Minimum Probability of Error in Sparse Wideband Channels

Gautham Hariharan and Akbar M. Sayeed

Abstract— **This paper studies the impact of channel coherence and the role of channel state information (CSI) on the probability of error in wideband multipath fading channels. Inspired by recent ultra wideband channel measurement campaigns, we propose a *sparse* channel model for time and frequency selective fading in which the number of resolvable channel paths grow sub-linearly with signal space dimensions. As a result, the degrees of freedom (DoF) in the channel scale sub-linearly. With perfect CSI, the DoF reflect the amount of channel diversity whereas with no CSI, the DoF represent the level of channel uncertainty. Based on this model, we investigate the reliability of wideband communication using error exponents. With perfect CSI at the receiver, it is seen that sparse channels are less reliable to communicate over than rich channels in which the paths scale linearly with signaling dimensions. When there is no receiver CSI, we analyze training-based communication schemes aimed at learning the sparse channel and our results reveal a fundamental tradeoff between channel learnability and diversity that affects error performance. We present numerical examples with realistic parameter sets and discuss the conditions under which the two effects can be balanced to obtain the best probability of error in any practical system.**

I. Introduction

Emerging applications in ultra wideband (UWB) radio and energy constrained networks (e.g. sensor networks) have led to renewed interest in exploring the fundamental performance limits of wideband/low-power communication from an information-theoretic perspective. The availability of channel state information at the receiver (CSIR) is an important factor that drastically impacts performance in the low power regime, as illustrated in [2], [3], [4]. A common theme that emerges from these papers is the critical impact of CSIR on the choice of signaling schemes for the multipath channel so that it achieves the capacity of the additive white Gaussian noise (AWGN) channel in the limit of infinite bandwidth. In particular, while [2] shows that spread-spectrum signals are a poor choice at large bandwidths, the authors in [3] show that peaky frequency shift keying signals achieve AWGN capacity in the wideband limit. Practical wideband communication systems however use a large but finite bandwidth. The main differences between the infinite bandwidth and the large, finite bandwidth scenarios is quantified precisely in the seminal work [4], where it is also shown that flash signals are necessary to achieve the infinite bandwidth capacity limit when there is no CSIR. But these signals, in addition to being peaky and hence impractical, achieve the wideband limit extremely slowly (first-order optimal but not second-order optimal in the language of [4]).

In spite of these advances, there has been little work on the impact on performance of the physical scattering environment and on how the wideband channel structure couples with the availability and the quality of CSIR. For example, a recent work [5] investigates the effect of channel path growth on capacity and concludes that direct sequence spread-spectrum with duty cycle signaling (peakiness in time) is sufficient to achieve the infinite bandwidth limit, if the number of channel paths grows sub-linearly with bandwidth. It is clear that the key to a more complete understanding lies in establishing a connection between the physical channel character and the degrees of freedom (DoF) in the channel and study its impact on capacity. The authors in [6] study the intermediate case between the coherent and non-coherent extremes under the assumption that the channel coherence time increases as we decrease the signal-to-noise ratio per degree of freedom. Using peaky Gaussian signaling and with a coherence time - SNR scaling of the form $T_{coh} = \frac{1}{\mathsf{SNR}^{\mu}}$, they conclude that $\mu > 1$ is necessary and sufficient for first-order optimality while $\mu > 2$ also ensures second-order optimality. However, as remarked by the authors themselves in [6, Lemma 2], there is no physical basis for such scaling in coherence time with SNR.

In this paper, we consider joint time-frequency signaling and propose a sparse multipath channel model which provides a natural mechanism for the scaling of channel coherence with signaling duration and bandwidth. The motivation for this model also arises from recent measurement results (see e.g. [13] and references therein) for UWB channels, which show that the number of propagation paths in the physical environment scale *sub-linearly* with bandwidth. This is in contrast to rich multipath, where the number of paths scale linearly with bandwidth, and which is the assumption in most existing works. We used this model recently in [7], where it is shown that for training-based communication schemes using *non-peaky* signals, the notions of joint time-frequency coherence and sparsity lead to dramatically reduced requirements on the channel coherence time necessary to achieve first- and second-order optimality.

