transform (FFT) is used to transmit data in parallel over a large number of orthogonal subcarriers. When an adequate number of subcarriers are used in conjunction with a cyclic prefix of adequate length, subcarrier orthogonality is maintained, even in the presence of frequency-selective fading. Orthogonality implies a lack of subcarrier interference and permits simple, high-performance data detection.

In time- and frequency-selective—or *doubly selective*—fading, however, the orthogonality of OFDM is lost, leading to subcarrier interference that greatly complicates optimal data detection [3]–[8]. Historically, OFDM has been applied to scenarios in which time selectivity can be effectively ignored, but future wireless applications are expected to operate at high transmit-frequencies, at high levels of mobility, and at high capacities, resulting in fading that is doubly selective. Thus, the primary advantage of classical OFDM—interference-free operation—will not carry over to important future applications.

The following arguments more clearly explain the potential for doubly selective channels in future OFDM applications. First, as communication systems are implemented in higher frequency bands, they employ smaller wavelengths, implying that the their sensitivity to physical movement grows proportionally [9]. In other words, effective rates of channel variation for a fixed mobile-speed increase. Second, increasing either the efficiency or the bandwidth of OFDM systems will increase their sensitivity to channel variation. This latter claim can be understood from the desire to use a large OFDM symbol length that allows, in turn, significant channel variation within a symbol. Large symbol lengths are motivated by the desire to i) reduce capacity loss due to insertion of redundant guard intervals and ii) maintain narrow subcarrier spacing (to ensure flat subcarrier fading) as system bandwidth increases. In fact, symbol lengths of 4096 [10] and 8192 [11] are now common.

While the vast majority of OFDM literature ignores intrasymbol channel variation, several OFDM modifications have been proposed to cope with the resulting inter-carrier interference (ICI). Armstrong *et al.* [12] suggested the use of polynomial cancellation coding (PCC) [13], which is typically used to suppress ICI from carrier frequency offset. Because channel-variation yields a different ICI structure than does frequency offset, however, the applicability of PCC is quite limited.

Jeon *et al.* [14] ignored "small" ICI coefficients to reduce the complexity of linear minimum mean-squared error (MMSE) symbol estimation, and Choi *et al.* [15] presented matched-filter, least-square, and MMSE estimators that incorporate decision feedback. In [16], Cai and Giannakis combined [14] and [15] and derived recursive algorithms for calculation of the estimator coefficients. Linnartz and Gorokhov [17] used a two-term Taylor series expansion to linearly approximate time-domain channel variations and, from this, designed a linear MMSE estimator. Stamoulis *et al.* [18] examined the multiple-antenna case and derived a bank of LTV filters that maximize ratio of signal energy to ICI-plus-noise energy. The symbol estimators [15]–[18] require $\geq \mathcal{O}(N^2)$ complexity, where N denotes the OFDM symbol length, making them impractical for large N.

In this paper, we propose $\mathcal{O}(N)$-complexity symbol estimation strategies for OFDM systems in the presence of doubly selective fading. Rather than simply ignoring small-valued ICI coefficients, as in [14] and [16], we use signal-to-interference-plus-noise ratio (SINR)-optimal low-complexity linear preprocessing to squeeze ICI into a few coefficients. Then, we propose low-complexity iterative symbol estimation schemes that leverage the ICI-shortened channel representation. Full channel knowledge is assumed throughout; doubly selective channel estimation is treated elsewhere (see, e.g., [14], [15], [18], and [19]).

Manuscript received February 2, 2003; revised May 16, 2003. Portions of this work were presented at the 2002 Asilomar, Conference on Signals, Systems, and Computers, Pacific Grove, CA. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Zhi Ding.

The author is with the Department of Electrical Engineering, The Ohio State University, Columbus, OH 43210 USA (e-mail: schniter.1@osu.edu).

Digital Object Identifier 10.1109/TSP.2004.823503

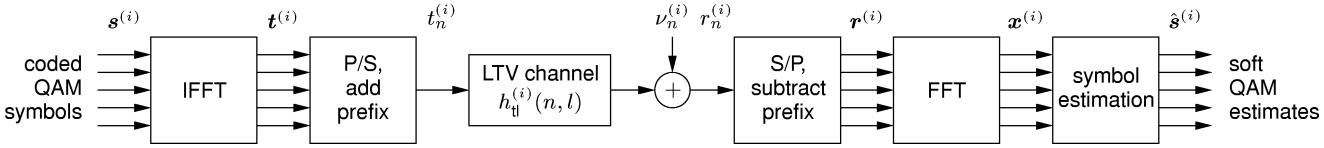

Fig. 1. OFDM system model.

Notation: We use $(\cdot)^t$ to denote transpose, $(\cdot)^*$ conjugate, and $(\cdot)^H$ conjugate transpose. $\mathcal{C}(\boldsymbol{b})$ denotes the circulant matrix with first column $\boldsymbol{b}$, $\mathcal{D}(\boldsymbol{b})$ the diagonal matrix created from vector $\boldsymbol{b}$, $\mathrm{diag}(\boldsymbol{B})$ the diagonal matrix with the same diagonal terms as matrix $\boldsymbol{B}$, and $\boldsymbol{I}$ the identity matrix. We use $[\boldsymbol{B}]_{m,n}$ to denote the element in the mth row and nth column of $\boldsymbol{B}$, where row/column indices begin with zero. $\|\cdot\|_F$ denotes the Frobenius norm, $(\cdot)^+$ the Moore–Penrose pseudo-inverse, and $\odot$ element-wise multiplication. Expectation is denoted by $\mathrm{E}\{\cdot\}$ and covariance by $\mathrm{Cov}\{\boldsymbol{b},\boldsymbol{c}\} := \mathrm{E}\{\boldsymbol{b}\boldsymbol{c}^H\} - \mathrm{E}\{\boldsymbol{b}\}\mathrm{E}\{\boldsymbol{c}^H\}$. Finally, $\delta(\cdot)$ denotes the Kronecker delta, $\langle\cdot\rangle_N$ the modulo-N operation, $*$ convolution, $\mathbb{R}$ the field of reals, and $\mathbb{Z}$ the set of integers.

II. System Model

Here, we review the OFDM system model, which is illustrated in Fig. 1. At each index $i \in \mathbb{Z}$, a set of N-coded QAM "frequency-domain" symbols $\{s_k^{(i)}\}$ is collected to form an OFDM symbol $\boldsymbol{s}^{(i)} = [s_0^{(i)},\ldots,s_{N-1}^{(i)}]^t$. The OFDM symbol is converted into the time-domain samples $\{t_n^{(i)}\}$ according the (norm-preserving) N-point inverse DFT operation

$$t_n^{(i)} = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} s_k^{(i)} e^{j\frac{2\pi}{N}kn}, \quad -N_p \le n < N \tag{1}$$

which are then serially transmitted over a noisy multipath channel. Note that $\{t_n^{(i)}\}$ incorporates a cyclic prefix of length $N_p \le N$. The multipath channel is modeled by the time-variant discrete impulse response $h_{\mathrm{tl}}(n,l)$, defined as the time-n response to an impulse applied at time $n-l$. A justification for this discrete model can be found in [18]. The channel response during the ith OFDM symbol interval is defined by

$$h_{\mathrm{tl}}^{(i)}(n,l) := h_{\mathrm{tl}}(iN + iN_p + n, l), \quad -N_p \le n < N.$$

Assuming a causal channel with maximum delay spread $N_h \le N_p$, the received samples collected during the ith OFDM symbol interval are

$$r_n^{(i)} = \sum_{l=0}^{N_h-1} h_{\mathrm{tl}}^{(i)}(n,l) t_{n-l}^{(i)} + \nu_n^{(i)}, \quad 0 \le n < N \tag{2}$$

where $\nu_n^{(i)}$ are samples of white Gaussian noise (AWGN) with variance σ^2. Note that $r_n^{(i)}$ contains contributions from only the ith transmitted symbol; this is a consequence of assuming that the multipath-corrupted cyclic prefix is discarded by the receiver. The receiver then computes an N-point DFT of $\{r_n^{(i)}\}$ [usually via the computationally-efficient fast Fourier transform (FFT) algorithm]:

$$x_d^{(i)} = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} r_n^{(i)} e^{-j\frac{2\pi}{N}dn}. \tag{3}$$

Using $\boldsymbol{F}$ to denote the N-point unitary DFT matrix, $\mathcal{H}_{\mathrm{tl}}^{(i)}$ to denote a (time-variant, circular) convolution matrix such that $[\mathcal{H}_{\mathrm{tl}}^{(i)}]_{n,l} := h_{\mathrm{tl}}^{(i)}(n,\langle n-l\rangle_N)$, and defining $\boldsymbol{r}^{(i)} := [r_0^{(i)},\ldots,r_{N-1}^{(i)}]^t$, $\boldsymbol{t}^{(i)} := [t_0^{(i)},\ldots,t_{N-1}^{(i)}]^t$, and $\boldsymbol{\nu}^{(i)} := [\nu_0^{(i)},\ldots,\nu_{N-1}^{(i)}]^t$, (2) can be written in vector form as follows:

$$\begin{aligned}\boldsymbol{r}^{(i)} &= \mathcal{H}_{\mathrm{tl}}^{(i)}\boldsymbol{t}^{(i)} + \boldsymbol{\nu}^{(i)} \\ &= \mathcal{H}_{\mathrm{tl}}^{(i)}\boldsymbol{F}^H\boldsymbol{s}^{(i)} + \boldsymbol{\nu}^{(i)}.\end{aligned} \tag{4}$$

