

# Orthogonal Time–Frequency Signaling Over Doubly Dispersive Channels

Ke Liu, *Student Member, IEEE*, Tamer Kadous, *Member, IEEE*, and Akbar M. Sayeed, *Senior Member, IEEE*

**Abstract**—This paper develops a general framework for communication over doubly dispersive fading channels via an orthogonal short-time Fourier (STF) basis. The STF basis is generated from a prototype pulse via time–frequency shifts. In general, the orthogonality between basis functions is destroyed at the receiver due to channel dispersion. The starting point of this work is a pulse scale adaptation rule first proposed by Kozek to minimize the interference between the basis functions. We show that the average signal-to-interference-and-noise (SINR) ratio associated with different basis functions is identical and is maximized by the scale adaptation rule. The results in this paper highlight the critical impact of the channel spread factor, the product of multipath and Doppler spreads, on system performance. Smaller spread factors result in lesser interference such that a scale-adapted STF basis serves as an approximate eigenbasis for the channel. A highly effective iterative interference cancellation technique is proposed for mitigating the residual interference for larger spread factors. The approximate eigendecomposition leads to an intuitively appealing block-fading interpretation of the channel in terms of time–frequency coherence subspaces: the channel is highly correlated within each coherence subspace whereas it is approximately independent across different subspaces. The block-fading model also yields an approximate expression for the coherent channel capacity in terms of parallel flat-fading channels. The deviation of the capacity of doubly dispersive channels from that of flat-fading channels is quantified by studying the moments of the channel eigenvalue distribution. In particular, the difference between the moments of doubly dispersive and flat-fading channels is proportional to channel spread factor. The results in this paper indicate that the proposed STF signaling framework is applicable for spread factors as large as 0.01.

**Index Terms**—Capacity, doubly dispersive channels, Gabor systems, random banded matrices, time-varying multipath channels.

## I. INTRODUCTION

WIRELESS channels typically exhibit time-varying multipath fading and can be modeled as linear doubly dispersive stochastic channels. The signal experiences dispersion in both time and frequency as it passes through the channel. The multipath effect causes dispersion in time, while the time-varying channel gain associated with each path results in dispersion in frequency. As compared to our understanding of the

classical additive white Gaussian noise (AWGN) channel, modulation and coding for doubly dispersive fading channels tend to be quite challenging and different from that for the AWGN channel.

A general approach to digital communication is *orthogonal signaling* in which transmitted symbols are modulated onto a set of orthonormal basis waveforms. An important class of such schemes is orthogonal frequency-division multiplexing (OFDM) or the discrete multitone modulation (DMT) [1]. In essence, OFDM divides the channel into many small frequency bands and is well suited for a time-dispersive (frequency-selective) channel [2]. For a slowly fading frequency-selective channel, longer symbol durations (narrow frequency bands) are desirable for mitigating the effect of multipath. However, longer symbol durations are more prone to frequency dispersion due to temporal channel variations. Thus, we see that in doubly dispersive channels, the symbol duration is constrained by conflicting requirements dictated by temporal and spectral dispersion. This suggests signaling over short-time Fourier (STF) basis functions whose time–frequency support is matched to channel characteristics.

STF signaling over doubly dispersive channels has been explored by several researchers [3]–[7]. We briefly review this work to put the results of this paper in proper perspective. An STF basis is generated from a given prototype pulse via time and frequency shifts (see (8)). It is also referred to as a Gabor basis or a Weyl–Heisenberg basis in the literature on time–frequency analysis. The time separation ( $T_0$ ) and frequency separation ( $F_0$ ) between STF basis functions critically affect the time–frequency characteristics of an STF basis. *Complete orthogonal* STF bases are only possible for  $T_0F_0 = 1$  (critical sampling) but are known to suffer from poor time–frequency localization (see, e.g., [8], [9]). If we relax the condition of critical sampling and consider  $T_0F_0 > 1$ , STF bases with better time–frequency localization properties can be generated, but they are necessarily *incomplete*. Several researchers have investigated such incomplete systems for doubly dispersive channels; a particular focus has been on *pulse shape* optimization to attain good time–frequency localization [3]–[6]. In this case, two sets of *biorthogonal bases* are used: one at the transmitter for modulating the symbols, and one at the receiver for recovering the symbols. The two bases are intimately related via a duality relationship (see, e.g., [10], [3]).

The time–frequency dispersion induced by a time-varying multipath channel destroys the orthogonality/biorthogonality condition in the above systems. As a result, there is interference between different basis functions at the receiver. A key motivation of the above works on biorthogonal systems is that

Manuscript received June 8, 2001; revised January 15, 2004. This work was supported in part by the National Science Foundation under CAREER Grant CCR-9875805. The material in this paper was presented at the IEEE International Symposium on Information Theory, Lausanne, Switzerland, June/July 2002 and at the IEEE Vehicular Technology Conference (VTC-Fall 2002), Vancouver, BC, Canada, September 2002.

K. Liu and A. M. Sayeed are with the Department of Electrical and Computer Engineering, University of Wisconsin–Madison, Madison, WI 53706 USA (e-mail: kliu@cae.wisc.edu; akbar@engr.wisc.edu).

T. Kadous is with Qualcomm Technologies and Ventures, Qualcomm Inc., San Diego, CA 92121 USA (e-mail: tkadous@qualcomm.com).

Communicated by D. N. C. Tse, Associate Editor for Communications.  
Digital Object Identifier 10.1109/TIT.2004.836931

bases with better time–frequency localization lead to lesser interference between basis functions. While these works study basis design quite thoroughly from a time–frequency localization perspective, the communication and information-theoretic aspects of STF signaling over dispersive channels are not fully explored. The most comprehensive study of incomplete bi-orthogonal systems in a communication context is done in [3]. However, the improvements due to pulse optimization are only quantified in terms of the mean-square error (MSE) in reconstructing the information symbols at the receiver. While biorthogonal systems have less interference compared to orthogonal systems, they suffer a loss in spectral efficiency compared to orthogonal systems (which have maximum spectral efficiency) by a factor of  $1/T_0F_0 < 1$ .<sup>1</sup> As we demonstrate in this paper, the spectral efficiency has a *linear* impact on overall system capacity whereas interference has a logarithmic impact (due to improvements in effective signal-to-interference-and-noise ratio (SINR)). Thus, it is attractive to consider orthogonal systems in conjunction with techniques for mitigating interference. In fact, the numerical results reported in [3] strongly suggest this: with a choice of  $T_0F_0 = 2$  (half the spectral efficiency of orthogonal systems), they report an MSE/interference improvement by a factor of 2 at most.

In this paper, we develop a general framework for orthogonal STF signaling over time-varying multipath channels and study its performance from both communication and information-theoretic viewpoints. Our analysis is based on the wide-sense stationary uncorrelated scattering (WSSUS) model for doubly dispersive channels [11], [12]. We consider complete orthogonal STF bases and assume that the pulse prototype is given *a priori*. A key to reliable communication is a fundamental understanding of the *interaction* between the signaling basis and the channel. In this context, the key channel parameters are the multipath spread  $T_m$  and the Doppler spread  $B_d$ . In particular, our focus is on *underspread* wireless channels (for which  $T_mB_d < 1$ ) since most radio-frequency wireless channels fall within this category [12]. The key corresponding basis parameters are the support of basis functions in time and in frequency. The starting point of our work are two attractive properties of orthogonal STF basis functions first reported in [4], [3]: 1) An appropriately chosen STF basis serves as a set of *approximate eigenfunctions* for underspread linear time-varying systems, and 2) the prototype pulse’s duration and bandwidth can be matched to the delay and Doppler spreads of the channel (pulse scale adaptation) to minimize the interference between basis functions.

The first contribution of this paper is a rigorous investigation of these two properties from a communication-theoretic viewpoint. Specifically, we derive an exact expression for the received SINR associated with each basis function and show that it is identical for all basis functions due to the channel stationarity in time and frequency in the WSSUS model. The SINR depends on an interference index, governed by the inter-

action between the channel and the pulse prototype, and maximizing the SINR is equivalent to minimizing the interference index. In [4], a similar interference minimization criterion was proposed from the viewpoint of “diagonalization” (eigenproperty) of the system operator, and in [3] it was proposed from the viewpoint of minimizing the MSE in the reconstruction of the transmitted symbols at the receiver. We then cast minimization of the interference index as an optimization problem over the pulse scale and analytically derive the optimal pulse scale adaptation rule for a class of pulses. The pulse scale adaptation rule was proposed in [4], [3] based on heuristic arguments by studying specific pulse shapes. To our knowledge, our derivation of the optimal pulse scaling rule is the most rigorous and most appropriate from a communication-theoretic viewpoint (SINR maximization). Furthermore, the derivation of the pulse scaling rule is the only overlap of our work with existing work on STF signaling.

We now provide a summary of the remaining results in this paper that build on optimal pulse scale adaptation. Overall, our results highlight the critical effect of the channel spread factor  $T_mB_d$  from a communication and information-theoretic viewpoint. For a given delay spread  $T_m$ , small spread factors correspond to *slowly* fading channels (small  $B_d$ ), while large spread factors correspond to *fast* fading channels (large  $B_d$ ). For sufficiently small spread factors, the residual interference can be made very small after pulse scale adaptation and thus the corresponding STF basis functions serve as approximate eigenfunctions of the channel. We show that the approximate eigendecomposition yields an intuitively appealing *block-fading* interpretation of the effect of the channel in terms of time–frequency coherence subspaces: the channel remains highly correlated within each coherence subspace whereas it is approximately uncorrelated across different coherence subspaces. The number of independent coherence subspaces equals the delay-Doppler diversity afforded by the channel which is proportional to  $T_mB_d$  [13]. On the other hand, channels with larger spread factors exhibit significant residual interference even after pulse-scale adaptation. For such rapidly time-varying channels, we propose a highly effective residual interference cancellation technique, sequential iterative interference cancellation (SIIC), which yields an impressive performance gain. Thus, the approximate block-fading model is applicable to larger values of  $T_mB_d$  when pulse scale adaptation is used in conjunction with interference cancellation at the receiver.

The block-fading channel interpretation in terms of time–frequency coherence subspaces also yields an approximate expression for the coherent capacity of doubly dispersive channels. Essentially, the capacity of the dispersive channel can be viewed as the capacity of  $D$  parallel independent flat-fading channels, where  $D$  is the number of independent coherence subspaces (and the level of delay-Doppler diversity). Both ergodic and outage capacities can be estimated using this block-fading model and also facilitate the comparison between orthogonal and biorthogonal systems from a capacity perspective. As the channel spread factor increases, the channel diversity  $D$  increases thereby improving outage capacity performance, whereas the ergodic capacity deviates from that of a flat-fading channel due to increased interference between

<sup>1</sup>The dimension of the space of signals with duration  $T$  and bandwidth  $W$  is approximately  $TW$  for large  $TW$ . The number of basis functions in an STF basis is given by  $N = TW/T_0F_0$ . Thus,  $N = TW$  for an orthogonal system whereas  $N' < TW$  for a biorthogonal system, yielding  $N'/N = 1/T_0F_0 < 1$  since  $T_0F_0 > 1$ .

basis functions. We quantify this deviation by investigating the moment behavior of the eigenvalue distribution associated with doubly dispersive channels using perturbation analysis. Our result bounds the discrepancy between the moments associated with doubly dispersive channels and flat-fading channels, and shows that the discrepancy is proportional to channel spread factor.

Our analytical and numerical results indicate that the proposed framework for orthogonal STF signaling is effective for values of  $T_m B_d$  as large as 0.01 and is most advantageous over conventional methods for channels with relatively large spread factors ( $10^{-3}$ – $10^{-2}$ ). We note that this range of spread factors covers all practical wireless communication channels and is also applicable to various other radio channels. For example, [12, Table 14-2-1] lists typical values of spread factor for several radio channels, whose spread factors are within the applicable range of our framework in most cases. For wireless cellular applications, the CDMA2000 “Vehicular B” channel gives a spread factor of  $10^{-4}$ – $10^{-3}$  at a maximum vehicle speed of 40 km/h.

The paper is organized as follows. Section II briefly reviews general orthogonal signaling over doubly dispersive channels and introduces STF signaling. Section III focuses on performance analysis of STF signaling and derivation of the pulse scale adaptation rule. Section IV discusses the proposed general time–frequency signaling framework, including the SIIC algorithm for interference cancellation and the block-fading interpretation in terms of time–frequency coherence subspaces. Information-theoretic aspects related to capacity of doubly dispersive channels are studied in Section V. Concluding remarks and pointers for future work are presented in Section VI and many of the proofs are relegated to the Appendix.

## II. SYSTEM MODEL

The (complex) baseband doubly dispersive channel can be modeled as a random linear operator  $\mathcal{H}$  with kernel  $\tilde{h}(\tau, t)$  [12], [11], [14]

$$\begin{aligned} \mathcal{H}(x(t)) &= \int_0^{T_m} \tilde{h}(\tau, t) x(t - \tau) d\tau \\ &= \int_0^{T_m} \int_{-B_d/2}^{B_d/2} h(\tau, \nu) x(t - \tau) e^{j2\pi\nu t} d\tau d\nu \quad (1) \end{aligned}$$

where  $x(t)$  is the channel input and the kernel  $h(\tau, \nu)$  is called the delay-Doppler spread function, which is a random process in both  $\tau$  and  $\nu$  [11]. The largest delay  $T_m$  produced by the channel is called the *multipath spread* and the largest Doppler shift  $B_d$  is called the *Doppler spread*. A wide variety of wireless environments can be fairly accurately described by the WSSUS model, under which different delays and Doppler shifts are uncorrelated

$$\mathbb{E}[h(\tau, \nu) h^*(\tau_1, \nu_1)] = S(\tau, \nu) \delta(\tau - \tau_1) \delta(\nu - \nu_1) \quad (2)$$

where  $*$  denotes the complex conjugation and the nonnegative  $S(\tau, \nu)$  is called the *scattering function* of the channel. Projections of the scattering function along  $\tau$  and  $\nu$  are called the delay

power profile and Doppler power profile, respectively. Without loss of generality, we assume channel multipath coefficients  $\tilde{h}(\tau, t)$  to be zero mean with total unit power, that is,

$$\int_0^{T_m} \int_{-B_d/2}^{B_d/2} S(\tau, \nu) d\tau d\nu = 1. \quad (3)$$

The Doppler spread  $B_d$  is a measure of time variation in the channel—the larger the value, the more rapidly the channel changes in time. Its reciprocal,  $T_{\text{coh}} = \frac{1}{B_d}$ , is called the *coherence time*, within which channel remains strongly correlated. Analogously, channel frequency response within the channel *coherence bandwidth*,  $B_{\text{coh}} = \frac{1}{T_m}$ , is strongly correlated. The product  $T_m B_d$  is called the channel *spread factor*. If  $T_m B_d < 1$ , the channel is said to be *underspread*; otherwise, it is *overspread*. The spread parameters critically control communication performance over doubly dispersive channels.

Let  $\{\phi_{lm}(t)\}_{(l,m) \in \mathcal{I}}$  be a (complex) orthonormal basis in  $L^2(\mathbf{R})$ . *Orthogonal signaling* modulates transmitted symbols  $x_{lm}$  onto the orthonormal basis by

$$x(t) = \sum_{l,m} x_{lm} \phi_{lm}(t). \quad (4)$$

Given signaling duration  $T$  and (two-sided) bandwidth  $B$ , the basis functions span a signal subspace with dimension approximately being  $N = \lceil TB \rceil$ , where  $\lceil x \rceil$  is the least integer not less than  $x$ . After matched filter processing at the receiver, the received symbol can be written as

$$y_{lm} = \sum_{l',m'} c_{lm,l'm'} x_{l'm'} + n_{lm} \quad (5)$$

or, equivalently, in a matrix form

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{n} \quad (6)$$

where transmit symbol  $x_{lm}$  has power  $E_s$  and the AWGN has variance  $\sigma^2$ , that is,  $n_{lm} \sim \text{CN}(0, \sigma^2)$ . The coefficient  $c_{lm,l'm'}$  of  $\mathbf{C}$  represents the coupling produced by the channel between the transmit basis function  $\phi_{l'm'}(t)$  and the receive basis function  $\phi_{lm}(t)$

$$c_{lm,l'm'} = \langle \mathcal{H}(\phi_{l'm'}(t)), \phi_{lm} \rangle = \int \mathcal{H}(\phi_{l'm'}(t)) \phi_{lm}^*(t) dt \quad (7)$$

where

$$\langle a(t), b(t) \rangle = \int a(t) b^*(t) dt$$

denotes the inner product and

$$\|a(t)\| = \sqrt{\langle a(t), a(t) \rangle}$$

is the induced  $L^2$ -norm.