The focus of this paper is on applying the sparse channel model to characterize the reliability of wideband communication, a measure of which is provided by the *reliability function* and the *random coding error exponent* [1]. Error exponents describe the rate at which the probability of error (P_e) decays with codeword length; they can be used to determine the code lengths required for a desired P_e and thus provide an estimate on the decoding complexity. For

This work was supported in part by the NSF under grant #CCF-0431088.
The authors are with the Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI 53706 USA. gauthamh@cae.wisc.edu akbar@engr.wisc.edu

the case of perfect CSIR and narrowband block Gaussian fading, random coding error exponents have been analyzed in [8], [9], [10]. While [9] investigated the average P_e using random coding and outage probability (slow fading), the authors in [10] provide several upper and lower P_e bounds and analyze the performance of coded diversity schemes. For the low SNR wideband non-coherent channel, [11], [12] have recently investigated the role of error exponents. The authors in [12] study the capacity and reliability of the non-coherent wideband multiple-antenna channel with coherence scaling similar to [6] and compute the random coding exponent using the capacity achieving peaky Gaussian signaling input. From a capacity point of view, channel uncertainty is the only limiting factor and peaky signals aim to reduce the penalty due to channel uncertainty. However from a system design perspective, given a fixed amount of transmitter resources and certain quality of service (QoS) requirements (threshold P_e, maximum delay, etc), it is not very clear how the channel and signal parameters interact in determining the error performance. For a given set of system constraints, we ask the follwoing question: what is the best strategy to obtain the minimum P_e? In this work, we identify a fundamental tradeoff between the ability to learn the multipath channel (channel estimation) and the level of delay-Doppler diversity (DoF) for the family of sparse channels. As we will see, no such phenomenon occurs in the case of rich channels.

We use orthogonal short-time Fourier (STF) basis functions [15] for signaling over the time-varying multipath channel. Without any peakiness in the input, we first explore the random coding error exponent with perfect CSIR. It is observed that sparse channels have smaller error exponents than rich channels, which can be attributed to the larger number of DoF in rich environments. Furthermore, reliability decreases, either as we increase sparsity or as we increase the signaling duration over which we code. For training-based communication schemes, our results reveal a fundamental tradeoff between *channel learnability* and diversity in sparse channels. The sparser the channel (larger the channel coherence), easier it is to learn the channel. However increased sparsity means that we communicate over fewer DoF, which affects reliability. We derive the necessary conditions in terms of the channel coherence dimension to obtain the optimal tradeoff (best P_e) and this provides the key to choosing signaling parameters matched to given channel conditions. We will also use the tradeoff analysis to characterize the minimum-energy-per-nat required to guarantee a certain reliability for any desired communication rate.

II. Multipath Channel Model

In this section we describe the sparse channel model and assumptions based on which our analysis is done. A physical discrete multipath channel can be modeled as

$$y(t) = \sum_n \beta_n x(t-\tau_n) e^{j2\pi\nu_n t} + w(t) \qquad (1)$$

where β_n, $\tau_n \in [0, T_m]$ and $\nu_n \in [-W_d/2, W_d/2]$ denote the complex path gain, delay and Doppler shift associated with the n-th path and $w(t)$ is complex AWGN. T_m and W_d represent the delay and Doppler spreads produced by the channel. We assume an underspread channel, $T_m W_d < 1$, which is valid for most radio channels.

A. Delay-Doppler Diversity

For a signaling duration T and bandwidth W, the channel admits the following decomposition [14] in terms of resolvable multipath delays and Doppler shifts as illustrated in Fig. 1(a)

$$\begin{aligned} y(t) &= \sum_{\ell=0}^{\lceil T_m W \rceil} \sum_{m=-\lceil TW_d/2 \rceil}^{\lceil TW_d/2 \rceil} h_{\ell,m} x(t-\ell/W) e^{j2\pi mt/T} \\ h_{l,m} &\approx \sum_{n \in A_{l,m}} \beta_n \end{aligned} \qquad (2)$$

where $A_{l,m} = \{n : \ell/W - 1/2W < \tau_n \leq \ell/W + 1/2W,\ m/T - 1/2T < \nu_n \leq m/T + 1/2T\}$ is the set of all paths whose delays and Doppler shifts lie within the (ℓ, m)-th delay-Doppler resolution bin. The number of resolvable paths (the number of dominant non-vanishing $h_{l,m}$) signifies the delay-Doppler diversity, D, afforded by the channel (the number of statistically independent degrees of freedom (DoF))