Defining the *subcarrier coupling matrix* $\mathcal{H}_{\mathrm{df}}^{(i)}$

$$\mathcal{H}_{\mathrm{df}}^{(i)} := \boldsymbol{F}\mathcal{H}_{\mathrm{tl}}^{(i)}\boldsymbol{F}^H \tag{5}$$

and $\boldsymbol{x}^{(i)} := [x_0^{(i)},\ldots,x_{N-1}^{(i)}]^t$, (3) can be written

$$\boldsymbol{x}^{(i)} = \boldsymbol{F}\boldsymbol{r}^{(i)} \tag{6}$$

$$= \mathcal{H}_{\mathrm{df}}^{(i)}\boldsymbol{s}^{(i)} + \boldsymbol{w}^{(i)} \tag{7}$$

where $\boldsymbol{w}^{(i)} = \boldsymbol{F}\boldsymbol{\nu}^{(i)}$. Using $h_{\mathrm{tl}}^{(i)}(n,l) = 0$ for $l \ge N_h$, it is straightforward to show that $[\mathcal{H}_{\mathrm{df}}^{(i)}]_{d,k} = h_{\mathrm{df}}^{(i)}(d-k,k)$, where

$$h_{\mathrm{df}}^{(i)}(d,k) := \frac{1}{N}\sum_{n=0}^{N-1}\sum_{l=0}^{N-1} h_{\mathrm{tl}}^{(i)}(n,l) e^{-j\frac{2\pi}{N}(lk+dn)}. \tag{8}$$

Note that $\{h_{\mathrm{df}}^{(i)}(0,:)\}$ appears on the main diagonal of $\mathcal{H}_{\mathrm{df}}^{(i)}$, $\{h_{\mathrm{df}}^{(i)}(1,:)\}$ on the first sub-diagonal, $\{h_{\mathrm{df}}^{(i)}(-1,:)\}$ on the first super-diagonal, and so on. This fact and (7) imply that $h_{\mathrm{df}}^{(i)}(d,k)$ can be interpreted as the frequency-domain response, at subcarrier $k+d$, to a frequency-domain impulse centered at subcarrier k. In $h_{\mathrm{df}}^{(i)}(d,k)$, d can be interpreted as "Doppler" index and k as the "frequency" index. In $h_{\mathrm{tl}}^{(i)}(n,l)$, n can be interpreted as the "time" index and l as the "lag" index.

We assume the typical wide-sense stationary uncorrelated scattering (WSSUS) model [9] such that

$$\mathrm{E}\{h_{\mathrm{tl}}(n,l)h_{\mathrm{tl}}^*(n-q,l-m)\} = r_{\mathrm{t}}(q)\sigma_l^2\delta(m). \tag{9}$$

In (9), $r_{\mathrm{t}}(q)$ denotes the normalized tap autocorrelation (where $r_{\mathrm{t}}(0)=1$), and σ_l^2 denotes the variance of the lth tap.

III. ICI-Generating Mechanism

A nondiagonal subcarrier coupling matrix introduces ICI, complicating the symbol estimation task. To understand properties of the ICI, we examine the variance of the subcarrier coupling coefficients $\{h_{\mathrm{df}}^{(i)}(d,k)\}$. Using (8) and (9), dropping the symbol index i for brevity, and defining the N-point rectangular window

$$u_n := \begin{cases} 1, & 0 \le n < N \\ 0, & \text{else} \end{cases}$$

we find (using doubly infinite sums unless otherwise noted)

$$\begin{aligned}\mathrm{E}\left\{|h_{\mathrm{df}}(d,k)|^2\right\} &= \frac{1}{N^2}\sum_{n,l,m,p} u_n u_m \mathrm{E}\left\{h_{\mathrm{tl}}(n,l)h_{\mathrm{tl}}^*(m,p)\right\} \\ &\quad \times e^{j\frac{2\pi}{N}(pk-lk+md-nd)} \\ &= \frac{1}{N^2}\sum_l \sigma_l^2 \sum_{n,m} u_n u_m r_{\mathrm{t}}(n-m) \\ &\quad \times e^{j\frac{2\pi}{N}(md-nd)} \\ &= \frac{1}{N^2}\sum_l \sigma_l^2 \sum_q \left(\sum_n u_n u_{n-q}\right) r_{\mathrm{t}}(q) \\ &\quad \times e^{-j\frac{2\pi}{N}qd}.\end{aligned}$$

Note that $\mathrm{E}\{|h_{\mathrm{df}}(d,k)|^2\}$ is not a function of k. With the definition of the $(2N-1)$-point triangular window

$$v_q := \begin{cases} N-|q|, & -N < q < N \\ 0, & \text{else} \end{cases}$$

we can write

$$\mathrm{E}\left\{|h_{\mathrm{df}}(d,\cdot)|^2\right\} = \frac{1}{N^2}\sum_l \sigma_l^2 \sum_q v_q r_{\mathrm{t}}(q) e^{-j\left(\frac{2\pi}{N}d\right)q} \quad (10)$$

$$= \left(S(\phi) * V(\phi)\right)\big|_{\phi=\frac{2\pi}{N}d} \cdot \sum_l \sigma_l^2. \quad (11)$$

In (11), $S(\phi)$ denotes the Doppler spectrum

$$S(\phi) := \sum_q r_{\mathrm{t}}(q) e^{-j\phi q}, \quad \phi \in \mathbb{R}$$

and $V(\phi)$ the DTFT of $v_n N^{-2}$:

$$\begin{aligned} V(\phi) &:= \frac{1}{N^2}\sum_q v_q e^{-j\phi q}, \quad \phi \in \mathbb{R} \\ &= \left(\frac{\sin\left(\frac{\phi N}{2}\right)}{N\sin\left(\frac{\phi}{2}\right)}\right)^2 \end{aligned}$$

commonly known as a Dirichlet sinc.

Equation (11) gives an interpretation of the ICI-generating mechanism that has been depicted in Fig. 2. Essentially, the Doppler spectrum $S(\phi)$ is convolved with the Dirichlet sinc $V(\phi)$ and then sampled on the regular grid $\{\phi : \phi = (2\pi/N)d, d \in \mathbb{Z}\}$. With a linear time-invariant (LTI) channel,[1] i.e., zero Doppler spread, the nulls of $S(\phi) * V(\phi)$ fall on the grid, implying

$$\mathrm{E}\left\{|h_{\mathrm{df}}(d,\cdot)|^2\right\} = \delta\left(\langle d\rangle_N\right)\sum_l \sigma_l^2.$$

With a linear time-variant (LTV) channel, i.e., nonzero Doppler spread, the nulls of $S(\phi) * V(\phi)$ no longer fall on the grid, implying ICI.

In the case of Rayleigh fading [9], we have

$$\begin{aligned} r_{\mathrm{t}}(q) &= J_0(2\pi f_{\mathrm{d}} q) \\ S(\phi) &= \begin{cases} \frac{1}{\sqrt{(2\pi f_{\mathrm{d}})^2 - \phi^2}}, & |\phi| < 2\pi f_{\mathrm{d}} \\ 0, & |\phi| > 2\pi f_{\mathrm{d}} \end{cases} \end{aligned}$$

[1]The LTI case can also be understood by evaluating (10) with $r_{\mathrm{t}}(q) = 1$.

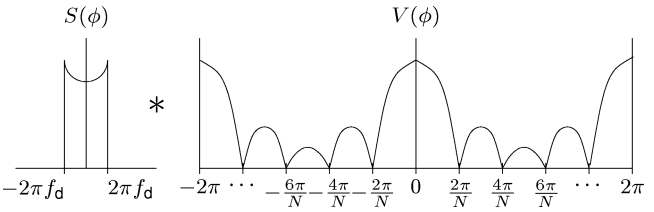

Fig. 2. Illustration of ICI generation mechanism (11). The spectra are convolved and then sampled on the regular grid $\{\phi = (2\pi/N)d, d \in \mathbb{Z}\}$. Without doppler spread, the samples fall on the sinc nulls, indicating zero ICI.