Ideally, choosing  $\{\phi_{lm}(t)\}$  to be eigenfunctions of the channel will render a *diagonal* channel matrix  $\mathbf{C}$ , in which case the channel is said to be *flat fading*, that is, the channel effect for each basis function reduces to a multiplicative scalar and there is no interference among basis functions. However, unlike time-invariant linear channels for which sinusoids are always eigenfunctions, there are no *fixed* eigenfunctions for general

*time-varying* linear channels. If the transmitter has the knowledge of channel realization, it may choose to dynamically adapt its transmit basis to channel eigenfunctions. But this procedure can be very undesirable, if not totally impractical, from the viewpoint of implementation and system complexity.

Instead, we focus on STF basis for it is matched to the time–frequency characteristics of doubly dispersive channel: the channel produces time and frequency shifts of the transmit signal and the STF basis is generated from a prototype function via time and frequency shifts. We shall demonstrate that STF basis functions exhibits good signaling performance, thus presenting an attractive choice for underspread channels for which they serve as approximate eigenfunctions.

*Definition 1:* The STF basis is defined as

$$\phi_{lm}(t) = g(t - lT_0)e^{j2\pi mF_0t}, \quad (l, m) \in \mathbf{Z}^2 \quad (8)$$

where  $g(t)$  is the normalized ( $\|g\| = 1$ ) prototype pulse,  $T_0$  is the time separation, and  $F_0$  is the frequency separation between basis functions. The mean of  $g(t)$  in time and frequency are defined to be

$$\mu_t = \int t|g(t)|^2 dt, \quad \mu_f = \int f|\hat{g}(f)|^2 df \quad (9)$$

where  $\hat{g}(f)$  is the Fourier transform of  $g(t)$ . The variance in time and frequency are defined correspondingly as

$$\sigma_t^2 = \int (t - \mu_t)^2 |g(t)|^2 dt, \quad \sigma_f^2 = \int (f - \mu_f)^2 |\hat{g}(f)|^2 df. \quad (10)$$

The product  $\kappa = \sigma_t \sigma_f$  is called the *time–frequency spread* of the pulse. Without loss of generality, pulse  $g(t)$  is assumed to be centered in time and frequency, that is,  $\mu_t = 0$  and  $\mu_f = 0$ .

*Remark 1:* The variance in frequency for a rectangular pulse does not exist due to slow decay of its spectrum. In this case, we define  $\sigma_t = T_0$  and  $\sigma_f = \frac{1}{\sigma_t}$ .

The STF basis falls within the framework of the so-called *Gabor systems* (see, e.g., [15], [6]). The pulse time–frequency spread  $\kappa$  is a measure of its time–frequency localization—the smaller the spread, the better the localization. The Heisenberg’s uncertainty principle (see, e.g., [15]) states that

$$\kappa \geq \frac{1}{4\pi} \quad (11)$$

with the equality attained by Gaussian pulses. Since STF basis functions are generated from the prototype via time–frequency shifts, they all have the same time–frequency spread parameters.

As we will see later, a simple *scaling* operation, which shrinks or dilates the pulse in time and frequency, has a profound effect on the performance of STF basis over doubly dispersive channels.

*Definition 2:* Scaling by parameter  $s > 0$  is a mapping defined as<sup>2</sup>

$$f(t) \mapsto \frac{1}{\sqrt{s}} f(t/s). \quad (12)$$

<sup>2</sup>Note that the scaling operation preserves pulse energy.

If the original STF basis function has parameters  $\sigma_t, \sigma_f, T_0$ , and  $F_0$ , then after scaling those parameters become  $s\sigma_t, \frac{\sigma_f}{s}, sT_0$ , and  $\frac{F_0}{s}$ , respectively. However, note that the products  $\kappa = \sigma_t \sigma_f$  (pulse spread) and  $T_0 F_0$  are unchanged.

The product  $T_0 F_0$  is critical to the *completeness* of STF basis in  $L^2(\mathbf{R})$ . For *undercritical* sampling,  $T_0 F_0 > 1$ , orthonormal STF bases exist but are not complete. For critical sampling,  $T_0 F_0 = 1$ , complete orthonormal STF bases exist if the prototype  $g(t)$  satisfies certain conditions [9]. Unfortunately, a complete STF basis has poor time–frequency localization due to the Balian–Low theorem [8], [9]. For *overcritical* sampling,  $T_0 F_0 < 1$ , the basis becomes linearly dependent or redundant. We will primarily consider critically sampled ( $T_0 F_0 = 1$ ) case since it yields complete orthonormal systems. However, the relatively poor time–frequency localization of basis functions in this case makes interference minimization even more critical.

From communications viewpoint, time–frequency localization of basis functions alone does not determine communication performance. It is the interaction between the basis and the channel that determines signaling performance over doubly dispersive channels. It turns out that the pulse *ambiguity function* is key to understanding the interaction between the basis and the channel [3].

*Definition 3:* The ambiguity function of the pulse  $g(t)$ ,  $A_g(\tau, \nu)$ , is defined as

$$\begin{aligned} A_g(\tau, \nu) &= \langle g(t), g(t - \tau) e^{j2\pi \nu t} \rangle \\ &= \int g(t) g^*(t - \tau) e^{-j2\pi \nu t} dt. \end{aligned} \quad (13)$$

One can easily verify the following well-known properties.

*Proposition 1:*

$$A_g(0, 0) = \|g\|^2, \quad A_g(\tau, \nu) = e^{-j2\pi \tau \nu} A_g^*(-\tau, -\nu). \quad (14)$$

Moreover, the following holds for orthonormal basis functions:

$$A_g(0, 0) = 1 \text{ and } A_g(lT_0, mF_0) = 0, \quad \forall (l, m) \neq (0, 0). \quad (15)$$

### III. PERFORMANCE ANALYSIS AND INTERFERENCE MINIMIZATION

In general, interference exists among basis functions in doubly dispersive channels. We rewrite (6) as

$$y_{lm} = c_{lm,lm} x_{lm} + \sum_{l' \neq l \text{ or } m' \neq m} c_{lm,l'm'} x_{l'm'} + n_{lm} \quad (16)$$

where the second term represents the interference toward symbol  $x_{lm}$  by other symbols. The interference term in (16) encompasses two types of interference. One is the *intersymbol interference* (ISI), which exists between adjacent time slots ( $l' \neq l$ ) and is caused by channel delay spread, the other is the *intercarrier interference* (ICI), which happens between adjacent frequencies ( $m' \neq m$ ) and is induced by channel Doppler spread. In principle, better time–frequency localization of the basis offers better immunity against channel dispersion. Well-localized undercritical Gabor systems have been constructed in [3], [5], [6]. Since an undercritical basis ( $T_0 F_0 > 1$ )

often incurs a loss in transmission rate, we are primarily concerned with a complete STF basis ( $T_0 F_0 = 1$ ). We also assume that the prototype pulse has been given *a priori*. (Pulse prototype design in the context of this work has been studied in [4], [3], [6].) We are interested in the impact of interference on system performance and methods of reducing interference for a given prototype.

#### A. Signal-to-Interference-and-Noise Ratio

We first calculate the entries of the channel matrix  $\mathbf{C}$  in (6).

*Proposition 2:*

$$c_{lm,l'm'} = \int \int h(\tau, \nu) A_g^* \cdot ((l' - l)T_0 + \tau, (m' - m)F_0 + \nu) \cdot e^{-j2\pi m' F_0 \tau} e^{j2\pi l T_0 \nu} d\tau d\nu \quad (17)$$

$$\mathbb{E}[|c_{lm,l'm'}|^2] = \int \int S(\tau, \nu) \cdot \left| A_g((l' - l)T_0 + \tau, (m' - m)F_0 + \nu) \right|^2 \cdot d\tau d\nu. \quad (18)$$

*Proof:* See Appendix I.  $\square$

Because of the time-frequency shift structure of the STF basis and the WSSUS channel assumption, the variance of channel coefficients is invariant under time-frequency shifts, which can be seen from (18) to depend on the difference of time-frequency indices ( $l - l'$  and  $m - m'$ ). Consequently, the entries on the main diagonal of  $\mathbf{C}$  all have the same variance  $\eta$

$$\eta = \mathbb{E}|c_{lm,lm}|^2 = \int_0^{T_m} \int_{-B_d/2}^{B_d/2} S(\tau, \nu) |A_g(\tau, \nu)|^2 d\tau d\nu. \quad (19)$$

Our next result quantifies the amount of interference associated with STF signaling over doubly dispersive channels and relates it to channel and basis parameters. We assume that the channel coefficients are independent of transmitted information symbols.

*Lemma 1:* Define the SINR of receive symbol  $y_{lm}$  as

$$\text{SINR}_{lm} = \frac{\eta E_s}{\sigma^2 + E_s \sum_{l' \neq l \text{ or } m' \neq m} \mathbb{E}|c_{lm,l'm'}|^2}. \quad (20)$$

Then the following holds:

$$\text{SINR} = \text{SINR}_{lm} = \frac{\eta E_s}{\sigma^2 + (1 - \eta) E_s} = \frac{\eta E_s}{\sigma^2 + \rho E_s} \quad (21)$$

where  $\eta$  is given in (19) and  $\rho = 1 - \eta$  is called the *interference index*.

*Proof:* The completeness of the underlining orthonormal basis implies that

$$\begin{aligned} \mathcal{H}(\phi_{lm}(t)) &= \sum_{l'm'} \langle \mathcal{H}(\phi_{lm}(t)), \phi_{l'm'} \rangle \phi_{l'm'}(t) \\ &= \sum_{l'm'} c_{l'm',lm} \phi_{l'm'}(t) \end{aligned} \quad (22)$$

and hence,

$$\|\mathcal{H}(\phi_{lm}(t))\|^2 = \int |\mathcal{H}(\phi_{lm}(t))|^2 dt = \sum_{l'm'} |c_{l'm',lm}|^2. \quad (23)$$

It follows from the definition of  $\mathcal{H}$  and power normalization of (3) that

$$\begin{aligned} \mathbb{E}[\|\mathcal{H}(\phi_{lm}(t))\|^2] &= \mathbb{E} \int h(\tau, \nu) \phi_{lm}(t - \tau) e^{j2\pi \nu t} h^*(\tau_1, \nu_1) \phi_{lm}^*(t - \tau_1) \\ &\quad \cdot e^{-j2\pi \nu_1 t} d\tau d\nu d\tau_1 d\nu_1 dt \\ &= \int \int S(\tau, \nu) \left( \int |\phi_{lm}(t - \tau)|^2 dt \right) d\tau d\nu = 1. \end{aligned}$$

So we have

$$\sum_{l'm'} \mathbb{E}|c_{l'm',lm}|^2 = 1.$$

Next we show  $\sum_{l'm'} \mathbb{E}|c_{lm,l'm'}|^2 = \sum_{l'm'} \mathbb{E}|c_{l'm',lm}|^2$

$$\begin{aligned} \sum_{l'm'} \mathbb{E}|c_{l'm',lm}|^2 &= \sum_{l'm'} \int \int S(\tau, \nu) |A_g((l - l')T_0 + \tau, (m - m')F_0 + \nu)|^2 \\ &= \int \int S(\tau, \nu) \sum_{k_1, k_2} |A_g(k_1 T_0 + \tau, k_2 F_0 + \nu)|^2 \\ \sum_{l'm'} \mathbb{E}|c_{lm,l'm'}|^2 &= \sum_{l'm'} \int \int S(\tau, \nu) |A_g((l' - l)T_0 + \tau, (m' - m)F_0 + \nu)|^2 \\ &= \int \int S(\tau, \nu) \sum_{k_1, k_2} |A_g(k_1 T_0 + \tau, k_2 F_0 + \nu)|^2 \end{aligned}$$

from which the lemma follows.  $\square$

*Remark 2:* The average distortion for a given basis function  $\phi_{lm}(t)$  is

$$\epsilon_{lm} = \mathbb{E} \left[ \min_a \|\mathcal{H}(\phi_{lm}(t)) - a \phi_{lm}(t)\|^2 \right]. \quad (24)$$

One can show that  $\epsilon = 1 - \eta = \rho$ , that is, the interference index is equal to the average distortion encountered by STF basis functions [16]. Thus, maximizing SINR is equivalent to minimizing the distortion of basis functions passing through the doubly dispersive channel.

#### B. Interference Minimization

Fig. 1 illustrates the basic notion of pulse scaling in the time-frequency domain to minimize interference. If the channel is nonselective in time and frequency ( $T_m = 0$  and  $B_d = 0$ ),  $\mathcal{H}$  degenerates to a random scalar and hence any choice of orthonormal basis will completely avoid interference. When the channel is time selective only ( $B_d \neq 0$  and  $T_m = 0$ ), the optimal pulse is peaky in time, which is analogous to

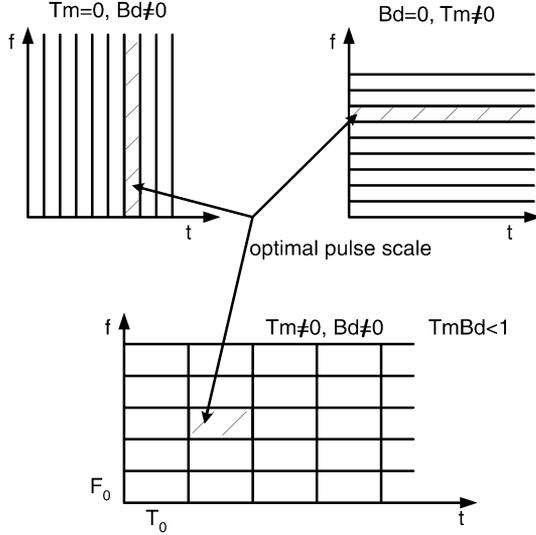


Fig. 1. A schematic illustrating the optimal pulse scaling.

time-division multiplexing (TDM). When the channel is frequency selective only ( $T_m \neq 0$  and  $B_d = 0$ ), the optimal pulse is peaky in frequency, which corresponds to frequency-division multiplexing (FDM). For the general case ( $T_m B_d \neq 0$ ), the optimal pulse scale varies between the above two extremes.

When a channel is dispersive in both time and frequency, thus causing both ISI and ICI, the pulse scale can significantly affect system performance. A long pulse in time can reduce ISI but is prone to frequency dispersion due to temporal channel variations. On the other hand, a short pulse helps reduce ICI but is at the cost of higher ISI due to relatively large delay spread with respect to pulse duration. Therefore, one can expect an optimal pulse scale that minimizes both ICI and ISI jointly. Such a pulse adaptation, given channel statistics and pulse prototype, can be cast systematically as the following optimization problem in light of Lemma 1:

$$\eta_{\text{opt}} = \max_{\sigma_t} \int_0^{T_m} \int_{-B_d/2}^{B_d/2} S(\tau, \nu) |A_g^{\sigma_t}(\tau, \nu)|^2 d\tau d\nu \quad \text{subject to } \sigma_t \sigma_f = \kappa \quad (25)$$

where  $\kappa \geq \frac{1}{4\pi}$  is the pulse time-frequency spread, a constant determined by the given pulse prototype. We note that the pulse ambiguity function is parameterized by pulse scale.