$$\begin{aligned} D &= D_T D_W \leq D_{max} = D_{T,max} D_{W,max} \\ D_{T,max} &= \lceil TW_d \rceil \ , \ D_{W,max} = \lceil T_m W \rceil \end{aligned} \qquad (3)$$

where $D_{T,max}$ denotes the maximum number of resolvable Doppler shifts (maximum Doppler/time diversity) and $D_{W,max}$ denotes maximum number of resolvable delays (maximum delay/frequency diversity). Both $D_{T,max}$ and $D_{W,max}$ increase linearly with T and W, respectively, and represent a rich multipath environment in which each delay-Doppler resolution bin in Fig. 1(a) is populated with a path. On the other hand, as illustrated by the dotted resolution bins in Fig. 1(a), physical multipath channels get sparser with increasing W due to fewer than $D_{W,max}$ resolvable delays and with increasing T due to fewer than $D_{T,max}$ resolvable Doppler shifts. We model such sparse multipath channels with sub-linear scaling in D:

$$D_T \sim (\lceil TW_d \rceil)^{\delta_1} \ , \ D_W \sim (\lceil WT_m \rceil)^{\delta_2} \ , \ \delta_1, \delta_2 \in [0,1] \qquad (4)$$

where the smaller the value of δ the slower (sparser) the growth in the resolvable paths in the corresponding domain. Note that this also implies that the number of DoF, $D = D_T D_W$, scale sub-linearly with the number of signal space dimensions denoted by $N = TW$.

B. Time-Frequency Coherence

We consider signaling using an orthonormal short-time Fourier (STF) basis [15] that serves as an approximate eigen basis for sufficiently underspread channels ([15], [16]) and naturally relates delay-Doppler diversity to coherence in time-frequency. Representing (1) with respect to the STF basis yields an N-dimensional matrix system equation

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \qquad (5)$$

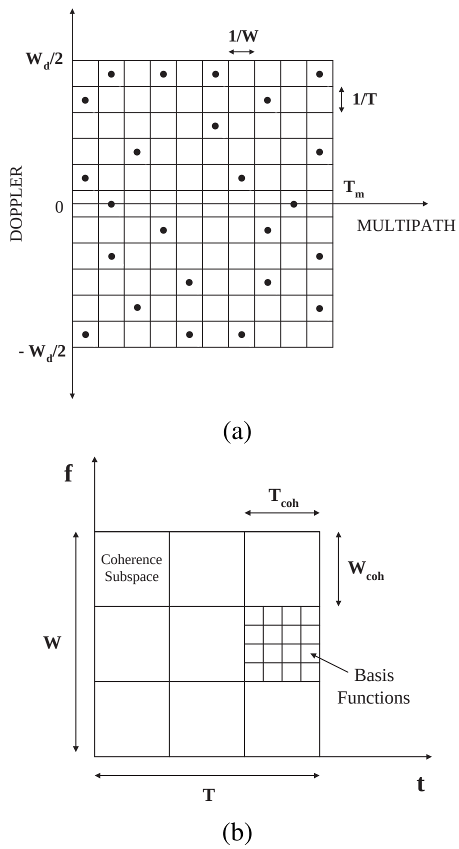

Fig. 1. *(a) Delay-doppler sampling commensurate with signaling duration and bandwidth. (b) Time-frequency coherence subspaces in short-time Fourier signaling.*

in which the $N \times N$ channel matrix $\mathbf{H}$ is approximately diagonal and $\mathbf{w}$ is complex AWGN vector whose components are i.i.d $\mathcal{CN}(0,1)$.

$$\mathbf{H} = \text{diag}\Big[\underbrace{h_{1,1}\cdots h_{1,N_c}}_{\text{Subspace } 1},\ \underbrace{h_{2,1}\cdots h_{2,N_c}}_{\text{Subspace } 2}\ \cdots\ \underbrace{h_{D,1}\cdots h_{D,N_c}}_{\text{Subspace } D}\Big] \tag{6}$$