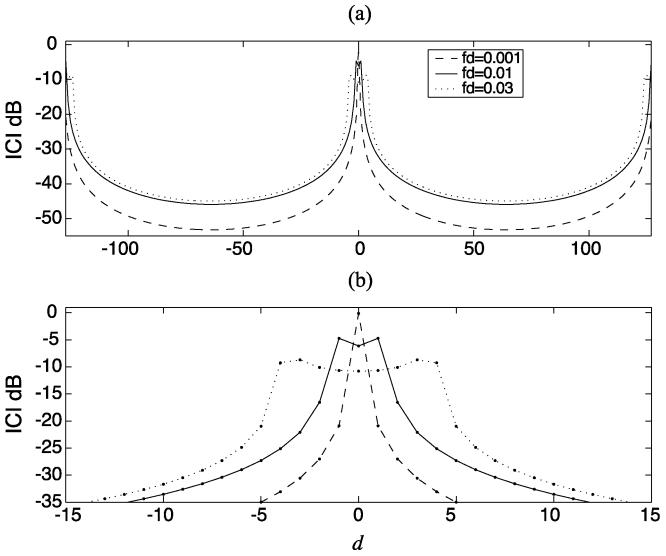

Fig. 3. ICI variance $\mathrm{E}\{|h_{\mathrm{df}}(d,\cdot)|^2\}$ versus d for $N = 128$, WSSUS Rayleigh fading, and various f_{d}. Subplot (b) provides a zoomed view.

where $J_0(\cdot)$ denotes the zeroth-order Bessel function of the first kind, and f_{d} denotes the maximum Doppler frequency normalized to the *signaling* rate (rather than the OFDM symbol rate). Fig. 3 plots $\mathrm{E}\{|h_{\mathrm{df}}(d,\cdot)|^2\}$ as a function of d, assuming $N = 128$, Rayleigh fading, and various f_{d}. Here, we see that even a Doppler frequency equal to approximately one DFT bin width (i.e., $f_{\mathrm{d}} = 0.01 \approx 1/N$) induces widespread ICI. This finding contradicts the claim (e.g., [14], [16], [20]) that the approximation $h_{\mathrm{df}}(d,\cdot) \approx 0$ for $d \geq \lceil f_{\mathrm{d}} N \rceil$ results in an accurate channel model. In fact, Section V demonstrates the significant performance degradations that result from equalizer design based on this approximation. We stress that finite-duration observation effects (manifested as the sinc in Fig. 2) play a critical role in ICI generation and cannot be ignored; the *time-frequency uncertainty principle* [21] strikes again.

IV. Symbol Estimation

From the observation $\boldsymbol{x}^{(i)}$ in (7), the receiver attempts to detect the true symbol $\boldsymbol{s}^{(i)}$. We assume a detection procedure which consists of an estimation (i.e., equalization) stage following by a decoding stage. As the decoding procedure depends on the specific coding scheme employed by the transmitter, it is outside the scope of this paper. Instead, we focus on symbol estimation. Since the decoding performance is expected to be proportional to the *subcarrier-averaged* SINR [22], we employ this criterion in the design of our estimators.

A. Classical Methods

The linear MMSE and zero-forcing (ZF) estimates [15] are given in (12) and (13), respectively, assuming $\mathrm{E}\{\boldsymbol{s}^{(i)}\} = \boldsymbol{0} = \mathrm{E}\{\boldsymbol{w}^{(i)}\}$, $\mathrm{E}\{\boldsymbol{s}^{(i)}\boldsymbol{s}^{(i)H}\} = \boldsymbol{I}$, $\mathrm{E}\{\boldsymbol{s}^{(i)}\boldsymbol{w}^{(i)H}\} = \boldsymbol{0}$, $\mathrm{E}\{\boldsymbol{w}^{(i)}\boldsymbol{w}^{(i)H}\} = \sigma^2\boldsymbol{I}$, and knowledge of the channel.

$$\hat{\boldsymbol{s}}^{(i)}_{\mathrm{mmse}} = \left(\mathcal{H}^{(i)H}_{\mathrm{df}}\mathcal{H}^{(i)}_{\mathrm{df}} + \sigma^2\boldsymbol{I}\right)^{-1}\mathcal{H}^{(i)H}_{\mathrm{df}}\boldsymbol{x}^{(i)} \tag{12}$$

$$\hat{\boldsymbol{s}}^{(i)}_{\mathrm{zf}} = \left(\mathcal{H}^{(i)}_{\mathrm{df}}\right)^{+}\boldsymbol{x}^{(i)}. \tag{13}$$

With an LTI channel, $\mathcal{H}^{(i)}_{\mathrm{df}}$ is diagonal, and both (12) and (13) can be implemented in $\mathcal{O}(N)$ operations; this simple "frequency-domain equalization" is the classical motivation for the use of OFDM. With an LTV channel, the MMSE and ZF estimators (12) and (13) require nontrivial matrix inversion. Inversion algorithms that make use of the Hermitian Toeplitz structure in (12) still require at least $\mathcal{O}(N^2)$ operations [23], making them impractical for large N.

B. Linear Preprocessing

In place of $\mathcal{O}(N^2)$ matrix multiplication, we propose low-complexity $\mathcal{O}(N)$ linear preprocessing that renders the ICI response *sparse*, thereby simplifying subsequent symbol estimation. The ICI-generating mechanism described in Section III suggests preprocessing that "squeezes" the significant coefficients of $\mathcal{H}_{\mathrm{df}}$ into the $2D+1$ central diagonals, a $D \times D$ lower triangular matrix in the bottom-left corner, and a $D \times D$ upper triangular matrix in the top-right corner, illustrated by the shaded regions in Fig. 4. The parameter $D \in \{0, \ldots, (N/2)-1\}$ controls the target ICI-response length: larger D corresponds to a longer ICI span and, thus, increased estimation complexity. In Section V, we find that $D = \lceil f_{\mathrm{d}}N \rceil + 1$ is an appropriate choice for Rayleigh fading. In general, D should be chosen proportional to the width of the Doppler spectrum $S(\phi)$.

ICI-response shortening can be regarded as the frequency-domain dual of inter-symbol interference (ISI)-response shortening, which is a well-known means of reducing the complexity of maximum likelihood sequence detection (MLSD) in single-carrier systems [24].

1) Time-Domain Windowing: While single-carrier systems typically achieve ISI-shortening via convolutive linear filtering, we leverage the receiver's FFT operation to achieve ICI-shortening via fast convolution [25], thereby saving computations when N is large. In matrix notation, the fast convolution property can be written [26]

$$\mathcal{C}(\boldsymbol{g}) = \boldsymbol{F}\mathcal{D}(\sqrt{N}\boldsymbol{F}^H\boldsymbol{g})\boldsymbol{F}^H. \tag{14}$$

Using $\boldsymbol{\beta}$ to denote the shortening filter's impulse response, the ICI-shortened observation takes the form

$$\begin{aligned}\breve{\boldsymbol{x}}^{(i)} &= \mathcal{C}(\boldsymbol{\beta})\boldsymbol{x}^{(i)} \\ &= \mathcal{C}(\boldsymbol{\beta})\mathcal{H}^{(i)}_{\mathrm{df}}\boldsymbol{s}^{(i)} + \mathcal{C}(\boldsymbol{\beta})\boldsymbol{w}^{(i)}\end{aligned} \tag{15}$$

where we desire that $\mathcal{C}(\boldsymbol{\beta})\mathcal{H}^{(i)}_{\mathrm{df}}$ has the structure illustrated in Fig. 4. Since the operation $\mathcal{C}(\boldsymbol{\beta})$ applies the same filtering to each column of $\mathcal{H}^{(i)}_{\mathrm{df}}$, ICI-shortening will not be accomplished perfectly. Still, Section V shows that good results can be obtained. While perfect ICI shortening is possible [consider, e.g., the ZF estimator (13)], it generally requires $\mathcal{O}(N^3)$ operations.

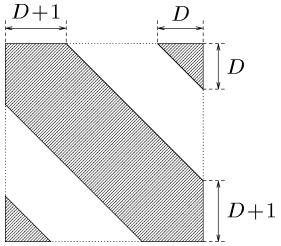

Fig. 4. Desired structure of windowed subcarrier coupling matrix $\breve{\mathcal{H}}^{(i)} = \mathcal{C}(\mathcal{B})\mathcal{H}^{(i)}_{\mathrm{df}}$.

Defining the time-domain vector

$$\boldsymbol{b} := \sqrt{N}\boldsymbol{F}^H\boldsymbol{\beta}$$

(14) implies that $\mathcal{C}(\boldsymbol{\beta}) = \boldsymbol{F}\mathcal{D}(\boldsymbol{b})\boldsymbol{F}^H$ so that

$$\breve{\boldsymbol{x}}^{(i)} = \boldsymbol{F}\mathcal{D}(\boldsymbol{b})\boldsymbol{r}^{(i)} \tag{16}$$

which indicates that the linear preprocessing operation $\mathcal{C}(\boldsymbol{\beta})$ can be implemented by an N-point windowing of the time-domain observation $\boldsymbol{r}^{(i)}$. While other low-complexity linear preprocessing methods are possible (e.g., replacing $\mathcal{D}(\boldsymbol{b})$ with a tridiagonal matrix), we restrict our attention to (16).