Most of the pulse energy is concentrated within a time duration  $T_0$  and a frequency band  $F_0$  (such as may be measured from the support of the pulse ambiguity function). So, if  $T_0$  is large compared to the channel delay spread while small compared to coherence time, it will approximately experience flat fading, thus resulting in small distortion. Analogous conditions can also be expressed in the frequency domain as well. In summary, the following conditions are necessary for small interference:

$$T_m \leq T_0 \leq T_{\text{coh}} = \frac{1}{B_d}, \quad B_d \leq F_0 \leq B_{\text{coh}} = \frac{1}{T_m}. \quad (26)$$

Multiplying together both inequalities in (26) and using  $T_0 F_0 = 1$ , we obtain  $T_m B_d \leq 1$ , which implies that basis design for

low interference is feasible for underspread channels but not for general overspread ( $T_m B_d > 1$ ) channels.

Solving the optimization problem in (25) often involves complicated numerical methods. The above intuitive discussion on interference minimization suggests an “equal” footing for time and frequency. We next give a simple scaling rule whose derivation is relegated to Appendix II. Also shown in Appendix II, the integral in (25) can be written as a function of  $a$  where  $a = \frac{T_m}{\sigma_t}$ .

*Theorem 1 (Optimal Pulse Scale):* Assume flat multipath and Doppler power profiles for the channel. Further assume that the pulse is separable and symmetric in time and frequency, that is,

$$|A_g(\tau, \nu)|^2 = \psi\left(\frac{\tau}{\sigma_t}\right)\psi\left(\frac{\nu}{\sigma_f}\right), \quad \sigma_t \sigma_f = \kappa \quad (27)$$

where  $\psi(x) \geq 0$  is differentiable. Define

$$f(x) = \psi'(x) + \frac{\psi(x)}{x} - \frac{\psi^2(x)}{\int_0^x \psi(t) dt}. \quad (28)$$

Then, a stationary point  $\bar{a}$  of (25) is <sup>3</sup>

$$\bar{a} = \frac{T_m}{\sigma_t} = \frac{B_d/2}{\sigma_f} = \sqrt{\frac{T_m B_d}{2\kappa}} \quad (29)$$

which is a local maximum if  $f(\bar{a}) < 0$ . A sufficient condition for local maximum is given by

$$\psi'(\bar{a}) < \frac{\psi(\bar{a})}{\bar{a}} (\psi(\bar{a}) - 1). \quad (30)$$

The corresponding optimal interference index  $\rho$  is given by

$$\rho = 1 - \eta = 1 - \frac{1}{\bar{a}^2} \left( \int_0^{\bar{a}} \psi(x) dx \right)^2. \quad (31)$$

*Remark 3:* When  $T_m B_d$  is sufficiently small, the minimal interference index in (31) can be approximated by

$$\rho \approx 1 - \psi^2 \left( \sqrt{\frac{T_m B_d}{2\kappa}} \right).$$

Since  $\psi(x)$  is a decreasing function in the neighborhood of 0, we see that smaller  $T_m B_d$  results in smaller interference. Moreover, the minimal interference depends only on the channel spread factor  $T_m B_d$  and the pulse spread factor  $\kappa$ .

*Remark 4:* The optimal scaling rule matches the duration and bandwidth of prototype pulse relatively to the channel delay and Doppler spread:  $\frac{\sigma_t}{\sigma_f} = \frac{T_m}{B_d/2}$ .

*Remark 5:* This work assumes no relative time delay between receiver and transmitter. It turns out one can optimize such time offset to further improve receiving signal-to-noise ratio (SNR). As studied in [17],  $T_m/2$  offset yields best performance, in which case the above scaling rule shall be modified accordingly as  $\frac{\sigma_t}{\sigma_f} = \frac{T_m}{B_d}$ .

We now apply the above methodology to study the performance of Gaussian and rectangular pulses. Due to space limi-

<sup>3</sup>The same relation has been derived in [4], [3] using different arguments.

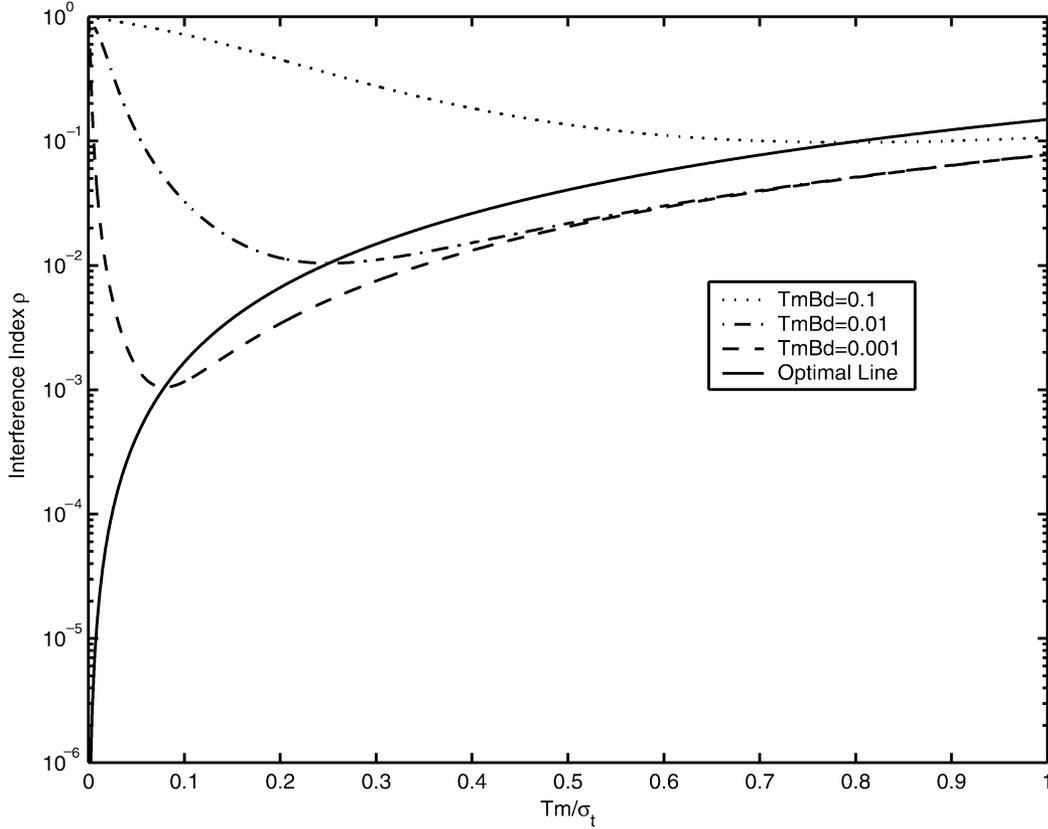


Fig. 2. Interference index of a Gaussian prototype as a function of time scale for different channel spread factors. Flat multipath and Doppler power profiles are used.

tation, the derivations of our analytical expressions are left out but can be found in [16].

The Gaussian pulse with variance  $\sigma_t^2$  and  $\sigma_f^2$  is defined as

$$g(t) = Ae^{-\frac{t^2}{4\sigma_t^2}}, \quad \sigma_t\sigma_f = \frac{1}{4\pi}$$

where the normalization factor  $A = (2\pi)^{-\frac{1}{4}}\sigma_t^{-\frac{1}{2}}$  is chosen to normalize pulse energy. Under flat delay and Doppler power profiles, one has

$$\eta = \frac{1}{2T_m B_d} \left(1 - 2Q\left(\frac{T_m}{\sigma_t\sqrt{2}}\right)\right) \left(1 - 2Q\left(\frac{B_d/2}{\sigma_f\sqrt{2}}\right)\right) \quad (32)$$

where  $Q(x) = \frac{1}{2\pi} \int_x^\infty e^{-t^2/2} dt$ . Here we emphasize that  $\eta$  in general is a function of  $T_m$  and  $B_d$ , but the *optimal* value of  $\eta$  depends approximately on the product  $T_m B_d$  as in Theorem 1.

In Fig. 2, we evaluate (32) for a Gaussian pulse and plot the interference index  $\rho$  versus the relative scale  $\frac{T_m}{\sigma_t}$  for different channel spread factors. Note that due to the time–frequency duality and optimal pulse scaling rule, it is the spread factor  $T_m B_d$  rather than the individual  $T_m$  or  $B_d$  that determines the optimal performance. The plots, being based on the relative scale  $T_m/\sigma_t$ , reflect system performance for arbitrary values of  $T_m > 0$  and  $B_d > 0$  such that  $T_m B_d$  is fixed. We also plot the  $\rho$  locus, denoted as the approximate optimal line, corresponding to the approximate scaling rule in Theorem 1. It is evident from the figure that this line intersects with the interference curves at their global minimum points. Moreover, the figure illustrates the

effect of the channel spread factor: the smaller the spread factor, the smaller the interference. The Gaussian pulse exhibits excellent immunity against channel distortion thanks to its excellent time–frequency localization. For example, for a spread factor of 0.001, the minimal interference is less than  $10^{-3}$ , which gives about 30-dB SINR floor in high-SNR regime. However, as  $T_0 F_0 \rightarrow 1$ , STF basis generated by a Gaussian pulse becomes unstable [18], thereby limiting its practical utility. Nevertheless, it provides an upper bound on performance.

The rectangular pulse with variance  $\sigma_t^2$  and  $\sigma_f^2$  ( $\sigma_t\sigma_f = 1$ ) is defined as

$$g(t) = \begin{cases} \frac{1}{\sqrt{\sigma_t}}, & |t| < \frac{\sigma_t}{2} \\ 0, & \text{otherwise.} \end{cases}$$

Under flat-delay and Doppler power profiles

$$\eta = \frac{1}{T_m B_d \sigma_t^2 \pi^2} \int_0^{B_d/2} \frac{1}{\nu^2} [T_m - \sigma_t \text{sinc}(2\nu\sigma_t) + (\sigma_t - T_m) \text{sinc}(2\nu(\sigma_t - T_m))] d\nu \quad (33)$$

where  $\text{sinc}(x) = \frac{\sin \pi x}{\pi x}$ . And under the more realistic “bathtub” Doppler profile

$$S_\nu(\nu) = \frac{2}{\pi B_d} \frac{1}{\sqrt{1 - (2\nu/B_d)^2}}, \quad |\nu| \leq B_d/2 \quad (34)$$

one has

$$\eta = \frac{1}{T_m \sigma_t^2 \pi^2} \int_0^{B_d/2} \frac{S_\nu(\nu)}{\nu^2} [T_m - \sigma_t \text{sinc}(2\nu\sigma_t) + (\sigma_t - T_m) \text{sinc}(2\nu(\sigma_t - T_m))] d\nu. \quad (35)$$

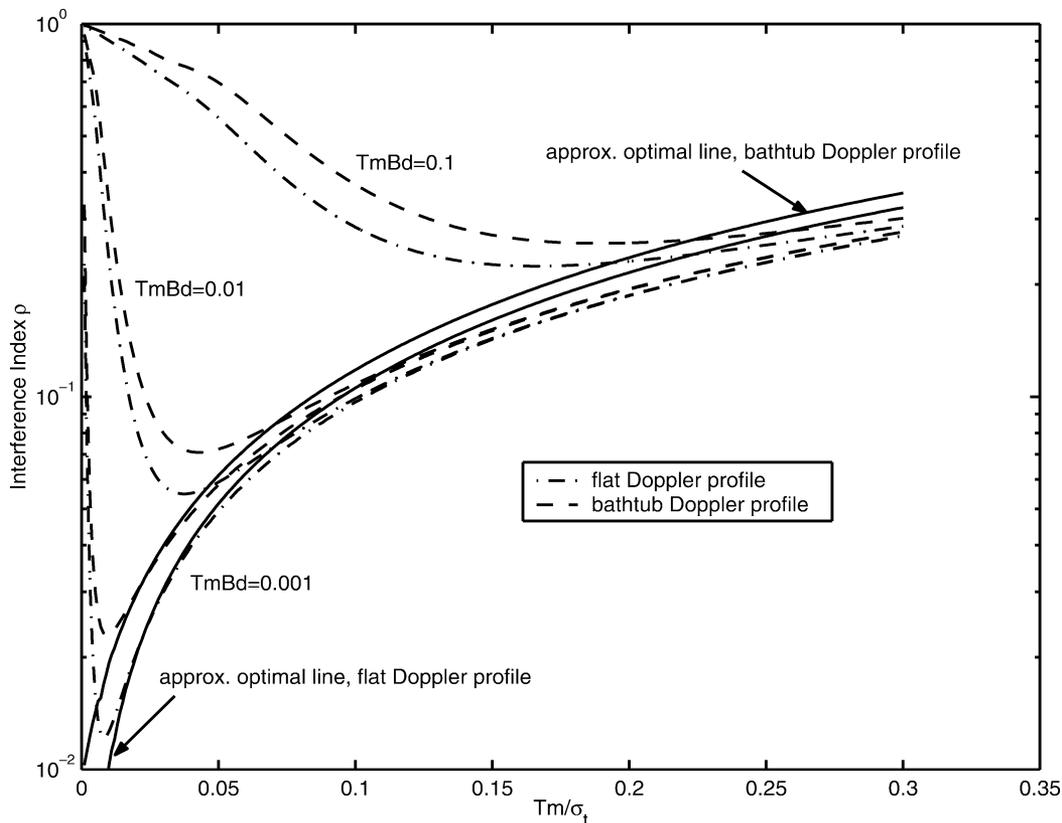


Fig. 3. Interference index for a rectangular pulse as a function of time scale for different channel spread factors. Flat multipath power profile is used. Both “bathtub” and flat Doppler profiles are used.

We plot the interference index  $\rho$  in Fig. 3 by numerically evaluating (33) and (35). The interference curves are shown to be very close for the two Doppler profiles. It is seen from the figure that the approximate optimal line intersects the two interference curves near their true minimum points.

Fig. 4 plots the (approximate) minimal interference index versus the channel spread factor  $T_m B_d$  for both rectangular and Gaussian pulses ( $B_d = 10$  Hz). The scale adaptation rule in Theorem 1 is used. The figure includes the analytical curves and the actual simulation data. The simulated channel is generated by the Jakes model [19], which inherently corresponds to a “bathtub” Doppler profile. It is seen that the minimum interference index is an increasing function of the spread factor. It is also evident that the Gaussian pulse outperforms the rectangular pulse. However, the performance gap vanishes as  $T_m B_d$  decreases. The analysis and simulation results for the rectangular pulse show that it undergoes rather significant channel distortion, even with optimal pulse scale, due to its poor time–frequency localization.

#### IV. GENERAL SIGNALING FRAMEWORK AND INTERFERENCE CANCELLATION

We have shown that an appropriate pulse scale can be chosen to match the channel spread parameters of underspread channels so as to minimize interference among STF basis functions. Thus, the channel matrix  $\mathbf{C}$  will have dominant diagonal entries corresponding to a strong signal energy component. For sufficiently small channel spread factors, interference can be made

so small that  $\mathbf{C}$  is approximately diagonal, in which case the STF basis functions serve as approximate eigenfunctions. The choice of an orthogonal STF basis is motivated by the fact that it is complete and hence preserves bandwidth efficiency. However, some residual interference remains even after pulse scale adaptation due to the relatively poor time–frequency localization of the STF basis. To mitigate the residual interference, we propose a low-complexity but highly effective interference cancellation technique, SIIC, which will be described later in this section.

##### A. General Signaling Framework

Pulse scale adaptation for interference minimization and the use of SIIC for residual interference removal suggest a general orthogonal time–frequency signaling framework for underspread channels as depicted in Fig. 5. Pulse design can be incorporated into the framework but that is beyond the scope of the paper. We focus on a single-user context to illustrate the framework. It is a natural extension of conventional OFDM signaling to doubly dispersive channels. The transmitter modulates the (coded) symbols onto the STF basis functions for communication over the channel. At the receiver, a bank of correlators or matched filters are used to generate the sufficient statistics for decoding the information symbols. Coherent communication requires channel estimation, which can be tackled in a variety of ways using our framework (see, e.g., [20]). Here we assume that perfect estimates of the matrix  $\mathbf{C}$  are available for simplicity.