We assume $\mathbf{H}$ to be exactly diagonal in this paper, as justified in [15]. The diagonal entries of $\mathbf{H}$ also admit an intuitive block fading interpretation in terms of *time-frequency coherence subspaces* illustrated in Fig. 1(b): the signal space is partitioned as $N = TW = N_c D$ where D represents the number of statistically independent TF coherence subspaces (delay-Doppler diversity; see (3)) and N_c represents the dimension of each coherence subspace. In the block fading model, the channel coefficients over the i-th coherence subspace $h_{i,1}\cdots h_{i,N_c}$ are assumed to be identical, h_i, and the channel coefficients over the different coherence subspaces are assumed to be i.i.d. zero-mean Gaussian random variables (Rayleigh fading). The variance of the channel coefficients is normalized to $\mathbf{E}[|h_i|^2] = \sum_n \mathbf{E}[|\beta_n|^2] = 1$. The dimension of each coherence subspace is given by

$$N_c = T_{coh}W_{coh} = \frac{T}{D_T}\cdot\frac{W}{D_W} = \frac{T^{1-\delta_1}}{W_d^{\delta_1}}\frac{W^{1-\delta_2}}{T_m^{\delta_2}} \geq \left\lceil\frac{1}{T_m W_d}\right\rceil \tag{7}$$

where $T_{coh} = T^{1-\delta_1}/W_d^{\delta_1}$ is the *coherence time* and $W_{coh} = W^{1-\delta_2}/T_m^{\delta_2}$ is the *coherence bandwidth* of the channel (see Fig. 1(b)). Note that $\delta = 1$ corresponds to a rich multipath channel in which $N_c = N_{coh,min} = 1/(T_m W_d)$ is fixed and $D = D_{max}$ increases linearly with N. The other extreme of $\delta = 0$ models an extremely sparse channel in which D is fixed and N_c scales linearly with N. For sparse channels ($\delta \in (0,1)$), both N_c and D *increase sub-linearly* with T and W. Thus while we obtain ergodic channel behavior with large signaling dimensions, the sub-linear growth in D enables us to track the channel and communicate reliably. For simplicity, we will assume $\delta_1 = \delta_2 = \delta$ in the following analysis.

It can be seen from (7) that channel coherence can scale in two possible ways, either as δ changes (effect of channel) or as we change T and(or) W (effect of signal space). Thus for a given δ, the signal space parameters can be suitably chosen to obtain a desired level of operational coherence. Using (7) and defining the parameter $\mathsf{SNR} = \frac{P}{N_0 W} = \frac{P}{W}$ (P is the transmit power and $N_0 = 1$ as in (5)) leads to $N_c = \frac{T^{1-\delta}}{(T_m W_d)^\delta}\frac{P^{1-\delta}}{\mathsf{SNR}^{1-\delta}}$. Furthermore, for any fixed P, by varying T with W according to $T = \frac{(k)^{\frac{1}{1-\delta}}(T_m W_d)^{\frac{\delta}{1-\delta}} W^\alpha}{P^{1+\alpha}}$ provides us with the following relationship between channel coherence dimension and SNR.

$$N_c = \frac{k}{\mathsf{SNR}^\mu} \tag{8}$$

$$\text{where} \quad \mu = (1+\alpha)(1-\delta)$$

As we will see later, the reliability of both sparse and rich channels is a function of only N_c and SNR and so a relationship of the form in (8) helps us analyze the low SNR asymptotics and illustrate the learnability vs. diversity tradeoff. However the key difference between our work and the assumptions in [6], [12] is that for a sparse channel we are able to provide a realizable and *natural* relationship between N_c and SNR that is physically justified.

III. Random Coding Error Exponent with Perfect CSIR

In this section we analyze the random coding error exponent for sparse and rich channels under the assumption of perfect CSIR. The reliability function of a channel is defined as [1, Chp.5]

$$E(R) = \lim_{N\to\infty}\sup\frac{-\log\ P_e(N,R)}{N}$$

where $P_e(N,R)$ is the average probability of error over an ensemble of codes in which each codeword spans the signal space dimensions $N = TW$ and communication takes place at a transmission rate R (in nats/s/Hz). For any fixed N, the random coding exponent, $E_r(N,R)$, and the sphere packing exponent, $E_{sp}(N,R)$, are lower and upper bounds respectively on $E(R)$. Furthermore, $E_r(N,R) = E_{sp}(N,R)$ for a range of rates $R_{cr} \leq R \leq C$, where R_{cr} is defined as the critical rate and C refers to channel capacity. We recall the random coding upper bound on P_e [1], [10] given by

$$P_e \leq e^{-N[E_r(N,R)]} \quad (9)$$

$$E_r(N,R) = \max_{0\leq\rho\leq 1} \max_{Q} \left[E_o(N,\rho,Q) - \rho R\right] \quad (10)$$

$$E_o(N,\rho,Q) = -\frac{1}{N}\log\left(E_{\mathbf{H}}\left[\int_{\mathbf{y}}\left[\int_{\mathbf{x}} q(\mathbf{x})p(\mathbf{y}|\mathbf{x},\mathbf{H})^{\frac{1}{1+\rho}}dx\right]^{1+\rho}dy\right]\right) \quad (11)$$