2) Max-SINR Window Design: The window coefficients $\boldsymbol{b}$ are designed to maximize the subcarrier-averaged SINR. To aid in the definition of "signal" and "interference," we partition the subcarrier coupling matrix into desired-ICI and undesired-ICI components

$$\mathcal{C}(\boldsymbol{\beta})\mathcal{H}^{(i)}_{\mathrm{df}} = \mathcal{M}_{\mathrm{D}}\left(\mathcal{C}(\boldsymbol{\beta})\mathcal{H}^{(i)}_{\mathrm{df}}\right) + \overline{\mathcal{M}}_{\mathrm{D}}\left(\mathcal{C}(\boldsymbol{\beta})\mathcal{H}^{(i)}_{\mathrm{df}}\right)$$

where $\mathcal{M}_{\mathrm{D}}(\cdot)$ denotes a mask operator that passes the shaded region and zeros the nonshaded region in Fig. 4, and $\overline{\mathcal{M}}_{\mathrm{D}}(\cdot)$ denotes its complement. It follows that $\breve{\boldsymbol{x}}^{(i)}$ can be written

$$\breve{\boldsymbol{x}}^{(i)} = \underbrace{\mathcal{M}_{\mathrm{D}}\left(\mathcal{C}(\boldsymbol{\beta})\mathcal{H}^{(i)}_{\mathrm{df}}\right)\boldsymbol{s}^{(i)}}_{\text{signal}} + \underbrace{\overline{\mathcal{M}}_{\mathrm{D}}\left(\mathcal{C}(\boldsymbol{\beta})\mathcal{H}^{(i)}_{\mathrm{df}}\right)\boldsymbol{s}^{(i)} + \mathcal{C}(\boldsymbol{\beta})\boldsymbol{w}^{(i)}}_{\text{interference + noise}}.$$

Using the assumptions on $\boldsymbol{s}^{(i)}$ and $\boldsymbol{w}^{(i)}$ from Section IV-A, signal energy $\mathcal{E}^{(i)}_s$ and noise-plus-interference energy $\mathcal{E}^{(i)}_{ni}$ become

$$\mathcal{E}^{(i)}_s = \left\|\mathcal{M}_{\mathrm{D}}\left(\mathcal{C}(\boldsymbol{\beta})\mathcal{H}^{(i)}_{\mathrm{df}}\right)\right\|^2_F \tag{17}$$

$$\mathcal{E}^{(i)}_{ni} = \left\|\overline{\mathcal{M}}_{\mathrm{D}}\left(\mathcal{C}(\boldsymbol{\beta})\mathcal{H}^{(i)}_{\mathrm{df}}\right)\right\|^2_F + \sigma^2\left\|\mathcal{C}(\boldsymbol{\beta})\right\|^2_F. \tag{18}$$

The window coefficients $\boldsymbol{b}^{(i)}_\star$ that maximize $\mathrm{SINR}^{(i)} := \mathcal{E}^{(i)}_s/\mathcal{E}^{(i)}_{ni}$ are derived in the Appendix, where it is found that

$$\begin{aligned}\boldsymbol{b}^{(i)}_\star &:= \arg\max_{\boldsymbol{b}} \mathrm{SINR}^{(i)}(\boldsymbol{b}) \\ &= \boldsymbol{v}_\star\left(\boldsymbol{A}\odot\hat{\boldsymbol{R}}^{(i)*}, \mathrm{diag}\left(\hat{\boldsymbol{R}}^{(i)}+\sigma^2\boldsymbol{I}\right) - \boldsymbol{A}\odot\hat{\boldsymbol{R}}^{(i)*}\right)\end{aligned} \tag{19}$$

where $\boldsymbol{v}_\star(\boldsymbol{B},\boldsymbol{C})$ denotes the principle generalized eigenvalue [23] of the matrix pair $(\boldsymbol{B},\boldsymbol{C})$, and

$$[\boldsymbol{A}]_{m,n} := \frac{\sin\left(\frac{\pi}{N}(2D+1)(n-m)\right)}{N\sin\left(\frac{\pi}{N}(n-m)\right)} \tag{20}$$

$$\hat{\boldsymbol{R}}^{(i)} := \boldsymbol{H}_{\mathrm{tl}}^{(i)}\boldsymbol{H}_{\mathrm{tl}}^{(i)H} \tag{21}$$

$$\left[\boldsymbol{H}_{\mathrm{tl}}^{(i)}\right]_{n,l} := h_{\mathrm{tl}}^{(i)}(n,l). \tag{22}$$

Since $\boldsymbol{b}_\star^{(i)}$ is a function of the channel realization $\boldsymbol{H}_{\mathrm{tl}}^{(i)}$, its calculation is impractical. We provide an alternative below.

3) Max Average-SINR Window: Denoting the channel-averaged SINR by $\overline{\mathrm{SINR}} := \mathrm{E}\{\mathcal{E}_s^{(i)}\}/\mathrm{E}\{\mathcal{E}_{ni}^{(i)}\}$, it follows from (35) that the window coefficients maximizing $\overline{\mathrm{SINR}}$ are

$$\bar{\boldsymbol{b}}_\star = \boldsymbol{v}_\star\left(\boldsymbol{A}\odot\boldsymbol{R},\left(\sigma^2+\sum_l \sigma_l^2\right)\boldsymbol{I}-\boldsymbol{A}\odot\boldsymbol{R}\right) \tag{23}$$

where $\bar{\boldsymbol{R}} := \mathrm{E}\{\hat{\boldsymbol{R}}^{(i)}\}$ is real symmetric with elements $[\bar{\boldsymbol{R}}]_{m,n} = r_{\mathrm{t}}(m-n)\sum_l \sigma_l^2$. Numerical results in Section V demonstrate that $\bar{\boldsymbol{b}}_\star$ and $\boldsymbol{b}_\star^{(i)}$ yield similar performance, even though $\bar{\boldsymbol{b}}_\star$ requires knowledge of only $r_{\mathrm{t}}(\cdot)$ and $\sigma^{-2}\sum_l \sigma_l^2$.

C. Iterative MMSE Symbol Estimation

Here, we present several high-performance low-complexity estimators of $\boldsymbol{s}^{(i)}$ from $\breve{\boldsymbol{x}}^{(i)}$ that leverage the ICI-shortened structure of $\mathcal{C}(\boldsymbol{\beta})\mathcal{H}_{\mathrm{df}}^{(i)}$. In one scheme, we estimate $s_k^{(i)}$ *sequentially* for $k = 0,1,2,\ldots,N-1$, incorporating the outcomes of previous estimates (and/or known pilots) as prior information for subsequent estimates. In doing so, we hope to avoid both the noise-enhancement of linear equalizers and the error-propagation of decision feedback equalizers [27]. After estimating $s_{N-1}^{(i)}$, we re-estimate $s_0^{(i)}$, and so on. In another scheme, we estimate $\{s_0^{(i)},\ldots,s_{N-1}^{(i)}\}$ in *parallel* and then use these estimates as prior information for the re-estimation of $\{s_0^{(i)},\ldots,s_{N-1}^{(i)}\}$. Decision-feedback versions of these schemes are also derived.

1) MMSE Estimation: A common component of our iterative estimation schemes is linear MMSE estimation incorporating priors. In the sequel, we use $\breve{\mathcal{H}}^{(i)} := \mathcal{C}(\boldsymbol{\beta})\mathcal{H}_{\mathrm{df}}^{(i)}$ and omit symbol-index superscripts $^{(i)}$ for brevity, turning (15) into

$$\breve{\boldsymbol{x}} = \breve{\mathcal{H}}\boldsymbol{s} + \mathcal{C}(\boldsymbol{\beta})\boldsymbol{w}. \tag{24}$$

The structure of $\breve{\mathcal{H}}$ (recall Fig. 4) implies that s_k contributes primarily to the observation elements

$$\breve{\boldsymbol{x}}_k := [\breve{x}_{k-D}\cdots\breve{x}_{k+D}]^t$$

where all indexing in this section is taken modulo-N. Leveraging the perfect ICI-shortening assumption[2] in (25)

$$\begin{aligned}
\breve{\mathcal{H}}_k &:= \begin{bmatrix} \breve{h}_{k-D,k-2D} & \cdots & \breve{h}_{k-D,k} & & \\ & \ddots & \vdots & \ddots & \\ & & \breve{h}_{k+D,k} & \cdots & \breve{h}_{k+D,k+2D} \end{bmatrix} \\
\boldsymbol{C}_k &:= \begin{bmatrix} \beta_{k-D} & \beta_{k-D-1} & \cdots & \beta_{k-D-N+1} \\ \beta_{k-D+1} & \beta_{k-D} & \cdots & \beta_{k-D-N+2} \\ \ddots & \ddots & \ddots & \ddots \\ \beta_{k+D} & \beta_{k+D-1} & \cdots & \beta_{k+D-N+1} \end{bmatrix} \\
\breve{\boldsymbol{h}}_k &:= [\breve{h}_{k-D,k}\ \cdots\ \breve{h}_{k+D,k}]^t \\
\boldsymbol{s}_k &:= [s_{k-2D}\ \cdots\ s_{k+2D}]^t
\end{aligned} \tag{25}$$

for $\breve{h}_{n,m} := [\breve{\mathcal{H}}]_{n,m}$, so that we can write

$$\breve{\boldsymbol{x}}_k = \breve{\mathcal{H}}_k\boldsymbol{s}_k + \boldsymbol{C}_k\boldsymbol{w}. \tag{26}$$

Note that as a consequence of modulo-N indexing, the elements of $\breve{\mathcal{H}}$ from the top-right and bottom-left shaded triangles in Fig. 4 are included in $\breve{\mathcal{H}}_k$; the perfect ICI-shortening assumption neglects only the nonshaded regions in Fig. 4.