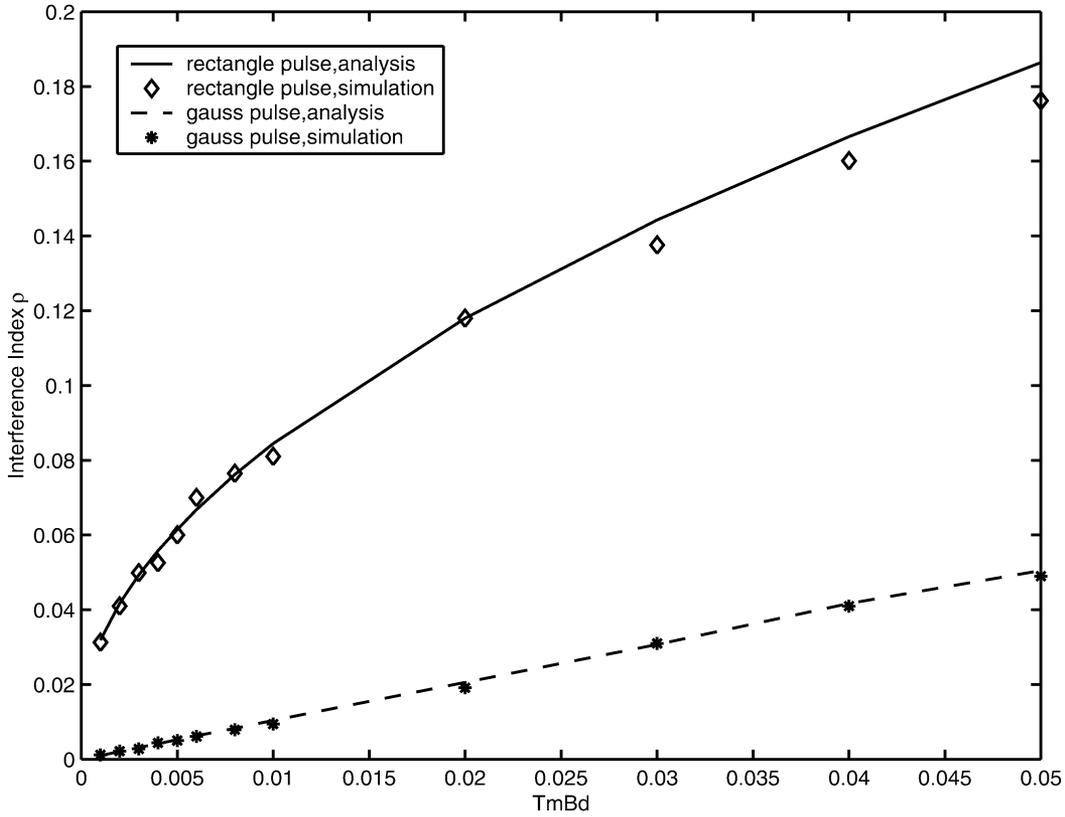


Fig. 4. Interference index for Gaussian and rectangular prototypes using the approximately optimal pulse scale. Flat multipath and “bathtub” Doppler profiles are used for analysis and simulation.

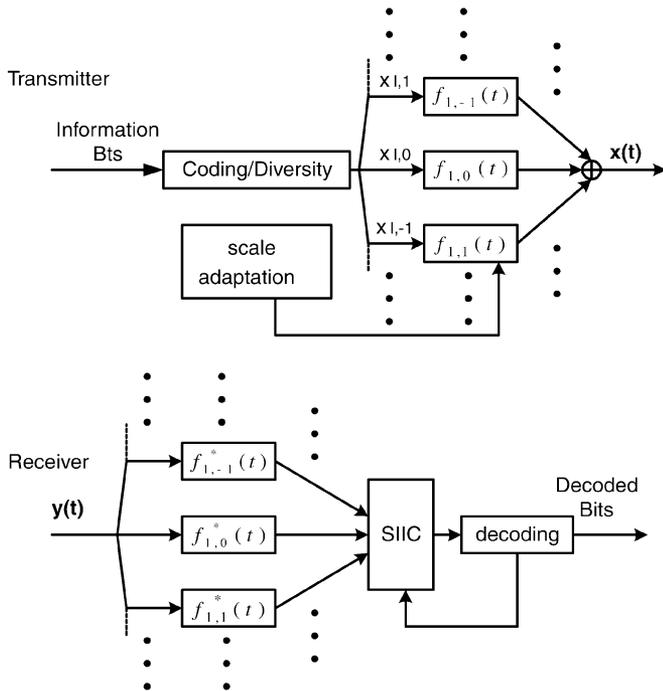


Fig. 5. General time-frequency signaling framework.

One important aspect of the STF signaling scheme is that it clearly reveals and facilitates the exploitation of channel diversity afforded by the time-varying multipath fading channel. The

auto-term coefficient of the  $(l, m)$ th receive signal is given in Proposition 2 as

$$c_{lm,lm} = \int \int h(\tau, \nu) A_g(\tau, \nu) e^{-j2\pi m F_0 \tau} e^{j2\pi l T_0 \nu} d\tau d\nu. \quad (36)$$

After pulse scale adaptation,  $\frac{T_m}{\sigma_t}$  and  $\frac{B_d/2}{\sigma_f}$  are much less than 1 for interference minimization. Under these conditions,  $A_g(\tau, \nu)$  can be approximated by 1 in the integration range, and hence,

$$c_{lm,lm} \approx \int \int h(\tau, \nu) e^{-j2\pi m F_0 \tau} e^{j2\pi l T_0 \nu} d\tau d\nu = H(mF_0, lT_0) \quad (37)$$

where the time-frequency kernel  $H(f, t)$  of the doubly dispersive channel is defined as

$$H(f, t) = \int \tilde{h}(\tau, t) e^{-j2\pi f \tau} d\tau = \int \int h(\tau, \nu) e^{-j2\pi f \tau} e^{j2\pi t \nu} d\tau d\nu. \quad (38)$$

It is easy to check that the correlation function of  $H(f, t)$  is related to the channel scattering function as

$$R_{f,t}(\Delta f, \Delta t) = \mathbb{E}[H(f, t) H^*(f - \Delta f, t - \Delta t)] = \int \int S(\tau, \nu) e^{-j2\pi \Delta f \tau} e^{j2\pi \Delta t \nu} d\tau d\nu. \quad (39)$$

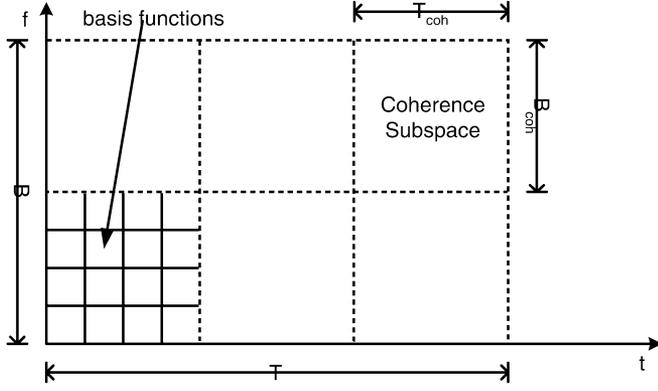


Fig. 6. A schematic illustrating the notion of time-frequency coherence subspaces.

Consequently, the correlation between different auto-term coefficients can be approximated by

$$\mathbb{E} [c_{lm,lm} c_{l_1 m_1, l_1 m_1}^*] \approx R_{f,t}((m - m_1)F_0, (l - l_1)T_0). \quad (40)$$

This demonstrates that the channel memory structure is preserved by the STF signaling scheme. When two basis functions are separated beyond coherence time  $T_{\text{coh}}$  and/or bandwidth  $B_{\text{coh}}$ , they encounter independent fading. On the other hand, when basis functions lie within a time-frequency region with support  $T_{\text{coh}} \times B_{\text{coh}}$ , their channel coefficients are strongly correlated. Therefore, we can partition the entire time-frequency plane by these  $T_{\text{coh}} \times B_{\text{coh}}$  coherence subspaces as illustrated in Fig. 6. Since  $T_0 F_0 = 1$ , the number of basis functions in a coherence subspace is

$$N_c = \lceil T_{\text{coh}} B_{\text{coh}} \rceil = \lceil 1/T_m B_d \rceil.$$

For a total signal duration  $T$  and bandwidth  $B$ , the total level of diversity  $D$  afforded by channel equals the number of coherence subspaces contained within the signal space. More precisely

$$D = \left\lceil \frac{B}{B_{\text{coh}}} \right\rceil \left\lceil \frac{T}{T_{\text{coh}}} \right\rceil = \lceil T_m B \rceil \lceil T B_d \rceil = D_m D_d \quad (41)$$

where  $D_m = \lceil T_m B \rceil$  and  $D_d = \lceil T B_d \rceil$  are, respectively, the levels of *multipath diversity* and *Doppler diversity* [13]. We note that the  $D$  is also the number of channel parameters to be estimated and that  $D \ll N$  for underspread channels. Codes can be designed under the proposed signaling framework to fully utilize channel diversity for reliable communication. For example, a simple time-frequency diversity scheme using this framework was proposed in [21].

### B. Residual Interference Cancellation

Generally speaking, both ISI and ICI contribute to interference in STF signaling. We assume interference can be well approximated by Gaussian distribution (by the usual central limit theorem argument provided the number of interfering basis functions is large). Therefore, the variance of zero-mean effective noise (interference plus channel noise) is actually

quantified by Lemma 1, where  $\eta E_s$  is the received signal power,  $(1 - \eta)E_s$  is the interference variance, and  $\sigma^2$  is the noise variance. As in Fig. 4, the interference index can be reduced to be much less than 0.1 for a wide range of channel parameters, which results in an SNR loss much less than 1 dB. However, the presence of a relatively significant interference can severely limit system performance, for as signal power increases, that is,  $E_s \rightarrow \infty$

$$\text{SINR} = \frac{\eta E_s}{\sigma^2 + (1 - \eta)E_s} \rightarrow \frac{\eta}{1 - \eta} \quad (42)$$

which is fixed by  $\eta$ . For moderate channel spread factors on the order of  $10^{-2}$  (whose interference index is around several percent), the limiting SINR is roughly 15–20 dB according to (42).

We consider binary phase-shift keying (BPSK) modulation with diversity order  $D$  to demonstrate the effect of interference on system performance. Diversity signaling of order  $D$  can be realized by a simple scheme that repeats the same bit on  $D$  different basis functions over disjoint coherence subspaces. The probability of bit error  $P_b$  with diversity order  $D$  is given by [12]

$$P_b = \left[ \frac{1}{2}(1 - \mu) \right]^D \sum_{k=0}^{D-1} \binom{D-1+k}{k} \left[ \frac{1}{2}(1 + \mu) \right]^k \quad (43)$$

where  $\mu = \sqrt{\frac{\gamma_c}{1 + \gamma_c}}$  and  $\gamma_c$  is the effective SNR of each basis function. In our case, it is SINR of each basis function due to the Gaussian assumption for interference.

We first show simulated performance of BPSK modulation without diversity for different channel spread factors. The transmitter and receiver are as in Fig. 5 except that the residual interference cancellation is not used.  $B_d = 10$  Hz is kept fixed and  $T_m$  is changed according to the spread factor  $T_m B_d$ . The STF basis is generated by a rectangular prototype whose scale is chosen by the scaling rule in Theorem 1. The system has finite bandwidth  $B$  such that the number of basis functions in frequency is 128; that is,  $\frac{B}{F_0} = 128$ . The channel is simulated via the Jakes model assuming flat multipath profile and “bathtub” Doppler profile. Fig. 7 plots both the simulated and analytical curves for  $P_b$  as a function of SNR. The figure clearly shows the existence of  $P_b$  floors due to interference—flattening of  $P_b$  as SNR increases. For  $T_m B_d = 0.01$ , the transition point is between 15 and 20 dB. Furthermore, it can be seen that the  $P_b$  floor decreases as  $T_m B_d$  decreases due to reduced interference.

Next we present in detail our proposed SIIC technique for residual interference reduction. The SIIC is a decision feedback type algorithm that jointly removes both ISI and ICI in time-frequency domain. It works sequentially in time. For each time slot, it first cancels the ISI caused by channel delay spread, then it iteratively cancels the ICI among different frequencies within the same time slot, analogous to the parallel interference cancellation (PIC) algorithms in multiuser detection applications (see, e.g., [22], [23]). Due to its decision feedback nature, the performance of SIIC is critically affected by the quality of the decisions used in interference cancellation.

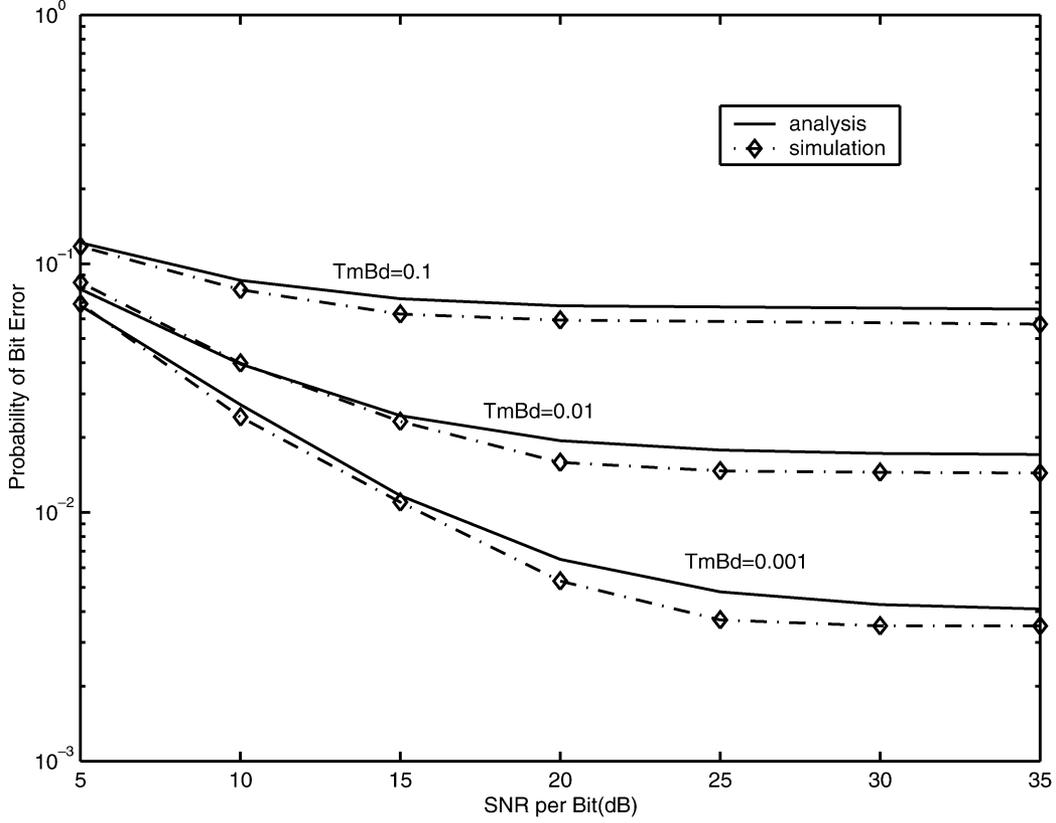


Fig. 7. Performance of STF signaling with BPSK modulation for different spread factors. Rectangular prototype, flat multipath, and “bathtub” Doppler profiles. The pulse scale is determined by approximate pulse scale rule.