Closed form expressions for the error exponent in (10) are unknown, mainly because of the lack of knowledge about the optimal input distribution. For temporal block fading, [8], [9] have previously investigated error exponents with Gaussian inputs and [8, Th. 22] shows that although not optimal, the Gaussian input provides a tight lower bound. For the doubly block fading model considered here, we fix the input to be i.i.d. complex-valued Gaussian for the analysis. In addition to the fact that this is the capacity achieving input (see [7]), it makes the calculations tractable and provides us with a tight lower bound to $E_r(N,R)$. For the system model in (5) and (6), the integrals in (11) can be evaluated to obtain

$$\begin{aligned} E_o(N,\rho) &= -\frac{1}{N}\log\left(E_{\mathbf{H}}\left[\prod_{i=1}^{N}\left(1+\frac{\mathsf{SNR}|h_i|^2}{1+\rho}\right)^{-\rho}\right]\right) \\ &\overset{(a)}{=} -\frac{1}{N}\log\left(\prod_{k=1}^{D}E_{\mathbf{H}}\left[\left(1+\frac{\mathsf{SNR}|h_k|^2}{1+\rho}\right)^{-N_c\rho}\right]\right) \\ &\overset{(b)}{=} -\frac{1}{N_c}\log\left(E_{\mathbf{H}}\left[\left(1+\frac{\mathsf{SNR}|h|^2}{1+\rho}\right)^{-N_c\rho}\right]\right) \quad (12) \end{aligned}$$

where (a) follows from the block structure of the channel in (6) and (b) is due to the fact that $\{h_k\}$ are i.i.d. The expectation in (12) can be computed for arbitrary SNR in terms of the incomplete Gamma function [17]

$$E_o(N,\rho) = -\frac{1}{N_c}\log\left[e^{\frac{1+\rho}{\mathsf{SNR}}}\left(\frac{\mathsf{SNR}}{1+\rho}\right)^{-N_c\rho}\Gamma\left(1-N_c\rho,\frac{1+\rho}{\mathsf{SNR}}\right)\right] \quad (13)$$

and the optimization over ρ can be performed numerically to compute the best exponent $E_r(N,R)$ (as is done for example in [10]). We will avoid the numerical optimization and instead obtain a lower bound to $E_o(N,\rho)$ in (12) and use the lower bound to compute the optimal ρ^* and analytically characterize the error exponent. This line of thought was previously advocated in [8] and more recently [12] uses a similar approach. The key factor that is exploited in deriving these bounds is the fact that we are in the $\mathsf{SNR}\to 0$ regime. First though, we present an upper bound.

Lemma 1. *$E_o(N,\rho)$ in (12) is upper bounded by*

$$E_o(N,\rho) \leq E_o^{UB} = \frac{1}{N_c}\log\left(1+\frac{N_c\mathsf{SNR}\rho}{1+\rho}\right) \quad (14)$$

Proof. Notice that we can rewrite (12) as

$$\begin{aligned} E_o(N,\rho) &= -\frac{1}{N_c}\log\left(E_{\mathbf{H}}\left[\left(1+\frac{\mathsf{SNR}|h|^2}{1+\rho}\right)^{-N_c\rho}\right]\right) \\ &= -\frac{1}{N_c}\log\left(E_{\mathbf{H}}\left[\exp\left(-N_c\rho\log\left(1+\frac{\mathsf{SNR}|h|^2}{1+\rho}\right)\right)\right]\right) \quad (15) \end{aligned}$$

The result follows by applying the log inequality $\log\left(1+\frac{\mathsf{SNR}|h|^2}{1+\rho}\right) \leq \left(\frac{\mathsf{SNR}|h|^2}{1+\rho}\right)$ and then computing the expectation in closed form. □

The following lemma presents a lower bound to $E_o(N,\rho)$.