The MMSE linear estimate of s_k given $\breve{\boldsymbol{x}}_k$ is [28]

$$\hat{s}_k = \mathrm{E}\{s_k\} + \mathrm{Cov}(s_k,\breve{\boldsymbol{x}}_k)\mathrm{Cov}(\breve{\boldsymbol{x}}_k,\breve{\boldsymbol{x}}_k)^{-1}\left(\breve{\boldsymbol{x}}_k - \mathrm{E}\{\breve{\boldsymbol{x}}_k\}\right).$$

If we assume $\mathrm{E}\{\boldsymbol{w}\} = \boldsymbol{0}$, $\mathrm{E}\{\boldsymbol{w}\boldsymbol{w}^H\} = \sigma^2\boldsymbol{I}$, $\mathrm{E}\{\boldsymbol{s}\boldsymbol{w}^H\} = \boldsymbol{0}$, and independence among $\{s_k\}$, and if we define $\bar{s}_k := \mathrm{E}\{s_k\}$, $v_k := \mathrm{Cov}(s_k,s_k)$, $\bar{\boldsymbol{s}}_k := [\bar{s}_{k-2D},\ldots,\bar{s}_{k+2D}]^t$, and $\boldsymbol{v}_k := [v_{k-2D},\ldots,v_{k+2D}]^t$, then it is straightforward to show that

$$\begin{aligned}
\mathrm{E}\{\breve{\boldsymbol{x}}_k\} &= \breve{\mathcal{H}}_k\bar{\boldsymbol{s}}_k \\
\mathrm{Cov}(s_k,\breve{\boldsymbol{x}}_k) &= v_k\breve{\boldsymbol{h}}_k^H \\
\mathrm{Cov}(\breve{\boldsymbol{x}}_k,\breve{\boldsymbol{x}}_k) &= \sigma^2\boldsymbol{C}_k\boldsymbol{C}_k^H + \breve{\mathcal{H}}_k\mathcal{D}(\boldsymbol{v}_k)\breve{\mathcal{H}}_k^H
\end{aligned}$$

giving the MMSE linear estimate

$$\boldsymbol{f}_k = \left(\sigma^2\boldsymbol{C}_k\boldsymbol{C}_k^H + \breve{\mathcal{H}}_k\mathcal{D}(\boldsymbol{v}_k)\breve{\mathcal{H}}_k^H\right)^{-1}\breve{\boldsymbol{h}}_k v_k \tag{27}$$

$$\hat{s}_k = \bar{s}_k + \boldsymbol{f}_k^H(\breve{\boldsymbol{x}}_k - \breve{\mathcal{H}}_k\bar{\boldsymbol{s}}_k). \tag{28}$$

We choose to use only *extrinsic* information, i.e., only the priors from $\{s_d, d\neq k\}$ when estimating s_k. This can be accomplished using (27) and (28) with $\bar{s}_k = 0$, $v_k = 1$, and

$$\bar{\boldsymbol{s}}_k := [\bar{s}_{k-2D},\ldots,\bar{s}_{k-1},0,\bar{s}_{k+1},\ldots,\bar{s}_{k+2D}]^t \tag{29}$$

$$\boldsymbol{v}_k := [v_{k-2D},\ldots,v_{k-1},1,v_{k+1},\ldots,v_{k+2D}]^t. \tag{30}$$

2) Updating the Priors: The symbol estimate $\hat{s}_k$ can be used to update $\bar{s}_k$ and v_k. For simplicity, we consider only independent and identically distributed (i.i.d.) binary phase shift keying

[2]Section V examines the implications of this assumption.

(BPSK) symbols $s_k \in \mathcal{B} := \{-1, +1\}$; quadrature amplitude modulation (QAM) extensions are straightforward. Assuming a conditionally Gaussian model for the estimates:

$$p(\hat{s}_k|s_k = b) \approx \frac{1}{\sigma_k(b)} \cdot \phi\left(\frac{\hat{s}_k - \mu_k(b)}{\sigma_k(b)}\right)$$

where $\phi(w) := e^{-w^2}/\sqrt{\pi}$ is the proper complex Gaussian density, and defining $\mu_k(b) := \mathrm{E}\{\hat{s}_k|s_k = b\}$ and $\sigma_k^2(b) := \mathrm{Cov}(\hat{s}_k, \hat{s}_k|s_k = b)$, it can be shown that

$$\begin{aligned}\mu_k(b) &= \boldsymbol{f}_k^H \breve{\boldsymbol{h}}_k b \\ \sigma_k^2(b) &= \boldsymbol{f}_k^H \breve{\boldsymbol{h}}_k \left(1 - \breve{\boldsymbol{h}}_k^H \boldsymbol{f}_k\right).\end{aligned}$$

If we define the prior and posterior log-likelihood ratios (LLRs) as $L(s_k) := \ln(P(s_k = +1)/P(s_k = -1))$ and $L(s_k|\hat{s}_k) := \ln(P(s_k = +1|\hat{s}_k)/P(s_k = -1|\hat{s}_k))$, respectively, their difference can be expressed as

$$\begin{aligned}\Delta L(\hat{s}_k) &:= L(s_k|\hat{s}_k) - L(s_k) \\ &= \ln \frac{p(\hat{s}_k|s_k = +1)}{p(\hat{s}_k|s_k = -1)} \\ &= \ln\left(\frac{\sigma_k(-1)}{\sigma_k(+1)} \cdot \frac{\phi\left(\frac{\hat{s}_k - \mu_k(+1)}{\sigma_k(+1)}\right)}{\phi\left(\frac{\hat{s}_k - \mu_k(-1)}{\sigma_k(-1)}\right)}\right) \\ &= \frac{4\mathrm{Re}(\hat{s}_k)}{1 - \breve{\boldsymbol{h}}_k^H \boldsymbol{f}_k}. \qquad (31)\end{aligned}$$

The posterior LLR leads to an update of the priors:

$$\begin{aligned}\bar{s}_{k,\mathrm{new}} &= \sum_{b \in \mathcal{B}} b \cdot P(s_k = s|\hat{s}_k) \\ &= P(s_k = +1|\hat{s}_k) - P(s_k = -1|\hat{s}_k) \\ &= \frac{e^{L(s_k|\hat{s}_k)}}{e^{L(s_k|\hat{s}_k)} + 1} - \frac{1}{e^{L(s_k|\hat{s}_k)} + 1} \\ &= \tanh\left(\frac{L(s_k|\hat{s}_k)}{2}\right) \qquad (32) \\ v_{k,\mathrm{new}} &= \sum_{b \in \mathcal{B}} (b - \mathrm{E}\{s_k|\hat{s}_k\})^2 \cdot P(s_k = b|\hat{s}_k) \\ &= 1 - \bar{s}_{k,\mathrm{new}}^2 \qquad (33) \\ L_{\mathrm{new}}(s_k) &= L(s_k) + \Delta L(\hat{s}_k) \qquad (34)\end{aligned}$$

which, in turn, can be used to estimate $\{s_{d \neq k}\}$ via (27)–(30).

3) Iterative Joint Estimators: To initialize the iterative algorithms, we set $\bar{s}_k = 0$ and $v_k = 1$ for indices k that do not correspond to pilots. For k corresponding to pilots, $\bar{s}_k$ are assigned the pilot amplitudes and $v_k = 0$. Various methods of iterative OFDM-symbol estimation are proposed below.

In *block iterative estimation* (BIE), we calculate the entire batch of estimates $\{\hat{s}_k, k = 0, \ldots, N-1\}$ via (27)–(30) before updating the priors via (31)–(33). Using updated priors, a new batch of estimates is computed via (27)–(30), and so on. The algorithm terminates when the LLRs surpass a threshold, or a specified number of iterations have elapsed.

In *sequential iterative estimation* (SIE), we calculate $\hat{s}_0$ via (27)–(30) and then immediately update the priors $\bar{s}_0$ and v_0 via (31)–(33). Next, we calculate $\hat{s}_1$ and then immediately update $\bar{s}_1$ and v_1. This continues until $\hat{s}_{N-1}$, $\bar{s}_{N-1,\mathrm{new}}$, and $v_{N-1,\mathrm{new}}$ have been computed, then repeats again, starting with $\hat{s}_0$. The algorithm terminates when the LLRs surpass a threshold or a specified number of iterations have elapsed.

Block decision feedback (BDF) operates identically to BIE, except that $\bar{s}_{k,\mathrm{new}} = \mathrm{sgn}(\hat{s}_k)$ and $v_{k,\mathrm{new}} = 0$. Computation of LLRs is not necessary, and the algorithm terminates when $\{\bar{s}_k\}$ converge or a specified number of iterations have elapsed.

Sequential decision feedback (SDF) [29] operates identically to SIE, except that $\bar{s}_{k,\mathrm{new}} = \mathrm{sgn}(\hat{s}_k)$ and $v_{k,\mathrm{new}} = 0$. Here too, computation of LLRs is not necessary, and the algorithm terminates when $\{\bar{s}_k\}$ converge or a specified number of iterations have elapsed.

4) Incorporating the Decoder: After the symbol estimation algorithm terminates, we have the choice of passing LLRs $\{L(s_k|\hat{s}_k)\}$, soft estimates $\{\hat{s}_k\}$, or hard estimates $\{\mathrm{sgn}(\hat{s}_k)\}$ to the decoder. After decoding, updated LLRs could be passed back to the estimator, forming an outer loop of iteration. This latter scheme is a form of *turbo equalization* [30], which we discuss further in Section V.