In the following, uncoded systems are considered to illustrate the SIIC algorithm. The algorithm can be readily extended to incorporate many coding scheme by modifying its decision generating component. Stacking all the frequency variables corresponding to the same time index  $l$ , that is, let

$$\mathbf{x}_l = (\dots, x_{l,m-1}, x_{lm}, x_{l,m+1}, \dots)^T$$

and

$$\mathbf{y}_l = (\dots, y_{l,m-1}, y_{lm}, y_{l,m+1}, \dots)^T$$

the matrix channel (6) can be written as

$$\mathbf{y}_l = \sum_{d=0}^L \mathbf{C}_{l,l-d} \mathbf{x}_{l-d} + \mathbf{n}_l \quad (44)$$

where  $L$  is the multipath spread in symbols. The  $\mathbf{C}_{l,l-d}$  with  $d \neq 0$  in (44) represents the ISI from previous symbols, while  $\mathbf{C}_{l,l}$  represents the ICI among different frequencies at the same time slot  $l$ . The receiver is assumed to have perfect knowledge about matrices  $\mathbf{C}_{l,l-d}$ . For rectangular pulses,  $T_m \ll \sigma_t$  after pulse scale adaptation to minimize interference. Thus,  $L = 1$  and hence only adjacent time slots interfere. The residual ICI signal can be calculated by

$$\mathbf{y}_l^{\text{ICI}} = (\mathbf{C}_{l,l} - \mathbf{\Lambda}_{l,l}) \mathbf{x}_l \quad (45)$$

where  $\mathbf{\Lambda}_{l,l}$  is a diagonal matrix whose diagonal elements are the same as those of  $\mathbf{C}_{l,l}$ .

*Algorithm 1 (SIIC):* The SIIC algorithm with  $I$  iterations proceeds as follows.

- 1) Initiate iteration index  $i = 1$ . For time slot  $l$ , perform ISI cancellation by calculating the ISI using decoded bits from previous  $L$  time slots and then subtracting it from  $\mathbf{y}_l$

$$\mathbf{y}_l^{(1)} = \mathbf{y}_l - \sum_{d=1}^L \mathbf{C}_{l,l-d} \hat{\mathbf{x}}_{l-d}$$

where  $\hat{\mathbf{x}}_{l-d}$  are previous bit decisions and  $\mathbf{y}_l^{(1)}$  is the ISI-removed signal.

- 2) Decode  $\mathbf{y}_l^{(i)}$  assuming no interference. This generates the  $i$ th bit decisions  $\hat{\mathbf{x}}_l^{(i)}$ .
- 3) Reconstruct the  $i$ th residual ICI signal  $\mathbf{y}_l^{\text{ICI},i}$  by using (45) and subtracting it from  $\mathbf{y}_l^{(i)}$  to generate the signal  $\mathbf{y}_l^{(i+1)}$  for the next iteration

$$\mathbf{y}_l^{(i+1)} = \mathbf{y}_l^{(i)} - (\mathbf{C}_{l,l} - \mathbf{\Lambda}_{l,l}) \hat{\mathbf{x}}_l^{(i)}$$

- 4) Iterate Steps 2 and 3 for  $I$  times.

*Remark 6:* Many practical slow-fading channels exhibit relatively small Doppler dispersion, so the ICI effect is small for these systems. The optimal pulse scaling for this scenario coincides with OFDM signaling, where a cyclic prefix is used to combat ISI due to multipath delay. In this case, the SIIC can be avoided because of the negligible residual interference in the system. In fact, SIIC is intended for fast-fading channels with large multipath dispersion.

*Remark 7:* The exact performance analysis of the SIIC algorithm seems difficult. But approximation can be done by tracking each step of the algorithm and approximating the

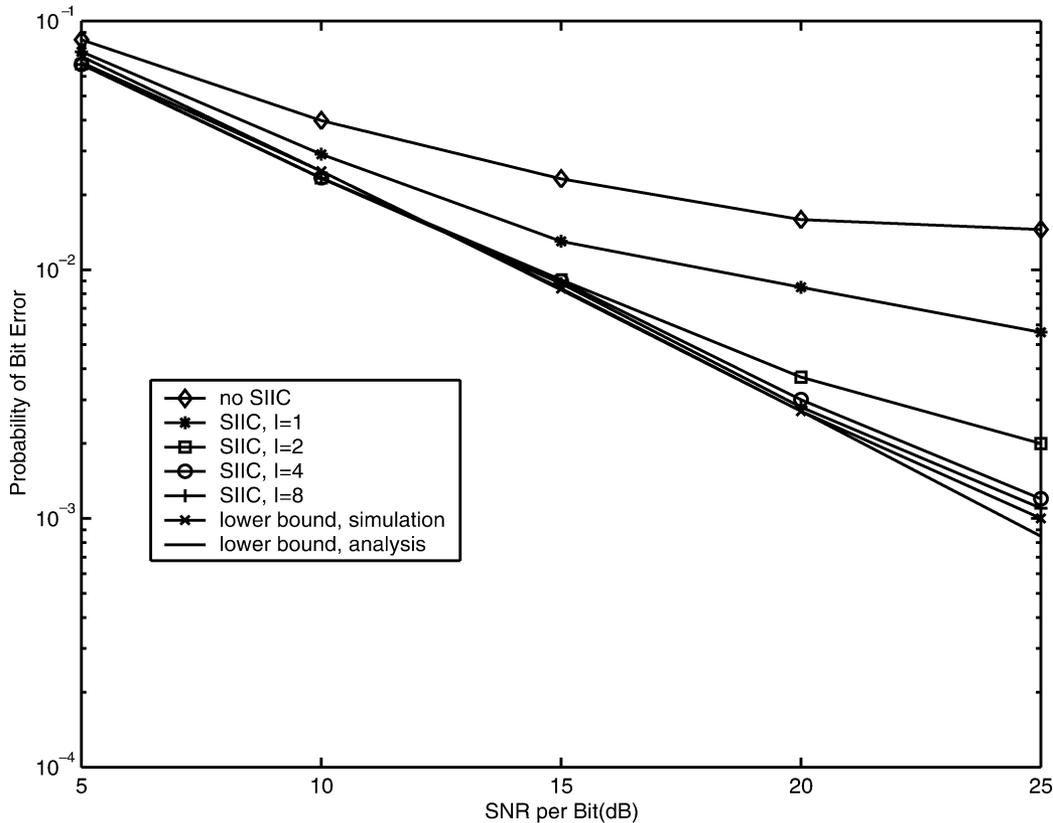


Fig. 8. SIIC performance for spread factor 0.01. No diversity signaling is used. The lower bound for the bite-error rate (BER) corresponds to perfect interference cancellation.

residual interference as Gaussian noise. We refer to [24] for such an analysis technique.

Fig. 8 shows the performance of SIIC-based reception of the BPSK modulation with no diversity over a channel with a spread factor of 0.01. The channel simulation setup is the same as in Fig. 7. We also plot the SIIC performance if perfect decisions are supplied, which serves as a lower bound for  $P_b$ . As evident from the figure, the SIIC algorithm demonstrates an impressive performance enhancement. A large performance gain can be achieved by only a small number of iterations. For example, the performance for  $I = 2$  (ISI removal and one ICI removal) is already very close to the lower bound for SNR ranging from 5 to 25 dB. However, the performance gain diminishes as the number of iteration increases, which is typical behavior of decision feedback techniques.

Shown in Fig. 9 is the SIIC performance for essentially the same BPSK simulation setup but with two-level of diversity. Comparing it with Fig. 8, we see that SIIC benefits also from diversity due to the enhancement in decision feedback quality, which is critical to overall performance. For instance, two iterations of SIIC are sufficient to attain the performance lower bound as shown in the figure.

## V. SYSTEM CAPACITY

In this section, we study doubly dispersive channels and our proposed signaling framework from an information-theoretic

viewpoint. We focus on the impact of channel parameters, especially  $T_m B_d$ , on system capacity.

### A. A Moment Theorem for Doubly Dispersive Channels

We shall work with a discrete description of doubly dispersive channels for capacity analysis. In this context, the channel output  $r_n$  and input  $s_n$  are related by

$$r_n = \sum_{l=0}^L h_l(n) s_{n-l} + w_n \quad (46)$$

where  $L + 1$  is the total number of resolvable paths and noise  $w_n \sim \text{CN}(0, \sigma^2)$  is white in time. The path coefficient process  $\{h_l(n)\}$  is the discrete analog of the continuous process  $\hat{h}(\tau, t)$  in (1) with the path index  $l$  corresponding to  $\tau$  and the time index  $n$  corresponding to  $t$ . In view of the WSSUS assumption on doubly dispersive channels, we may assume that different path processes are independent zero-mean complex Gaussian and that sum of total variance is 1, that is,  $\sum_{l=0}^L \sigma_l^2 = 1$  where  $\sigma_l^2$  is the variance of the  $l$ th wide-sense stationary (WSS) process  $\{h_l(n)\}$ .

The channel equation in (46) can be concisely rewritten in a multiple-input multiple-output (MIMO) form as

$$\mathbf{r} = \mathbf{H}\mathbf{s} + \mathbf{w} \quad (47)$$

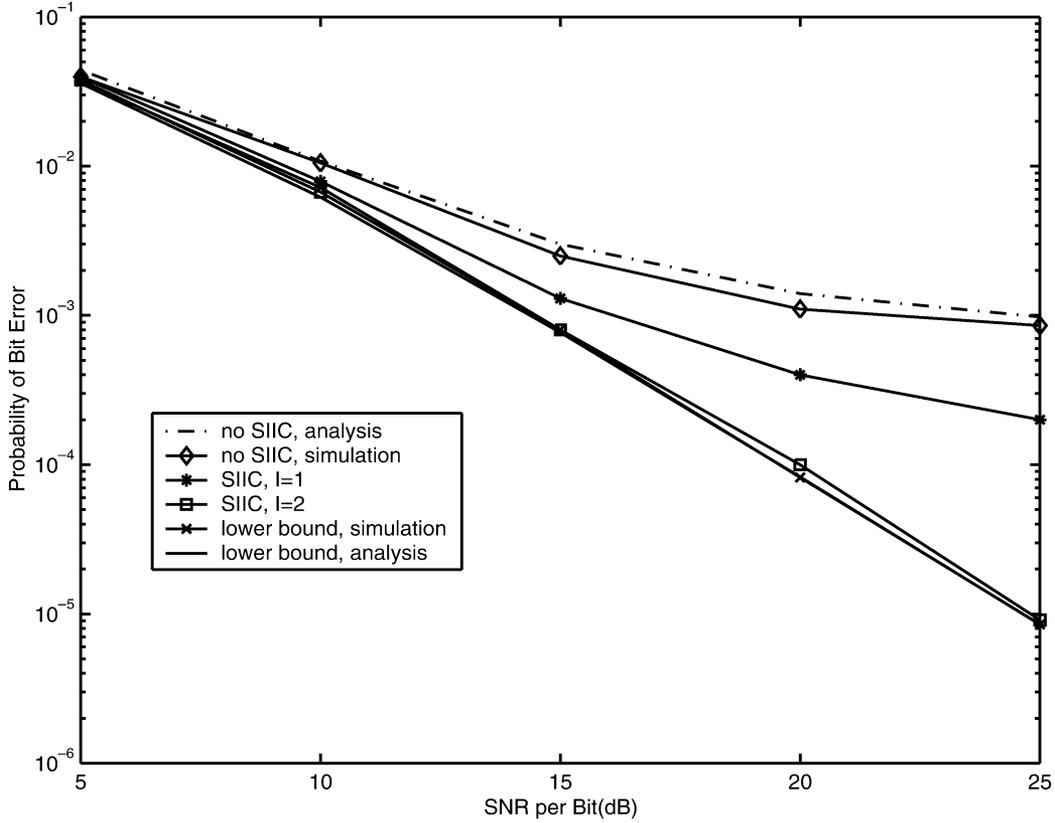


Fig. 9. SIIC performance with diversity signaling of order 2.

where the semi-infinite channel matrix is given by

$$\mathbf{H} = \begin{pmatrix} h_0(1) & & & & \\ h_1(2) & h_0(2) & & & \\ \vdots & \vdots & \ddots & & \\ h_L(L+1) & \vdots & h_1(L+1) & h_0(L+1) & \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (48)$$

The matrix  $\mathbf{H}$  has only an  $L + 1$  number of diagonal lines, each corresponding to a path coefficient process. Unlike time-invariant linear channels, doubly dispersive channels cause variations along the diagonals of  $\mathbf{H}$ , whose statistical characteristics are strongly connected to channel Doppler spread  $B_d$ , or equivalently, to channel coherence time  $T_{\text{coh}}$ . The larger  $B_d$ , the larger the variation. Since the discrete channel essentially comes from discretizing the underlining (continuous) doubly dispersive channel, one can similarly define the notion of channel coherence length  $K = \lceil T_{\text{coh}}B \rceil$  where the signaling bandwidth  $B$  is assumed to be sufficiently large. Therefore, diagonal entries within a range of  $K$  remain strongly correlated. Since  $L = \lceil T_m B \rceil$ , one has

$$\frac{L}{K} = \frac{\lceil T_m B \rceil}{\lceil T_{\text{coh}} B \rceil} \approx T_m B_d. \quad (49)$$

Hence, the channel spread factor  $T_m B_d$  can also be seen as a measure of relative ratio between the number of diagonal lines of  $\mathbf{H}$  and the ‘‘constant’’ length of those lines.

When channel state information is available at the receiver but not at the transmitter, the ergodic channel capacity per dimension given power constraint  $P$  is given by [25], [26]

$$C = \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}[\log \det(\mathbf{I}_N + P\mathbf{B}_N)] \quad (50)$$

where  $\mathbf{B}_N = \mathbf{H}_N \mathbf{H}_N^H$  and the subscript  $N$  means a truncation of the corresponding infinite-dimensional matrix to dimension  $N$ . Denote by  $F(\lambda)$  the limiting empirical distribution of eigenvalues  $\{\lambda_{i,N}\}$  of random matrix  $\mathbf{B}_N$

$$F(\lambda) = \lim_{N \rightarrow \infty} F_N(\lambda) = \lim_{N \rightarrow \infty} \frac{1}{N} \#(\lambda_{i,N} \leq \lambda) \quad (51)$$

where  $\#(\mathcal{A})$  denotes the cardinality of a set  $\mathcal{A}$ . As shown in [26], the capacity  $C$  in (50) can also be expressed as

$$C = \int \log(1 + P\lambda) dF(\lambda). \quad (52)$$

It is easy to observe that the random matrix  $\mathbf{B}$  also exhibits a finite banded structure as  $\mathbf{H}$ . The eigenvalue distribution of such banded random matrices is therefore a key to evaluating channel capacity. We refer readers to a recent excellent review on spectral properties of large random matrices [27], where several types of random matrices have been analyzed. But the problem for banded random matrices seems elusive and remains largely unsolved [28]. Although at this stage we are not able to give the limiting eigenvalue distribution of  $\mathbf{B}_N$ , we have characterized its moment behavior.

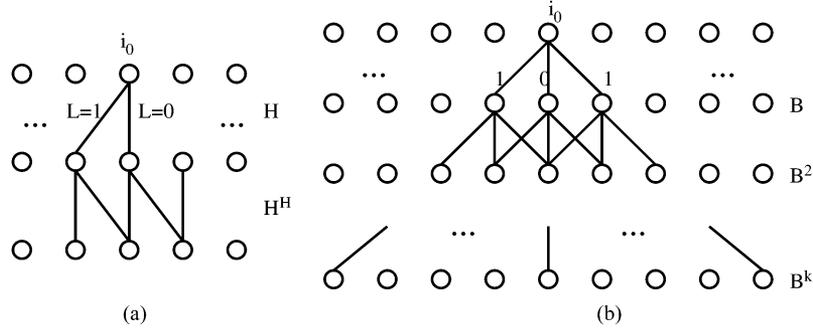


Fig. 10. A schematic illustrating matrix products using MBG representation: (a) corresponds to  $\mathbf{H}\mathbf{H}^H$  and (b) corresponds to  $\mathbf{B}^k$  where  $L = 1$  and local branches relative to the  $i_0$ th row are shown.

The  $k$ th moment  $M_{k,N}$  of the empirical distribution  $F_N(\lambda)$  is

$$M_{k,N} = \int \lambda^k dF_N(\lambda) = \frac{1}{N} \sum_{i=1}^N \lambda_{i,N} = \frac{1}{N} \text{tr}(\mathbf{B}_N^k) \quad (53)$$

where  $\text{tr}(\cdot)$  denotes the trace of a matrix. As  $N \rightarrow \infty$

$$\begin{aligned} M_{k,N} &\rightarrow M_k = \int \lambda^k dF(\lambda) \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N b_{i,i,k,N} \\ &= \mathbb{E}[b_{i_0,i_0,k}] \quad \text{a.s.} \end{aligned} \quad (54)$$

where  $b_{i,j,k,N}$  is the  $(i,j)$ th entry in  $\mathbf{B}_N^k$  and  $b_{i_0,i_0,k}$  represents a typical entry in  $\mathbf{B}^k$  [26].<sup>4</sup> Therefore, the problem boils down to evaluating entries in higher matrix products.