Lemma 2. *$E_o(N,\rho)$ in (12) can be lower bounded by*

$$E_o(N,\rho) \geq E_o^{LB} = \frac{1}{N_c}\log\left(1+\frac{N_c\mathsf{SNR}\left(1-\mathsf{SNR}^{1-\epsilon}\right)\rho}{1+\rho}\right) - o(1) \quad (16)$$

for $\epsilon > 0$ and sufficiently small.

Proof. We provide a sketch of the proof. The expectation in (12) equals

$$I = \int_0^\infty \left(1+\frac{\mathsf{SNR}v}{1+\rho}\right)^{-N_c\rho}\exp[-v]\,\mathrm{d}v$$

where $V = |h|^2$ with p.d.f. $f_V(v) = \exp[-v],\ v > 0$. By rewriting the integrand as in (15) and splitting the integral into two parts

$$\begin{aligned} I = &\int_0^c \exp\left[-N_c\rho\log\left(1+\frac{\mathsf{SNR}v}{1+\rho}\right)\right]\exp[-v]\ \mathrm{d}v \\ &+ \int_c^\infty \exp\left[-N_c\rho\log\left(1+\frac{\mathsf{SNR}v}{1+\rho}\right)\right]\exp[-v]\ \mathrm{d}v \end{aligned}$$

where $c = \left(\frac{1+\rho}{\mathsf{SNR}^\epsilon}\right)$ for small enough $\epsilon > 0$. By simplifying and further bounding the two integrals it can be shown that

$$I \leq I_1 + I_2$$

where

$$\begin{aligned} I_1 &= \left[1+\frac{N_c\rho\mathsf{SNR}\left(1-\mathsf{SNR}^{1-\epsilon}\right)}{1+\rho}\right]^{-1} \\ I_2 &= \exp\left[-\left(\frac{1+\rho}{\mathsf{SNR}^\epsilon}\right) - N_c\rho\mathsf{SNR}^{1-\epsilon}\left(1-\mathsf{SNR}^{1-\epsilon}\right)\right] \end{aligned}$$

Next, using the relation in (8) $N_c = \frac{k}{\mathsf{SNR}^\mu}$ ($\mu = 0$ corresponds to a rich channel and $\mu > 0$ is a sparse channel) and defining

$$\begin{aligned} f(\mathsf{SNR}) &= N_c\,\mathsf{SNR}^{1-\epsilon}\left(1-\mathsf{SNR}^{1-\epsilon}\right) \\ &= k\,\mathsf{SNR}^{1-\epsilon-\mu}\left(1-\mathsf{SNR}^{1-\epsilon}\right) \end{aligned}$$

we have

$$\lim_{\mathsf{SNR}\to 0}\left(\frac{I_2}{I_1}\right) = \lim_{\mathsf{SNR}\to 0}\left[\left(1+\frac{f(\mathsf{SNR})\mathsf{SNR}^\epsilon\rho}{1+\rho}\right)\ \exp\left[-\left(\frac{1+\rho}{\mathsf{SNR}^\epsilon}\right) - \rho f(\mathsf{SNR})\right]\right] \quad (17)$$

The limit is now analyzed in the following cases: Case 1 : $0 \leq \mu < (1-\epsilon)$, Case 2 : $(1-\epsilon) \leq \mu < 1$ and Case 3 : $\mu \geq 1$. In all three cases the limit in (17) equals zero and as a consequence, we have $I_2 = o(I_1)$. This implies that

$$I \leq I_1 + o(I_1) = I_1(1+o(1))$$

Putting things together into (12) completes the proof. □

For our numerical results to follow we will ignore the $o(1)$ terms that are negligble. The approximation improves depending on the value of ϵ, although this comes at the expense of a very small loss in tightness in the upper bound. The following theorem provides the required P_e random coding upper bound using the result of Lemma 2.

Theorem 1. *The probability of error for the N-dimensional channel in (5), using i.i.d. Gaussian input and maximum-likelihood decoding at the receiver, is upper bounded by*

$$P_e \leq e^{-N\left[E_r^{LB}(N,R)\right]}$$

where $E_r^{LB}(N,R)$ *is given in (18)*

A. Perfect CSIR: Sparse vs Rich Channels

We present numerical results for a realistic parameter set to illustrate the findings of Theorem 1. Figure 2 shows the random coding error exponent, normalized by the SNR as a function of transmission rate (normalized by R_{max}) for fixed N and varying δ. It can be observed that the error exponent is largest for the rich channel and decreases monotonically as the channel gets sparse (smaller δ). This is because with perfect CSIR, we would get a lower P_e for a channel with higher diversity ($\delta = 1$) than one with lower diversity ($0 < \delta < 1$). Conversely, for a desired P_e, we will need to transmit at a lower rate in the sparser channel to meet the P_e requirement.