5) Computational Complexity: The implementation complexity of the BIE, SIE, BDF, and SDF algorithms is dominated by the $(2D + 1) \times (2D + 1)$ Hermitian matrix inversion in (27). As this requires only $\mathcal{O}(D^2)$ operations, a total of $\mathcal{O}(D^2 N)$ operations is needed per iteration. It should be noted that $\boldsymbol{C}_k \boldsymbol{C}_k^H$ is fixed for all k and equal to a subblock of $\mathcal{C}(\boldsymbol{F}(\boldsymbol{b} \odot \boldsymbol{b}^*)/\sqrt{N})$, the latter of which can be precomputed for realization-independent $\boldsymbol{b}$.

6) Relation to Other Known Schemes: The iterative algorithms proposed in Section IV-C are related to, yet distinct from, a number of existing algorithms. SIE is perhaps closest to the estimation stage in the "turbo equalization" scheme of Tüchler *et al.* [31]. Unlike SIE, however, [31] assumes an LTI channel in white noise and inserts a decoding iteration after each equalization iteration. SDF is reminiscent of the "successive detection" scheme proposed for V-BLAST-coded signals in [32] and OFDM reception in [15], although SDF does not require their (computationally-expensive) symbol ordering procedure. In addition, SDF allows for multiple iterations so that hard decisions are given an opportunity to converge. BIE bears some similarity to the "partial interference cancellation" scheme proposed by Divsalar *et al.* [33] for CDMA reception, although [33] is based on matched-filtering rather than MMSE estimation (e.g., there is no autocorrelation-matrix inverse in [33]). Finally, BDF is reminiscent of hard parallel interference cancellation (HPIC) schemes, like "multistage detection" [34], for CDMA reception. HPIC schemes, however, also use matched filtering in place of MMSE estimation. MMSE estimation is considered too computationally expensive for practical CDMA applications because their system matrices do not have the sparse banded structure in Fig. 4.

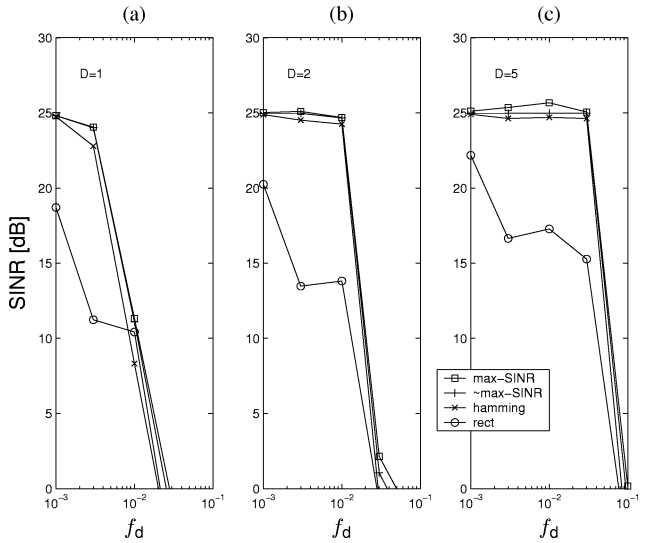

Fig. 5. SINR versus $f_{\rm d}$ for $N = 128$, SNR $= 25$ dB, WSSUS Rayleigh fading, and various D.

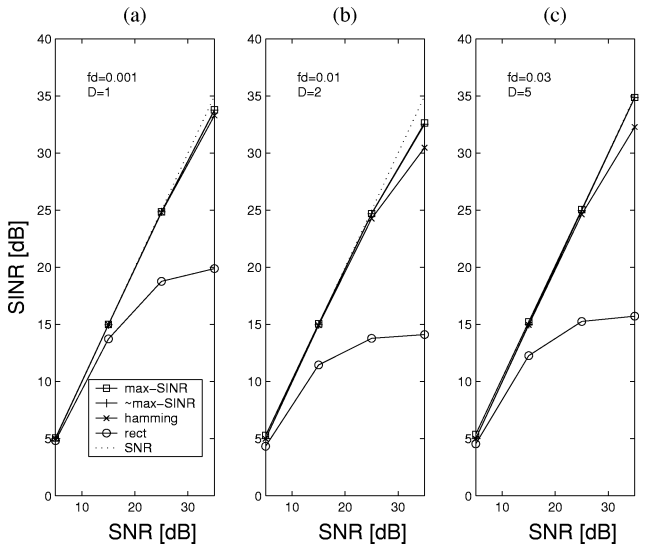

Fig. 6. SINR versus SNR for $N = 128$, WSSUS Rayleigh fading, and various $\{f_{\rm d}, D\}$.

V. Numerical Results and Discussion

Here, we discuss the proposed algorithms and study the results of numerical simulations. All experiments employed $N = 128$ i.i.d., unit-variance BPSK symbols per OFDM symbol, SNR^{-1}-variance circular AWGN noise, and an energy-preserving WSSUS Rayleigh-fading channel with $\sigma_l^2 = N_h^{-1}$ (for $0 \le l < N_h$) and $N_h = N/4$. Perfect channel knowledge was assumed, and no pilots were employed.

As a benchmark, consider symbol estimation given perfect knowledge of interfering symbols. This generates the so-called *matched filter bound* (MFB). The MFB does not use assume perfect ICI-shortening; it makes use of the unwindowed observation $\boldsymbol{x}^{(i)}$ in (7). Consider also an *approximate MFB* (AMFB) in which a masked version of $\breve{\mathcal{H}}^{(i)} = \mathcal{C}(\boldsymbol{\beta})\mathcal{H}_{\rm df}^{(i)}$ [akin to (25)] is used for estimation. This bound can be calculated by the estimation (27), (28) with $\bar{s}_k = 0$, $v_k = 1$, $\bar{\boldsymbol{s}}_k := [s_{k-2D}, \ldots, s_{k-1}, 0, s_{k+1}, \ldots, s_{k+2D}]^t$, and $\boldsymbol{v}_k := \boldsymbol{e}_{2D}$,

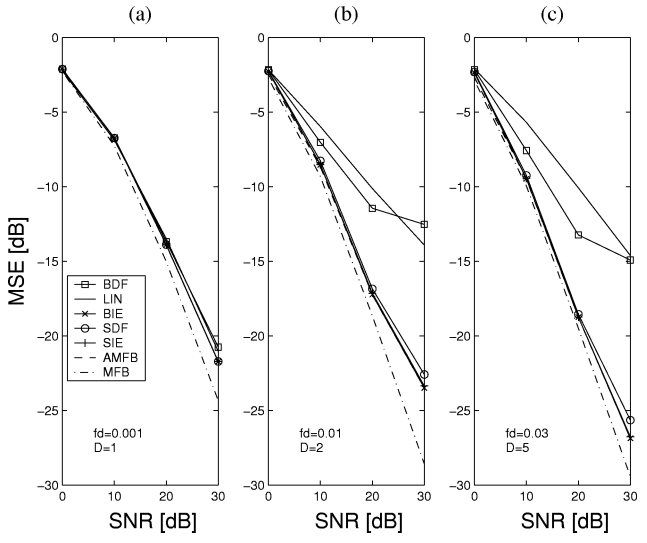

Fig. 7. MSE versus SNR after ten iterations for $N = 128$, WSSUS Rayleigh fading, and windowing with $\bar{\boldsymbol{b}}_\star$.

where $\boldsymbol{e}_d$ denotes a vector of zeros with a one in the dth position. The AMFB lower bounds the MSE performance of BIE, SIE, BDF, and SDF since they were designed around the perfectly shortened ICI assumption.

Figs. 5 and 6 investigate the subcarrier-averaged SINR performance of various windows averaged over 1000 channel realizations. Recall that SINR$^{(i)} = \mathcal{E}_s^{(i)}/\mathcal{E}_{ni}^{(i)}$ for $\mathcal{E}_s^{(i)}$, and $\mathcal{E}_{ni}^{(i)}$ defined in (17) and (18), respectively. Fig. 5 supports the rule $D \ge \lceil f_{\rm d} N \rceil + 1$ and verifies that although complexity increase discourages larger D, performance does not. Fig. 6 shows that with max-SINR windowing and proper selection of D, SINR $\approx$ SNR over the expected operating region. This suggests that the interference is dominated by channel noise and not residual ICI, i.e., that max-SINR windowing does indeed suppress undesired ICI.

Both Figs. 5 and 6 show that there is little difference between the performance of the max-SINR window $\boldsymbol{b}_\star^{(i)}$ and the max-average-SINR window $\bar{\boldsymbol{b}}_\star$. In fact, for typical SNRs and a conservative choice of D, there is little difference between $\boldsymbol{b}_\star^{(i)}$ and the Hamming window [25]. For high SNR or $D < \lceil f_{\rm d} N \rceil + 1$, however, the Hamming window is suboptimal. The rectangular window (i.e., the absence of windowing) is clearly suboptimal in all but the lowest SNR environments.

Figs. 7 and 8 compare the subcarrier-averaged MSE performance of the SDF, BDF, BIE, and SIE iterative symbol estimation algorithms proposed in Section IV-C3 to the MFB, AMFB, and linear MMSE estimator (12). Each trace represents the average of 5000 channel/data realizations. The iterative algorithms and AMFB employed the max-average-SINR window $\bar{\boldsymbol{b}}_\star$. Fig. 7 shows performance after ten iterations, whereas Fig. 8 shows performance after two. Note that the $\{f_{\rm d}, D\}$ pairs match those in Fig. 6.