For this purpose, we will briefly touch upon the usage of matrix bipartite graph (MBG). The MBG for a matrix  $\mathbf{A}$  is defined to be a bipartite graph consisting of two rows of vertices and downward arcs between them. Vertices in the top row correspond to row indices of  $\mathbf{A}$  and vertices in the bottom row to column indices of  $\mathbf{A}$ . The arc connecting the top vertex  $vt_i$  to the bottom vertex  $vb_j$  has value  $a_{ij}$ , the  $(i,j)$ th entry in  $\mathbf{A}$ . Let the MBG associated with matrix  $\mathbf{A}$  be  $\mathcal{G}_1 = (vt_i, vb_k, a_{ik})$  and the MBG with  $\mathbf{B}$  be  $\mathcal{G}_2 = (vt_k, vb_j, b_{kj})$ . Then, the MBG  $\mathcal{G}$  corresponding to  $\mathbf{C} = \mathbf{A}\mathbf{B}$  can be simply represented by cascading  $\mathcal{G}_1$  and  $\mathcal{G}_2$ , and the arc connecting  $vt_i$  to  $vb_j$  is the sum of all directed paths from  $vt_i$  of  $\mathcal{G}_1$  to  $vb_j$  of  $\mathcal{G}_2$ , or symbolically,  $\mathcal{G} = (vt_i, vb_j, \sum_k a_{ik}b_{kj})$ . Shown in Fig. 10 are the MBG representations of  $\mathbf{H}\mathbf{H}^H$  and  $\mathbf{B}^k$  whose  $i_0$  row is represented by a subgraph fanning out from the common top  $i_0$ th vertex. It is straightforward to observe that the finite bandedness of  $\mathbf{H}$  implies the finite bandedness of  $\mathbf{B}^k$ , which has  $2kL + 1$  number of diagonals ranging from  $-kL$  to  $kL$ . To facilitate exposition, we introduce a reindexing of matrix entries for banded matrix  $\mathbf{B}$  as

$$\tilde{b}_{i,l} = b_{i,i-l} \quad (55)$$

which denotes the entry at the intersection of the  $i$ th row and the  $l$ th diagonal line of  $\mathbf{B}$ .

Fig. 10 clearly reveals the local nature of computation of entries in  $\mathbf{B}^k$ . More specifically, the  $i_0$ th row of  $\mathbf{B}^k$  is only affected by rows of  $\mathbf{B}$  ranging from  $i_0 - kL$  to  $i_0 + kL$ , or equivalently,

<sup>4</sup> $i_0$  is sufficiently large for fixed  $k$  to avoid the edge effects.

by rows of  $\mathbf{H}$  ranging from  $i_0 - (k+1)L$  to  $i_0 + (k+1)L$ , denoted by  $\mathcal{R}(i_0, k)$ .

Now suppose that the channel coherence length  $K$  is sufficiently large compared to  $2(k+1)L + 1$  so that each diagonal line is almost constant within the row range  $\mathcal{R}(i_0, k)$ . One has

$$\tilde{b}_{i_0,l,1} = \sum_{m=0}^L \tilde{h}_{i_0,m} \tilde{h}_{i_0-l,m-l}^* \approx \sum_{m=0}^L \tilde{h}_{i_0,m} \tilde{h}_{i_0,m-l}^* = a(l) \quad (56)$$

where  $a(l) = \tilde{h}_{i_0} \circ \tilde{h}_{i_0}^*(l)$  denotes the correlation operation. Similarly, the  $i_0$ th row of  $\mathbf{B}^k$  is given by  $k$  iterated correlation of  $a(l)$ , that is,  $a_k(l) = a \circ \dots \circ a(l)$ ,  $k$  times. Let  $\hat{h}(\omega) = \sum_l \tilde{h}_{i_0,l} e^{-jl\omega}$  be the (discrete-time) Fourier transform of the sequence  $\tilde{h}_{i_0,l}$  with respect to the diagonal line index  $l$ . Since  $\tilde{h}_{i_0,l}$  are independent complex Gaussian with total variance normalized to 1, it is easy to check that  $\hat{h}(\omega) \sim \text{CN}(0, 1)$  is white in frequency  $\omega$ . Since the Fourier transform of  $a_k(l)$  is simply related to that of  $\tilde{h}_{i_0,l}$  by

$$\hat{a}_k(\omega) = |\hat{h}(\omega)|^{2k} \quad (57)$$

one has

$$M_k = \mathbb{E}[\tilde{b}_{i_0,0,k}] = \mathbb{E}[a_k(0)] = \mathbb{E}\left[\frac{1}{2\pi} \int_0^{2\pi} |\hat{h}(\omega)|^{2k} d\omega\right] = k! \quad (58)$$

that is, the eigenvalue distribution is exponential. In this case, the capacity of doubly dispersive channels degenerates to that of flat fading channels.

However, channel variation always exists for general doubly dispersive channels where the spread factor  $T_m B_d$  is nonzero but may be small in most practical situations. Suppose that variance of such variation in the local range  $\mathcal{R}(i_0, k)$  is uniformly upper-bounded by  $\delta$ , that is,

$$\sup_{i,j \in \mathcal{R}(i_0,k)} \mathbb{E}|\tilde{h}_{i,l} - \tilde{h}_{j,l}|^2 \leq \delta. \quad (59)$$

Since  $\mathbb{E}|\tilde{h}_{i,l} - \tilde{h}_{j,l}|^2 \leq \mathbb{E}(|\tilde{h}_{i,l}|^2 + |\tilde{h}_{j,l}|^2) \leq 4$ , one can choose  $\delta \leq 4$ .

*Theorem 2:* Given an integer  $k \geq 1$ , suppose the channel variance within the range  $\mathcal{R}(i_0, k)$  is upper-bounded by  $\delta$ . Then, the  $k$ th moment of limiting eigenvalue distribution satisfies

$$|M_k - k!| \leq (2L + 1)^{2k-1} C_k \delta^{1/2} \quad (60)$$

where the constant  $C_k$  depends on  $k$ .

*Proof:* See Appendix III.  $\square$

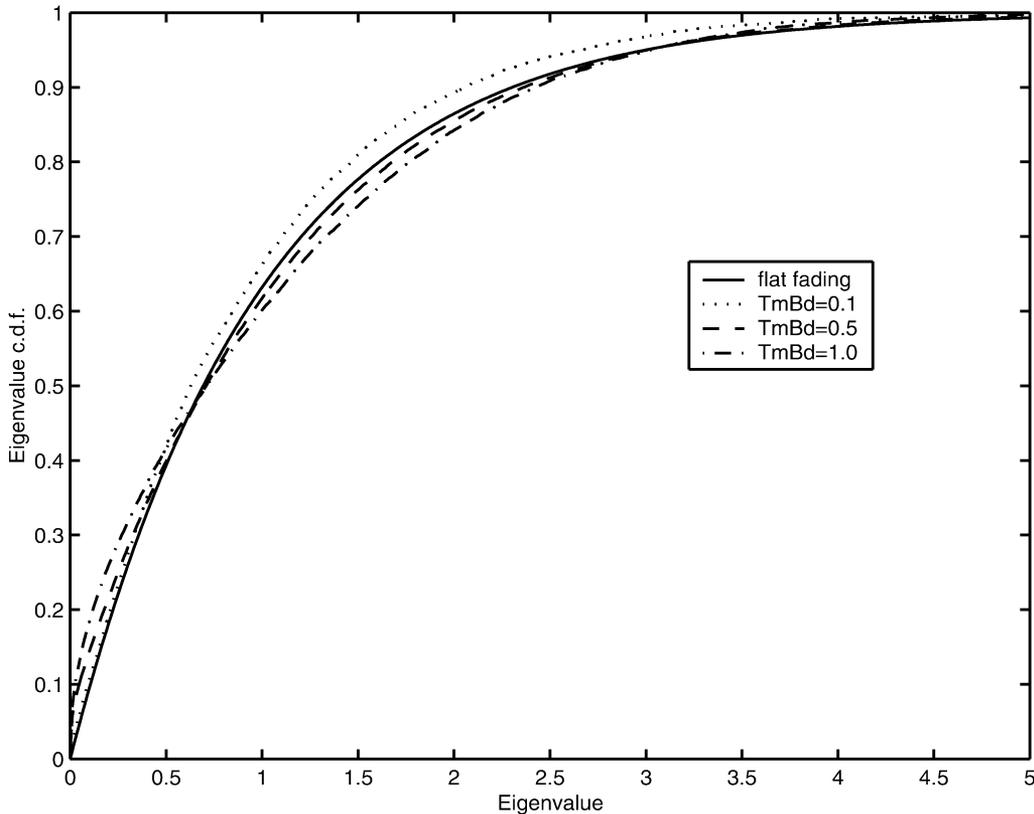


Fig. 11. Eigenvalue distribution of a random channel matrix corresponding to  $L + 1 = 4$  paths and  $N = 100\,000$  symbols.

Theorem 2 characterizes the effect of channel spread parameters on capacity. Since the time span in  $\mathcal{R}(i_0, k)$  is  $2(k+1)L+1$ , the channel variation is proportional to the relative ratio of the time span and the channel coherence time  $K$  and hence,

$$\delta \propto \frac{2(k+1)L+1}{K} \propto kT_m B_d \quad (61)$$

by (49). The smaller the channel spread  $T_m B_d$ , the smaller the  $\delta$  and thus the moment difference up to the  $k$ th order. Higher order moments tend to exhibit more deviation than lower ones. However, more and more moments of the eigenvalue distribution of doubly dispersive channels will agree with those of flat-fading channels provided that  $T_m B_d$  is small enough, which implies that capacity of doubly dispersive channels converges to that of flat-fading channels as channel spread decreases.

Fig. 11 plots the empirical eigenvalue distribution of simulated doubly dispersive channels for various spread factors. The distribution is seen to deviate from the exponential distribution for large  $T_m B_d$ . But the deviation is small as long as channel remains underspread. Shown in Fig. 12 is the channel capacity computed from the empirical eigenvalue distribution. As evident from the figure, capacity of underspread channels is upper-bounded by that of flat-fading channels and the difference between them is fairly small for small SNR values.

### B. Block Fading in the Time-Frequency Domain

The notion of coherence subspace in our signaling framework leads to a time-frequency *block-fading* view for doubly

dispersive channels. As illustrated in Fig. 6, channel coefficients within the same coherence subspace are assumed to be the same while they vary in an independent fashion from one subspace to another. The operational capacity of STF signaling is affected both by channel dispersion parameters and by the performance of interference cancellation technique such as SIIC in our framework. Here we treat interference as Gaussian noise to simplify the analysis.

The channel ergodic capacity per dimension is given by

$$I = \mathbb{E}[\log(1 + \lambda \text{SINR})] \quad (62)$$

where  $\lambda$  is exponentially distributed with unit mean. The SINR in the framework depends on many factors such as channel parameters, pulse characteristics, and the interference cancellation technique used. It is lower bounded by the minimum SINR without interference cancellation and upper-bounded by the maximum SINR with perfect interference cancellation, that is,

$$\frac{\eta E_s}{\sigma^2 + (1 - \eta)E_s} \leq \text{SINR} \leq \frac{\eta E_s}{\sigma^2}. \quad (63)$$

If the interference is left untreated, the system capacity will suffer from a limiting cap analogous to the error floor in the BER performance. More precisely, as  $E_s \rightarrow \infty$ ,  $\text{SINR} \rightarrow \frac{\eta}{1-\eta}$  and hence  $I \rightarrow \mathbb{E}[\log(1 + \lambda \frac{\eta}{1-\eta})]$ .

Fig. 13 plots the ergodic capacity of the STF signaling framework for a scale-adapted rectangular pulse over a channel with spread factor 0.01. The lower and upper bounds for SINR were estimated from the simulation data. The capacity of flat-fading channel is also plotted as a benchmark. As evident

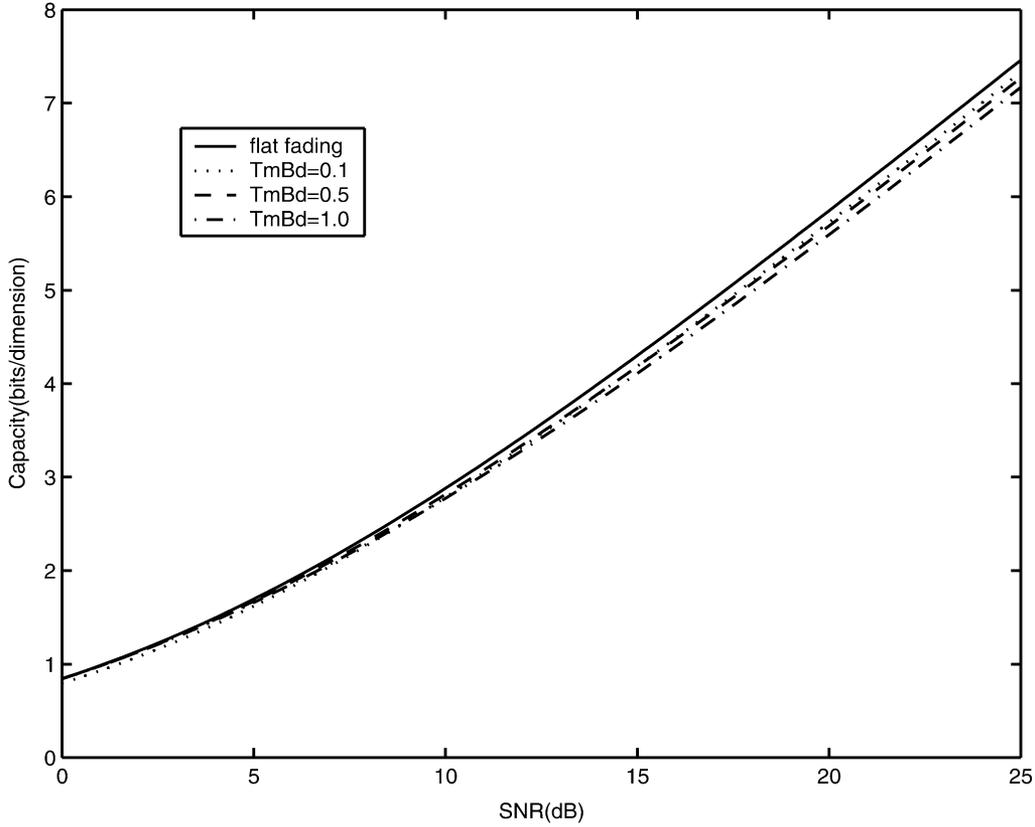


Fig. 12. Ergodic capacity for different spread factors calculated from the estimated eigenvalue distribution.

from the figure, system capacity levels off in the high-SNR region when no interference cancellation is used. In contrast, the SIIC effectively remedies the capacity bottleneck. In the figure, the mutual information curves for different iterations of the SIIC are plotted using the estimated residual SINR from simulation data. More iterations in SIIC narrow the capacity gap to the upper bound. Furthermore, system capacity benefits from small channel spread factors, which is demonstrated by Fig. 14, where system capacity corresponding to a spread factor 0.001 is plotted. Fig. 14 suggests that interference cancellation techniques may be avoided without incurring significant capacity loss when the channel spread factor is fairly small. For example, the lower bound in Fig. 14 almost coincides with the flat-fading capacity up to an SNR of 10 dB.

It is instructive to point out the difference between undercritical sampling ( $T_0F_0 > 1$ ) and critical sampling ( $T_0F_0 = 1$ ) with respect to system capacity. Undercritical sampling offers better time–frequency localization than the critical sampling orthogonal basis in our STF framework. But its spectral efficiency is reduced by a factor of  $T_0F_0 > 1$ . More specifically, the system capacity per dimension of an undercritical sampling basis is given by

$$I' = \frac{1}{T_0F_0} \mathbb{E} [\log(1 + d\lambda \text{SINR})] \quad (64)$$

where  $d$  is the SINR improvement over the critical sampling basis. Comparing it with the capacity of the critical sampling basis (62), we want to find out how large  $d$  needs to be in order to compensate for the loss in spectral efficiency such that it still

would attain the same capacity as that of (62). Using Jensen's inequality, one has

$$\begin{aligned} \mathbb{E}_\lambda \log(1 + \lambda \text{SINR}) &= \frac{1}{T_0F_0} \mathbb{E}_\lambda \log(1 + d\lambda \text{SINR}) \\ &\leq \frac{1}{T_0F_0} \log(1 + d\text{SINR}) \end{aligned} \quad (65)$$

which, together with  $T_0F_0 > 1$ , implies that

$$\begin{aligned} d &\geq \frac{1}{\text{SINR}} (e^{T_0F_0 \mathbb{E}_\lambda \log(1 + \lambda \text{SINR})} - 1) \\ &\geq \frac{1}{\text{SINR}} (e^{T_0F_0 \mathbb{E}_\lambda \log \lambda} + \text{SINR}^{T_0F_0}) \rightarrow \infty \\ &\quad \text{as SINR} \rightarrow \infty. \end{aligned} \quad (66)$$

The above computation illustrates the merit of critical sampling STF basis: it attains the largest spectral efficiency. The slight improvement on SINR by using undercritical sampling may not be warranted from capacity perspective.

Our framework also facilitates the evaluation of the *outage capacity* [29]. Given a total signal duration  $T$  and bandwidth  $B$ , the level of multipath-Doppler diversity is  $D = \lceil T_m B_d T B \rceil$  and the number of basis function in a coherence subspace is  $N_c = \lceil T_{\text{coh}} B_{\text{coh}} \rceil$  as discussed in Section IV. So, the maximal mutual information per dimension can be written as

$$I(\mathbf{\Lambda}) = \frac{1}{D} \sum_{d=1}^D \log(1 + \text{SINR} \lambda_d) \quad (67)$$

where  $\mathbf{\Lambda} = (\lambda_1, \dots, \lambda_D)$  is a vector of the fading coefficient for each coherence subspace. Then, the outage capacity is defined

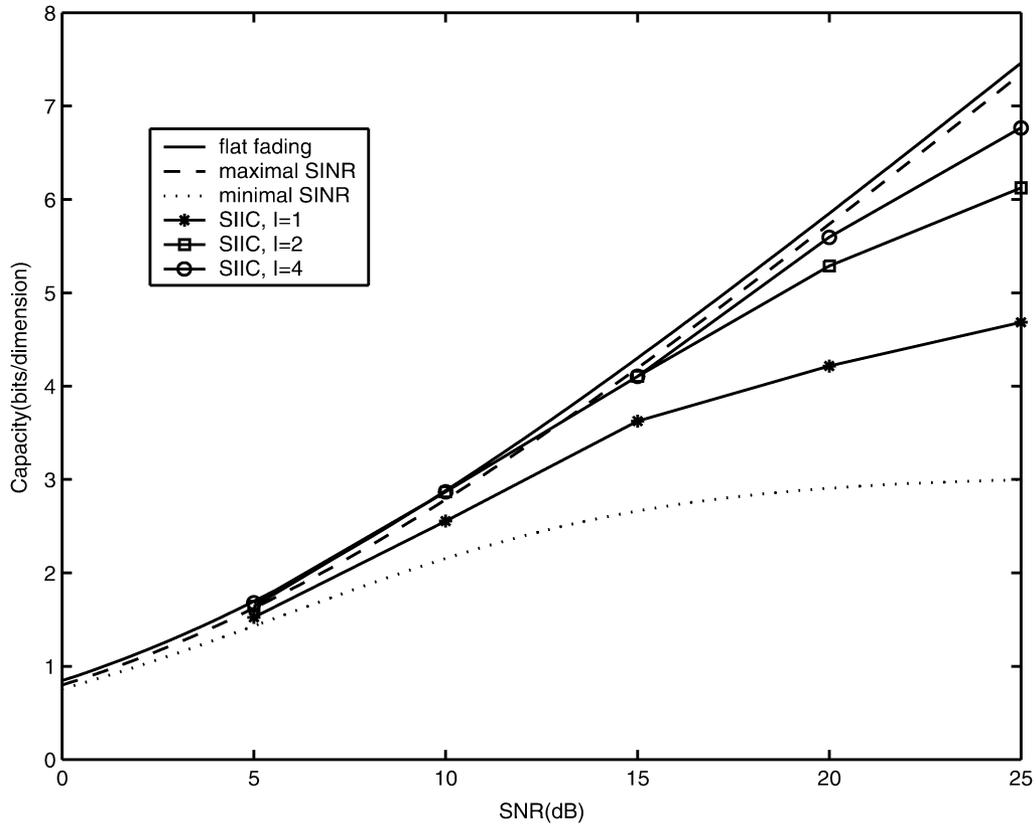


Fig. 13. Ergodic capacity of STF signaling for spread factor 0.01. The capacity depends on the residual SINR. Capacity when SIIC is used is calculated from the estimated residual SINR from the SIIC simulations.

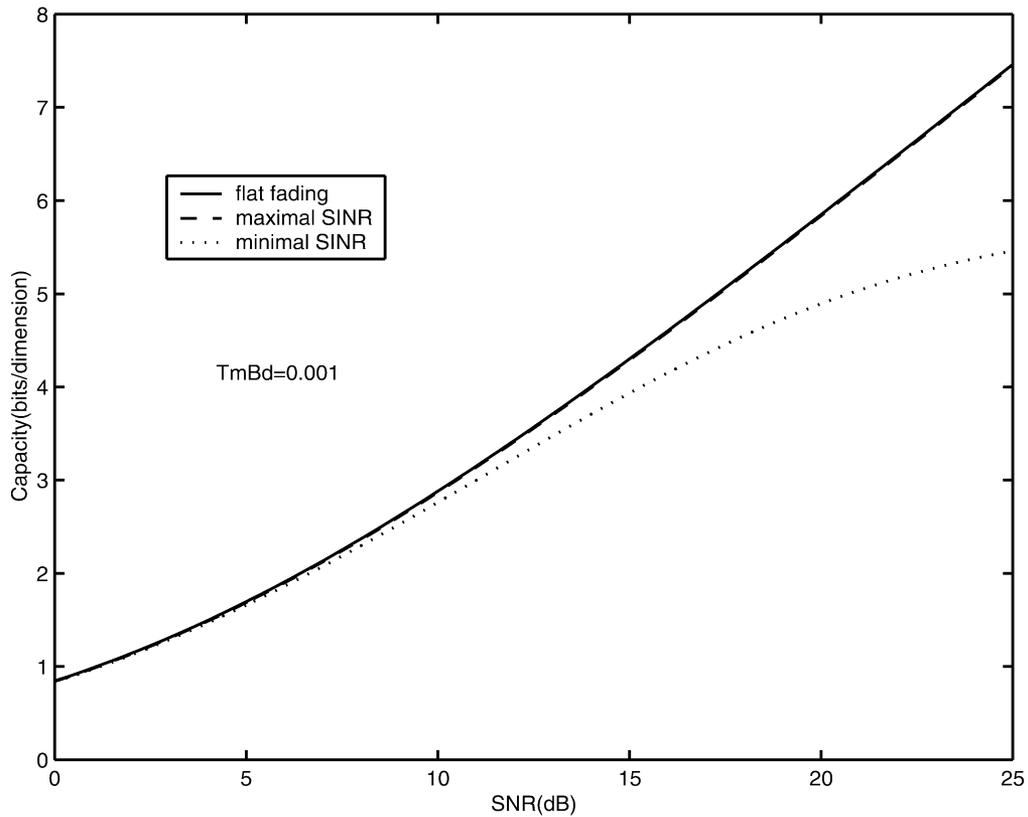


Fig. 14. Ergodic capacity of signaling framework for spread factor 0.001.

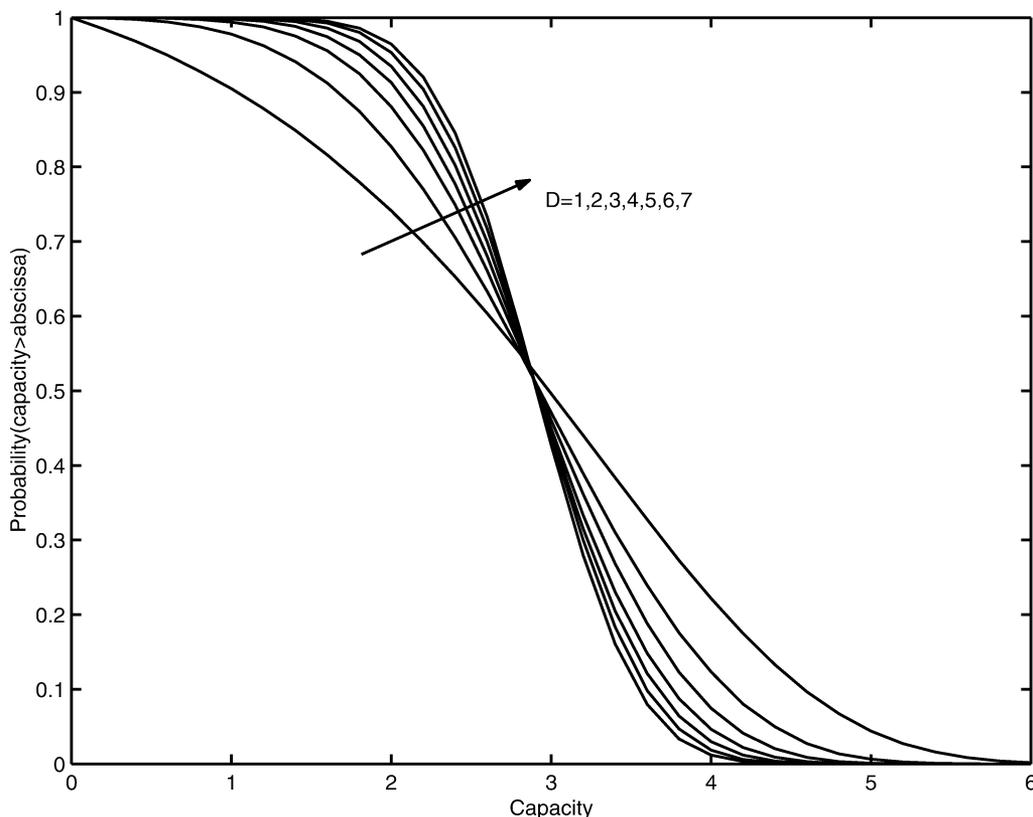


Fig. 15. Outage capacity curves for different levels of diversity at SNR = 10 dB.

to be the maximum rate that can be guaranteed with a certain outage probability  $P_{\text{out}}$ . More precisely

$$P_{\text{out}}(c) = \Pr(I(\Lambda) \leq c). \quad (68)$$

The outage performance curves for different levels of diversity are plotted in Fig. 15. Large channel spread factors give rise to more diversity in the system. Diversity improves outage performance although such improvement diminishes at high diversity level.

## VI. CONCLUSION

We have proposed a general framework for orthogonal signaling over doubly dispersive fading channels via STF basis functions. Our quantitative results reveal the effect of pulse time–frequency properties and channel characteristics on signaling performance. A simple scale adaptation rule is derived, also proposed in [3], [4], to match the pulse characteristics with the channel. For sufficiently small channel spread factors, a scale-adapted STF basis serves as a set of approximate eigenfunctions of the channel and yields a simple block-fading interpretation of the channel in terms of time–frequency coherence subspaces. We propose an efficient interference cancellation technique to further reduce the residual interference. System capacity of doubly dispersive channels is studied using random matrix theory, which reveals the important role of channel spread factor on capacity. In particular, the capacity deviation of doubly dispersive channels from flat-fading channels is proportional to channel spread factor.

In relation to the work on undercritically sampled biorthogonal systems ( $T_0 F_0 >$ ) [3], [5]–[7], the results in this paper demonstrate the attractiveness of using orthogonal STF bases for communication over doubly dispersive channels, despite their poor time–frequency localization properties. This is because undercritical systems result in a loss in spectral efficiency that affects overall system capacity linearly, whereas the improvement in SINR due to better time–frequency localization only yields a logarithmic gain in capacity. Furthermore, the simplicity and impressive performance of the proposed SIIC scheme suggests that orthogonal systems with interference cancellation could yield SINR performance comparable to biorthogonal systems. This is also significant from a practical perspective since the critically sampled case is much easier to implement in practice. The same set of basis functions are used at the transmitter and receiver and incorporating our results in existing systems requires a simple modification: using existing prototype pulses with appropriate scale adaptation that only requires knowledge of channel delay and Doppler spreads. In biorthogonal systems, two (dual) basis functions are needed at the transmitter and receiver and the determination of dual prototypes requires numerical optimization, which is always approximate and changes with channel statistics [3].

Finally, we note that other recent results on the noncoherent capacity of time-varying multipath channels suggest signaling waveforms that are peaky in time or frequency [30], [31] as opposed to noise-like spread-spectrum waveforms. The notion of

time–frequency coherence subspaces introduced in this paper suggests that partial spreading in time–frequency, commensurate with the dimension of the coherence subspaces, warrants further investigation in a noncoherent context. Furthermore, the block-fading interpretation in terms of time–frequency coherence subspaces facilitates the application of coding techniques for block-fading channels for communication over underspread doubly dispersive channels [32], [33].