Fig. 7 shows that when $f_{\rm d}$ is small, all iterative algorithms perform very close to the AMFB after convergence. For small $f_{\rm d}$, the linear MMSE estimator is also close to the AMFB. The difference between the AMFB and the true MFB can be interpreted as the cost of $\mathcal{O}(N)$ rather than $\mathcal{O}(N^2)$ estimation complexity. The MFB improvement with increasing $f_{\rm d}$ can be

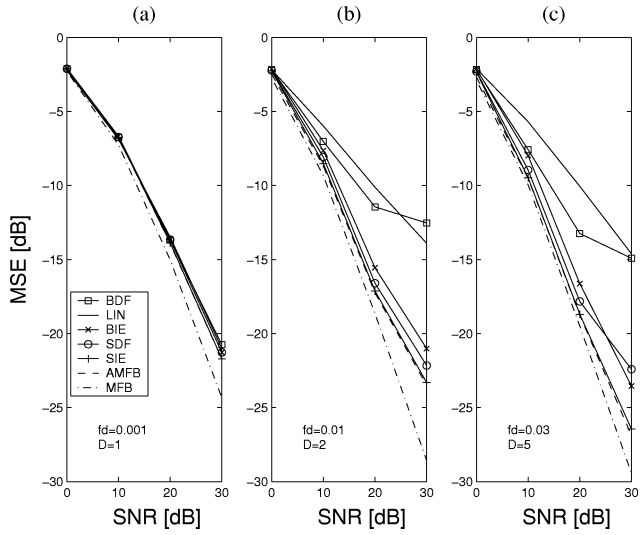

Fig. 8. MSE versus SNR after two iterations for $N = 128$, WSSUS Rayleigh fading, and windowing with $\hat{\boldsymbol{b}}_\star$.

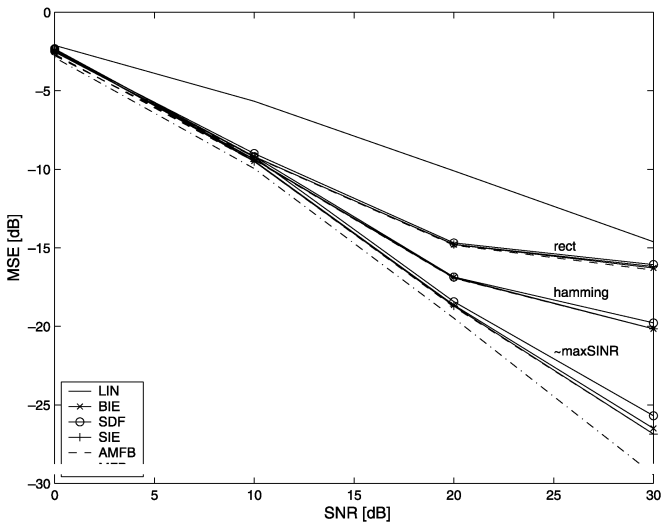

Fig. 9. MSE versus SNR after ten iterations for $N = 128$, WSSUS Rayleigh fading with $f_{\rm d} = 0.03$, and various windows.

attributed to the increase in diversity that comes with higher Doppler spread [35], [36]. This implies that when used properly, ICI can actually *enhance* symbol estimation performance.

For larger $f_{\rm d}$, we see performance differences between the algorithms. For example, iterative estimation significantly outperforms linear MMSE equalization. This is especially noteworthy given that the complexity of iterative estimation is far less than that of the linear MMSE estimation.[3] Fig. 8 shows that the SIE algorithm reaches, for all practical purposes, the AMFB in only two iterations. While at this point BIE remains 2–3 dB away, it also reaches the AMFB after three iterations (as observed in simulations not shown here). We surmise that SIE converges faster than BIE because it makes immediate use of the prior information on interfering symbols. The fact that equalization alone reaches the AMFB implies that incorporating symbol reliability information from a decoder (i.e., turbo equalization) would not improve the estimates. A more sophisticated linear preprocessing stage could, however, help close the gap between the AMFB and MFB. While the simple SDF algorithm performs nearly as well as SIE in the cases that we have examined, our experience with other (i.e., non-OFDM) channels suggests that this is generally not the case. Due to error propagation, BDF performs the worst of the four iterative algorithms.

Fig. 9 shows the MSE attained by the SIE, BIE, and SDF algorithms in comparison with the AMFB for various windows. For reference, it also shows the MSE attained by the MFB and linear MMSE estimator, neither of which employs windowing. The approximately max-SINR window $\hat{\boldsymbol{b}}_\star$ exhibits performance close to the MFB, as expected from Fig. 7(c). The hamming window suffers in performance, especially at high SNR, and the rectangular window performs by far the worst. In all cases, however, it is interesting to note that SIE reaches the AMFB: the best performance that can be expected given the window choice. Fig. 9 provides clear evidence that the "basis expansion model" (BEM) constructed from a *rectangularly*-windowed sinusoidal basis [16], [20] and truncated to include no more than $2D + 1$ elements, where $D = \lceil f_{\rm d} N \rceil + 1$, is ineffective in its description of doubly selective WSSUS Rayleigh channels; receivers designed around this BEM will perform poorly on true (i.e., nonapproximate) channels.

[3]Since the MSE of the $\mathcal{O}(N)$ (linear) equalizer in [14] is lower bounded by that of the linear MMSE estimator (12), it is clear that the $\mathcal{O}(N)$ equalizers SIE, BIE, and SDF significantly outperform the one in [14].

VI. Conclusions

Equalization of OFDM in doubly selective channels is complicated by the existence of ICI: The classical frequency-domain equalizer—a simple scaling of each sub-carrier—is no longer sufficient. Previously proposed doubly selective OFDM equalizers either approximate the linear MMSE estimator with an $\mathcal{O}(N)$ scheme, resulting in relatively poor performance, or require at least $\mathcal{O}(N^2)$ operations per OFDM symbol, making them infeasible for large symbol length N. In response, we proposed a low-complexity two-stage equalizer whose performance far surpasses the linear MMSE estimator. The first stage, requiring $\mathcal{O}(N)$ operations, applies SINR-optimal windowing to squeeze ICI into a range of $2D + 1$ subcarrier intervals. The second stage, requiring $\mathcal{O}(D^2 N)$ operations, uses iterative soft ICI-cancellation to estimate the frequency-domain symbols. Simulations indicate that our equalizer performs close to the MFB after only two iterations.

Appendix

In this Appendix, we use the property $\|\boldsymbol{F}\boldsymbol{B}\|_F = \|\boldsymbol{B}\|_F$ and the definitions $\mathcal{E}_t^{(i)} := \mathrm{E}\{\|\boldsymbol{x}^{(i)}\|^2\}$, $[\boldsymbol{H}_{\rm df}^{(i)}]_{d,k} := h_{\rm df}^{(i)}(d,k)$, and

$$\boldsymbol{P}_D := \begin{bmatrix} \boldsymbol{I}_{D+1} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\mathrm{I}}_D \end{bmatrix}.$$

Note that $\boldsymbol{H}_{\rm df}^{(i)} = \boldsymbol{F}\boldsymbol{H}_{\rm tl}^{(i)}\boldsymbol{F}$ for $\boldsymbol{H}_{\rm tl}^{(i)}$ defined in (22). For total energy $\mathcal{E}_t^{(i)} := \mathcal{E}_s^{(i)} + \mathcal{E}_{ni}^{(i)}$, we have

$$\mathrm{SINR}^{(i)} = \frac{\mathcal{E}_s^{(i)}}{\mathcal{E}_t^{(i)} - \mathcal{E}_s^{(i)}}.$$

Since $\mathcal{E}_t^{(i)}$ and $\mathcal{E}_s^{(i)}$ can be expressed as quadratic forms

$$\begin{aligned}
\mathcal{E}_{\mathrm{t}}^{(i)} &= \sum_{n=0}^{N-1} |b_n|^2 \mathrm{E}\left\{\left|r_n^{(i)}\right|^2\right\} \\
&= \boldsymbol{b}^H \mathrm{diag}\left(\mathrm{E}\left\{\boldsymbol{r}^{(i)}\boldsymbol{r}^{(i)H}\right\}\right)\boldsymbol{b} \\
&= \boldsymbol{b}^H \mathrm{diag}\left(\mathcal{H}_{\mathrm{tl}}^{(i)}\mathcal{H}_{\mathrm{tl}}^{(i)H} + \sigma^2\boldsymbol{I}\right)\boldsymbol{b} \\
&= \boldsymbol{b}^H \mathrm{diag}\left(\boldsymbol{H}_{\mathrm{tl}}^{(i)}\boldsymbol{H}_{\mathrm{tl}}^{(i)H} + \sigma^2\boldsymbol{I}\right)\boldsymbol{b} \\
\mathcal{E}_s^{(i)} &= \left\|\mathcal{M}_{\mathrm{D}}\left(\mathcal{C}(\boldsymbol{\beta})\mathcal{H}_{\mathrm{df}}^{(i)}\right)\right\|_F^2 \\
&= \left\|\boldsymbol{P}_D\mathcal{C}(\boldsymbol{\beta})\boldsymbol{H}_{\mathrm{df}}^{(i)}\right\|_F^2 \\
&= \left\|\boldsymbol{P}_D\boldsymbol{F}\mathcal{D}(\boldsymbol{b})\boldsymbol{F}^H\boldsymbol{F}\boldsymbol{H}_{\mathrm{tl}}^{(i)}\boldsymbol{F}\right\|_F^2 \\
&= \left\|\boldsymbol{P}_D\boldsymbol{F}\mathcal{D}(\boldsymbol{b})\boldsymbol{H}_{\mathrm{tl}}^{(i)}\right\|_F^2 \\
&= \sum_{l=0}^{N_h-1}\sum_{|k|\le D}\left|\frac{1}{\sqrt{N}}\sum_{n=0}^{N-1} b_n h_{\mathrm{tl}}^{(i)}(n,l)e^{-j\frac{2\pi}{N}kn}\right|^2 \\
&= \frac{1}{N}\sum_{n,m,l} b_n b_m^* \sum_{|k|\le D} e^{-j\frac{2\pi}{N}k(n-m)} h_{\mathrm{tl}}^{(i)}(n,l)h_{\mathrm{tl}}^{(i)*}(m,l) \\
&= \boldsymbol{b}^H\left(\boldsymbol{A}\odot\hat{\boldsymbol{R}}^{(i)*}\right)\boldsymbol{b}
\end{aligned}$$