### APPENDIX I PROOF OF PROPOSITION 2

From (7), we have

$$c_{lm,l'm'} = \int \int h(\tau, \nu) \cdot \left( \int \phi_{lm}(t) \phi_{l'm'}^*(t - \tau) e^{-j2\pi\nu t} dt \right)^* d\tau d\nu. \quad (69)$$

For the STF basis, we calculate the inner integral in (69)

$$\begin{aligned} I &= \int g(t - lT_0) e^{j2\pi m F_0 t} g^*(t - \tau - l'T_0) \\ &\quad \cdot e^{-j2\pi m' F_0 (t - \tau)} e^{-j2\pi\nu t} dt \\ &\stackrel{t - lT_0 = t'}{=} e^{j2\pi T_0 F_0 l(m - m')} e^{-j2\pi l T_0 \nu} e^{j2\pi m' F_0 \tau} \\ &\quad \cdot \int g(t') g^*(t' - ((l' - l)T_0 + \tau)) \\ &\quad \cdot e^{-j2\pi((m' - m)F_0 + \nu)t'} dt' \\ &\stackrel{T_0 F_0 = 1}{=} A_g((l' - l)T_0 + \tau, (m' - m)F_0 + \nu) e^{j2\pi m' F_0 \tau} \\ &\quad \cdot e^{-j2\pi l T_0 \nu}. \end{aligned}$$

Substituting this into (69) yields (17). Next we evaluate  $\mathbb{E}|c_{lm,l'm'}|^2$  to yield (18):

$$\begin{aligned} &\mathbb{E}|c_{lm,l'm'}|^2 \\ &= \mathbb{E} \int \int h(\tau, \nu) A_g^*((l' - l)T_0 + \tau, (m' - m)F_0 + \nu) \\ &\quad \cdot e^{-j2\pi m' F_0 \tau} e^{j2\pi l T_0 \nu} d\tau d\nu \\ &\quad \cdot \int \int h^*(\tau_1, \nu_1) \\ &\quad \cdot A_g((l' - l)T_0 + \tau_1, (m' - m)F_0 + \nu_1) \\ &\quad \cdot e^{j2\pi m' F_0 \tau_1} e^{-j2\pi l T_0 \nu_1} d\tau_1 d\nu_1 \\ &= \int \int \int \int S(\tau, \nu) \delta(\tau - \tau_1) \delta(\nu - \nu_1) \\ &\quad \cdot A_g^*((l' - l)T_0 + \tau, (m' - m)F_0 + \nu) e^{-j2\pi m' F_0 \tau} \\ &\quad \cdot e^{j2\pi l T_0 \nu} A_g((l' - l)T_0 + \tau_1, (m' - m)F_0 + \nu_1) \\ &\quad \cdot e^{j2\pi m' F_0 \tau_1} e^{-j2\pi l T_0 \nu_1} d\tau d\nu d\tau_1 d\nu_1 \\ &= \int \int S(\tau, \nu) \\ &\quad \cdot |A_g((l' - l)T_0 + \tau, (m' - m)F_0 + \nu)|^2 d\tau d\nu. \end{aligned}$$

### APPENDIX II PROOF OF THEOREM 1

It is easy to see from the definition of ambiguity function that  $\psi(0) = 1$  and  $\psi(x) = \psi(-x)$ . Then, (19) becomes

$$\begin{aligned} \eta &= \frac{2}{T_m B_d} \int_0^{T_m} \int_0^{B_d/2} \psi\left(\frac{\tau}{\sigma_t}\right) \psi\left(\frac{\nu}{\sigma_f}\right) d\tau d\nu \\ &= \frac{2\kappa}{T_m B_d} \int_0^a \psi(x) dx \int_0^b \psi(x) dx \end{aligned} \quad (70)$$

where  $a = \frac{T_m}{\sigma_t}$  and  $b = \frac{B_d/2}{\sigma_f}$ . Since  $ab = \frac{T_m B_d}{2\kappa} = c > 0$  is a constant, we rewrite (70) as

$$\eta(a) = \frac{1}{c} \int_0^a \psi(x) dx \int_0^{\frac{c}{a}} \psi(x) dx. \quad (71)$$

Taking the derivative with respect to  $a$ , we have

$$\eta'(a) = \frac{1}{c} \left[ \psi(a) \int_0^{\frac{c}{a}} \psi(x) dx - \frac{c}{a^2} \psi\left(\frac{c}{a}\right) \int_0^a \psi(x) dx \right] \quad (72)$$

from which  $\bar{a} = \frac{c}{\bar{a}} = \bar{b}$  is seen to be a stationary point ( $\eta'(\bar{a}) = 0$ ), thus proving (29). The second derivative at this point is given by

$$\begin{aligned} \eta''(\bar{a}) &= \frac{2}{c} \left[ \psi'(\bar{a}) \int_0^{\bar{a}} \psi(x) dx + \frac{\psi(\bar{a})}{\bar{a}} \int_0^{\bar{a}} \psi(x) dx - \psi^2(\bar{a}) \right] \\ &= \frac{2}{c} f(\bar{a}) \int_0^{\bar{a}} \psi(x) dx \end{aligned} \quad (73)$$

where  $f(x)$  is defined in (28). So,  $f(\bar{a}) < 0$  will guarantee  $\bar{a}$  to be a local maximum. Since  $\psi(x) \leq 1$ , one has  $\int_0^{\bar{a}} \psi(x) dx \leq \bar{a}$ . If condition (30) is satisfied, then

$$\begin{aligned} f(\bar{a}) &\leq \psi'(\bar{a}) + \frac{\psi(\bar{a})}{\bar{a}} - \frac{\psi^2(\bar{a})}{\bar{a}} \\ &= \psi'(\bar{a}) - \frac{\psi(\bar{a})}{\bar{a}} (\psi(\bar{a}) - 1) < 0. \end{aligned} \quad (74)$$

### APPENDIX III PROOF OF THEOREM 2

We bound the perturbation along the computation of entries in  $\mathbf{B}^k$ . Let  $\tilde{\varepsilon}_{i,l,k} = \tilde{b}_{i,l,k} - a_k(l)$  be the (random) difference between the true value  $\tilde{b}_{i,l,k}$  and the desired target  $a_k(l)$ . For convenience,  $k = 0$  corresponds to  $\mathbf{H}$ . One has

$$\begin{aligned} \tilde{b}_{i,l,1} &= \sum_{m=0}^L \tilde{h}_{i,m} \tilde{h}_{i-l,m-l}^* \\ &= \sum_{m=0}^L (\tilde{h}_{i_0,m} + \tilde{\varepsilon}_{i,m,0}) (\tilde{h}_{i_0,m-l}^* + \tilde{\varepsilon}_{i-l,m-l,0}^*) \\ &= a(l) + \tilde{\varepsilon}_{i,l,1} \end{aligned} \quad (75)$$

where

$$\tilde{\varepsilon}_{i,l,1} = \sum_{m=0}^L \left( \tilde{h}_{i_0,m} \tilde{\varepsilon}_{i-l,m-l,0}^* + \tilde{\varepsilon}_{i,m,0} \tilde{h}_{i_0,m-l}^* + \tilde{\varepsilon}_{i,m,0} \tilde{\varepsilon}_{i-l,m-l,0}^* \right). \quad (76)$$

Using Jensen's inequality  $(\sum_{i=1}^n x_i)^t \leq n^{t-1} \sum_{i=1}^n x_i^t$ , one has

$$\mathbb{E}|\tilde{\varepsilon}_{i,l,1}|^t \leq (3(L+1))^{t-1} \sum_{m=0}^L \left( \mathbb{E}|\tilde{h}_{i_0,m} \tilde{\varepsilon}_{i-l,m-l,0}|^t + \mathbb{E}|\tilde{\varepsilon}_{i,m,0} \tilde{h}_{i_0,m-l}|^t + \mathbb{E}|\tilde{\varepsilon}_{i,m,0} \tilde{\varepsilon}_{i-l,m-l,0}|^t \right). \quad (77)$$

By the Cauchy-Schwarz inequality  $\mathbb{E}|XY| \leq \sqrt{\mathbb{E}|X|^2 \mathbb{E}|Y|^2}$ , the first term in the summation can be bounded by

$$\begin{aligned} \mathbb{E}|\tilde{h}_{i_0,m} \tilde{\varepsilon}_{i-l,m-l,0}|^t &\leq \sqrt{\mathbb{E}|\tilde{h}_{i_0,m}|^{2t} \mathbb{E}|\tilde{\varepsilon}_{i-l,m-l,0}|^{2t}} \\ &\leq \sqrt{t! \delta^{t!}} = t! \delta^{t/2} \end{aligned} \quad (78)$$

because the variance of complex Gaussian random variables  $\tilde{h}_{i_0,m}$  and  $\tilde{\varepsilon}_{i-l,m-l,0}$  are bounded by 1 and  $\delta$ , respectively. Similarly

$$\mathbb{E}|\tilde{\varepsilon}_{i,m,0} \tilde{h}_{i_0,m-l}|^t \leq t! \delta^{t/2}$$

and

$$\mathbb{E}|\tilde{\varepsilon}_{i,m,0} \tilde{\varepsilon}_{i-l,m-l,0}|^t \leq t! \delta.$$

Therefore, the  $t$ th moment of  $\tilde{\varepsilon}_{i,l,1}$  is bounded by

$$\begin{aligned} \mathbb{E}|\tilde{\varepsilon}_{i,l,1}|^t &\leq (3(L+1))^t t! \delta^{t/2} (2 + \delta^{t/2}) \\ &\leq (2L+1)^t C_{t,1} \delta^{t/2} \end{aligned} \quad (79)$$

where  $C_{t,1}$  is a constant depending on  $t$  and  $\delta$  is upper-bounded by 4.

Before proceeding to compute  $\tilde{b}_{i,l,k}$ , we need to bound the moments of  $a_k(m)$

$$\begin{aligned} \mathbb{E}|a_k(m)|^{2t} &= \mathbb{E} \left| \frac{1}{2\pi} \int \hat{h}(\omega) |^{2k} e^{jm\omega} d\omega \right|^{2t} \\ &\leq \mathbb{E} \left( \frac{1}{2\pi} \int |\hat{h}(\omega)|^{2k} d\omega \right)^{2t} \\ &\leq \mathbb{E} \left[ \frac{1}{2\pi} \int |\hat{h}(\omega)|^{2(k+t)} d\omega \right] \\ &= (k+t)! \end{aligned} \quad (80)$$

where we have used the integral version of the Jensen's inequality to push power  $2t$  inside the integral.

Now we compute  $\tilde{b}_{i,l,2}$  as

$$\begin{aligned} \tilde{b}_{i,l,2} &= \sum_{m=-L}^L \tilde{b}_{i,m,1} \tilde{b}_{i-l,m-l,1}^* \\ &= \sum_{m=-L}^L (a(m) + \tilde{\varepsilon}_{i,m,1})(a^*(m-l) + \tilde{\varepsilon}_{i-l,m-l,1}^*) \\ &= a_2(l) + \tilde{\varepsilon}_{i,l,2} \end{aligned} \quad (81)$$

where

$$\tilde{\varepsilon}_{i,l,2} = \sum_{m=-L}^L \left( a(m) \tilde{\varepsilon}_{i-l,m-l,1}^* + \tilde{\varepsilon}_{i,m,1} a^*(m-l) + \tilde{\varepsilon}_{i,m,1} \tilde{\varepsilon}_{i-l,m-l,1}^* \right). \quad (82)$$

Similar as the case of  $\tilde{b}_{i,l,1}$ , one has, by using (79) and (80)

$$\mathbb{E}|\tilde{\varepsilon}_{i,l,2}|^t \leq (2L+1)^{3t} C_{t,2} \delta^{t/2} \quad (83)$$

where  $C_{t,2}$  is a constant depending on  $t$ .

In general,  $\tilde{b}_{i,l,k}$  is given by

$$\tilde{b}_{i,l,k} = \sum_{m=-L}^L \tilde{b}_{i,m,k-1} \tilde{b}_{i-l,m-l,1}. \quad (84)$$

Hence, repeating the above argument to give

$$\mathbb{E}|\tilde{\varepsilon}_{i,l,k}|^t = \mathbb{E}|\tilde{b}_{i,l,k} - a_k(l)|^t \leq (2L+1)^{2k-1} C_{t,k} \delta^{t/2} \quad (85)$$

where  $C_{t,k}$  is a constant depending on  $t$  and  $k$ . Since

$$|M_k - k!| = |\mathbb{E}[\tilde{b}_{i_0,0,k}] - \mathbb{E}[a_k(0)]| \leq \mathbb{E}|\tilde{\varepsilon}_{i_0,0,k}|$$

the theorem is proved by substituting  $t = 1$  in (85).

#### ACKNOWLEDGMENT

The authors gratefully appreciate comments from anonymous reviewers for improving the exposition.

#### REFERENCES

- [1] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Mag.*, vol. 28, pp. 5–14, May 1990.
- [2] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2619–2692, Oct. 1998.
- [3] W. Kozek and A. F. Molisch, "Nonorthogonal pulse shapes for multicarrier communications in doubly dispersive channels," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1579–1589, Oct. 1998.
- [4] W. Kozek, "Adaptation of Weyl-Heisenberg frames to underspread environments," in *Gabor Analysis and Algorithm: Theory and Applications*, H. G. Feichtinger and T. Strohmer, Eds. Boston, MA: Birkhäuser, 1997, pp. 323–352.
- [5] R. Haas and J. C. Belfiore, "A time-frequency well-localized pulse for multiple carrier transmission," *Wireless Personal Commun.*, vol. 5, pp. 1–18, 1997.
- [6] T. Strohmer, "Approximation of dual Gabor frames, window decay, and wireless communications," *Appl. Comp. Harm. Anal.*, vol. 11, no. 2, pp. 243–262, 2001.
- [7] H. Bölcskei, R. Koetter, and S. Mallik, "Coding and modulation for underspread fading channels," in *Proc. 2002 IEEE Intl. Symp. Information Theory*, Lausanne, Switzerland, June 2002, p. 358.
- [8] I. Daubechies, *Ten Lectures on Wavelets*, ser. cBMS-NSF Reg. Conf. Series in Applied Mathematics. Philadelphia, PA: SIAM, 1992.
- [9] H. Jensen, T. Høholdt, and J. Justesen, "Double series representation of bounded signals," *IEEE Trans. Inform. Theory*, vol. 34, pp. 613–623, July 1988.
- [10] A. J. E. M. Janssen, "Duality and bioorthogonality for Weyl-Heisenberg frames," *J. Fourier Anal. Appl.*, vol. 1, no. 4, pp. 403–436, 1995.
- [11] P. A. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun. Syst.*, vol. CS-11, pp. 360–393, Dec. 1963.
- [12] J. G. Proakis, *Digital Communications*: McGraw-Hill, 1995.
- [13] A. M. Sayeed and B. Aazhang, "Joint multipath-Doppler diversity in mobile wireless communications," *IEEE Trans. Commun.*, vol. 47, pp. 123–132, Jan. 1999.
- [14] R. S. Kennedy, *Performance Estimation of Fading Dispersive Channels*. New York: Wiley, 1968.
- [15] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. San Diego, CA: Academic, 1998.
- [16] K. Liu, "Orthogonal time-frequency signaling over doubly dispersive channels," M.Sc. thesis, Univ. Wisconsin-Madison, 2001.

- [17] P. Schniter, "A new approach to multicarrier pulse design for doubly dispersive channels," in *Proc. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2003.
- [18] I. Daubechies, "The wavelet transform, time-frequency localization, and signal analysis," *IEEE Trans. Inform. Theory*, vol. 36, pp. 961–1005, Sept. 1990.
- [19] G. Stüber, *Principles of Mobile Communication*, 2nd ed. Norwell, MA: Kluwer Academic, 2001.
- [20] M. R. Baissas and A. M. Sayeed, "Pilot-based estimation of time-varying multipath channels for coherent CDMA receivers," *IEEE Trans. Signal Processing*, vol. 50, pp. 2037–2049, Aug. 2002.
- [21] T. Kadous, K. Liu, and A. Sayeed, "Optimal time-frequency signaling for rapidly time-varying channels," in *Proc. 2001 IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP'01)*, May 2001.
- [22] S. Moshavi, "Multi-user detection for DS-CDMA communications," *IEEE Commun. Mag.*, pp. 124–136, Oct. 1996.
- [23] M. K. Varanasi and B. Aazhang, "Multistage detection in asynchronous code division multiple access channels," *IEEE Trans. Commun.*, vol. 38, pp. 509–519, Apr. 1990.
- [24] K. Liu and A. M. Sayeed, "Improved layered space-time processing in multiple antenna systems," in *Proc. 39th Annu. Allerton Conf. Communication, Control and Computing*, Monticello, IL, Oct. 2001, pp. 757–767.
- [25] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," Tech. Rep. Bell Labs., 1995.
- [26] V. Kafedziski, "Capacity of frequency selective fading channels with side information," in *Proc. 32nd Asilomar Conf. Signals, Systems and Computers*, vol. 2, 1998, pp. 1758–1762.
- [27] Z. D. Bai, "Methodologies in spectral analysis of large dimensional random matrices, a review," *Stat. Sinica*, vol. 9, pp. 611–677, 1999.
- [28] S. A. Molchanov, L. A. Pastur, and A. M. Khorunzhi, "Limiting eigenvalue distribution for band random matrices," *Theor. Math. Phys.*, vol. 90, pp. 108–118, 1992.
- [29] L. H. Ozarow, S. Shamai (Shitz), and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 359–378, May 1994.
- [30] R. G. Gallager and M. Medard, "Bandwidth scaling for fading channels," in *Proc. 1997 IEEE Intl. Symp. Information Theory*, Ulm, Germany, 1997, p. 471.
- [31] İ. E. Telatar and D. N. C. Tse, "Capacity and mutual information of wideband multipath fading channels," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1384–1400, July 2000.
- [32] R. Knopp and P. A. Humblet, "On coding for block fading channels," *IEEE Trans. Inform. Theory*, vol. 46, pp. 189–205, Jan. 2000.
- [33] G. Hariharan, Z. Hong, and A. Sayeed, "A full diversity scheme for the doubly dispersive channel under the orthogonal signaling framework," in *Proc. 41th Allerton Conf. Communication, control and Computing*, Monticello, IL, Oct. 2003, pp. 1486–1495.