where the matrices $\boldsymbol{A}$ and $\hat{\boldsymbol{R}}^{(i)}$ are defined in (20) and (21), respectively, SINR becomes

$$\mathbf{SINR}^{(i)}(\boldsymbol{b}) = \frac{\boldsymbol{b}^H\left(\boldsymbol{A}\odot\hat{\boldsymbol{R}}^{(i)*}\right)\boldsymbol{b}}{\boldsymbol{b}^H\left(\mathrm{diag}\left(\hat{\boldsymbol{R}}^{(i)}+\sigma^2\boldsymbol{I}\right)-\boldsymbol{A}\odot\hat{\boldsymbol{R}}^{(i)*}\right)\boldsymbol{b}} \tag{35}$$

and the maximizing coefficients $\boldsymbol{b}_\star^{(i)}$ are given in (19) as the solution to a generalized eigenvector problem [23].

ACKNOWLEDGMENT

The author thanks S. H. D'Silva for his involvement in preliminary stages of this research and H. El Gamal for useful discussions on iterative estimation.

REFERENCES

[1] S. B. Weinstein and P. M. Ebert, "Data transmission by frequency division multiplexing using the discrete Fourier transform," *IEEE Trans. Commun.*, vol. COM-19, pp. 628–634, Oct. 1971.

[2] L. J. Cimini Jr., "Analysis and simulation of a digital mobile radio channel using orthogonal frequency division multiplexing," *IEEE Trans. Commun.*, vol. COM-33, pp. 665–765, July 1985.

[3] M. Russell and G. L. Stüber, "Interchannel interference analysis of OFDM in a mobile environment," in *Proc. IEEE Veh. Tech. Conf.*, vol. 2, 1995, pp. 820–824.

[4] S. N. Diggavi, "Analysis of multicarrier transmission in time-varying channels," in *Proc. IEEE Int. Conf. Commun.*, vol. 3, 1997, pp. 1191–1195.

[5] P. Robertson and S. Kaiser, "The effects of Doppler spreads on OFDM(A) mobile radio systems," in *Proc. IEEE Veh. Tech. Conf.*, vol. 1, 1999, pp. 329–333.

[6] Y. H. Kim, I. Song, H. G. Kim, T. Chang, and H. M. Kim, "Performance analysis of a coded OFDM system in time-variant multipath Rayleigh fading channels," *IEEE Trans. Veh. Tech.*, vol. 48, pp. 1610–1615, Sept. 1999.

[7] B. Stantchev and G. Fettweis, "Time-variant distortions in OFDM," *IEEE Commun. Lett.*, vol. 4, pp. 312–314, Oct. 2000.

[8] Y. Li and L. J. Cimini Jr., "Bounds on the interchannel interference of OFDM in time-varying impairments," *IEEE Trans. Commun.*, vol. 49, pp. 401–404, Mar. 2001.

[9] W. C. Jakes, *Microwave Mobile Communications*. New York: Wiley, 1974.

[10] *Air Interface for Fixed Broadband Wireless Access Systems. Part A: Systems Between 2–11 GHz* (in IEEE Stand. 802.16, 01/01r1), July 2001.

[11] *Digital Video Broadcasting (DVB): Framing Structure, Channel Coding and Modulation for Digital Terrestrial Television* (in ETSI Stand. ETS 300 744), July 1999.

[12] J. Armstrong, P. M. Grant, and G. Povey, "Polynomial cancellation coding of OFDM to reduce intercarrier interference due to Doppler spread," in *Proc. IEEE Global Telecommun. Conf.*, vol. 5, 1998, pp. 2771–2776.

[13] Y. Zhao and S.-G. Häggman, "Intercarrier interference self-cancellation scheme for OFDM mobile communication sytems," *IEEE Trans. Commun.*, vol. 49, pp. 1185–1191, July 2001.

[14] 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 Trans. Commun.*, vol. 47, pp. 27–32, Jan. 1999.

[15] 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 Trans. Commun.*, vol. 49, pp. 1375–1387, Aug. 2001.

[16] X. Cai and G. B. Giannakis, "Low-complexity ICI suppression for OFDM over time-and frequency-selective Rayleigh fading channels," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Nov. 2002.

[17] J.-P. M. G. Linnartz and A. Gorokhov, "New equalization approach for OFDM over dispersive and rapidly time varying channel," in *Proc. IEEE Int. Symp. Personal Indoor Mobile Radio Commun.*, vol. 2, 2000, pp. 1375–1379.

[18] A. Stamoulis, S. N. Diggavi, and N. Al-Dhahir, "Intercarrier interference in MIMO OFDM," *IEEE Trans. Signal Processing*, vol. 50, pp. 2451–2464, Oct. 2002.

[19] P. Schniter, "Low-complexity estimation of doubly selective channels," in *Proc. IEEE Workshop Signal Processing Advances in Wireless Commun.*, 2003, to be published.

[20] X. Ma and G. B. Giannakis, "Designing maximum multipath-Doppler diversity transmissions over time- and frequency-selective wireless channels," in *Proc. Allerton Conf. Commun., Contr., Comput.*, 2001.

[21] M. Vetterli and J. Kovacević, *Wavlets and Subband Coding*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[22] H. Sari, G. Karam, and I. Jeanclaude, "Transmission techniques for digital terrestrial TV broadcasting," *IEEE Commun. Mag.*, pp. 100–109, Feb. 1995.

[23] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: John Hopkins Univ. Press, 1983.

[24] D. D. Falconer and F. R. Magee, "Adaptive channel memory truncation for maximum likelihood sequence estimation," *Bell Syst. Tech. J.*, vol. 52, pp. 1541–1562, Nov. 1973.

[25] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[26] P. Davis, *Circulant Matrices*. New York: Wiley, 1979.

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

[28] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York: Springer, 1994.

[29] P. Schniter and S. H. D'Silva, "Low-complexity detection of OFDM in doubly-dispersive channels," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Nov. 2002.

[30] C. Douillard, M. Jezequel, C. Berrou, A. Picart, P. Didier, and A. Glavieux, "Iterative correction of intersymbol interference: turbo equalization," *Eur. Trans. Telecommun.*, vol. 6, pp. 507–511, Sept.–Oct. 1995.

[31] M. Tüchler, A. Singer, and R. Koetter, "Minimum mean square error equalization using a priori information," *IEEE Trans. Signal Processing*, vol. 50, pp. 673–683, Mar. 2002.

[32] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. ISSSE*, Sept. 1998.

[33] D. Divsalar, M. K. Simon, and D. Raphaeli, "Improved parallel interference cancellation for CDMA," *IEEE Trans. Commun.*, vol. 46, pp. 258–268, Feb. 1998.

[34] M. K. Varanasi and B. Aazhang, "Multistage detection in asynchronous code-division multiple-access communications," *IEEE Trans. Commun.*, vol. 38, pp. 509–519, Apr. 1990.

[35] A. M. Sayeed and B. Aazhang, "Joint multipath-doppler diversity in mobile wireless communications," *IEEE Trans. Commun.*, vol. 47, pp. 123–132, Jan. 1999.

[36] N. J. Baas and D. P. Taylor, "Matched filter bounds for wireless communication over Rayleigh fading dispersive channels," *IEEE Trans. Commun.*, vol. 49, pp. 1525–1528, Sept. 2001.

Philip Schniter (M'92) was born in Evanston, IL, in 1970. He 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. In 2000, he received the Ph.D. degree in electrical engineering from Cornell University, Ithaca, NY.

From 1993 to 1996, he was with Tektronix Inc., Beaverton, OR, as a systems engineer. There, he worked on signal processing aspects of video and communication instrumentation design, including algorithms, software, and hardware architectures. He is currently an Assistant Professor with the Department of Electrical Engineering, The Ohio State University, Columbus. His areas of research include signal processing for communication systems and wireless sensor networks.

Dr. Schniter received the 1998 Schlumberger Fellowship and the 1998–1999 Intel Foundation Fellowship while pursuing the Ph.D. degree. He received 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's CAREER Award.