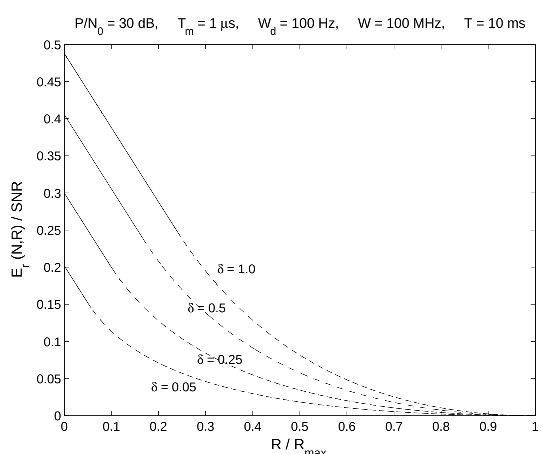

Fig. 2. $E_r(N,R)$ /SNR *vs. normalized transmission rate with perfect CSIR. Effect of varying* δ*. Solid line denotes the linear portion* ($\rho^* = 1$) *and the dashed line is the region where* $0 < \rho^* < 1$. R_{cr} *defines the boundary between the two regions.*

The effect of signaling duration, T, on the random coding exponent is illustrated in Fig. 3 where the normalized exponent is plotted for a fixed $\delta = 0.5$ and different values of T. Once again we notice that the error exponent of sparse channels decreases uniformly, both with rate and with increasing duration T. However, the error exponent of a rich channel is a constant and independent of T. The critical rate, R_{cr}, and the cut-off rate, defined as $R_o = E_r(N,0)$, also decrease, either as δ gets smaller as seen in Fig. 2 or as we increase T for fixed δ (Fig. 3). They eventually approach zero in either scenario, thereby characterizing the reliability completely for all transmission rates. Similar findings have been reported previously in [9], although the authors used a different model there; it being quasi-static and block fading in time ($\delta = 0$ case in our model). To summarize the perfect CSIR case, while sparsity has no impact on capacity, it significantly degrades the reliability of communication for all rates below capacity.

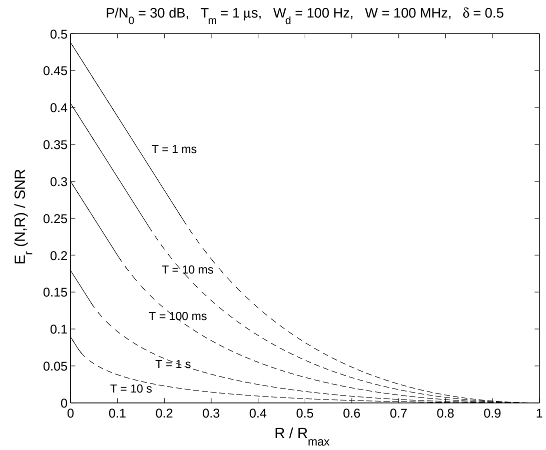

Fig. 3. $E_r(N,R)$ /SNR *vs. normalized transmission rate with perfect CSIR illustrating the effect of increasing signaling duration*

IV. Error Exponents for a Training-based Communication System

We now turn our attention to the more practical scenario where there is no CSIR *a priori*. We consider schemes that explicitly send training signals to learn the channel and then communicate using knowledge of the estimated channel coefficients. The training scheme used here is similar to the one in [6], adapted suitably for the STF signaling framework. The total energy available for training and communication per codeword is PT, of which a fraction η is used for training and the remaining fraction $(1-\eta)$ is used for communication. We will use the mean square error (MSE) of the estimated channel coefficient in each coherence subspace as a measure of the quality of learning. Since the MSE over one coherence subspace depends only on the training energy rather than on the number of dimensions, our scheme uses one signal space dimension in each coherence subspace for training and the remaining $(N_c - 1)$ for communication. The training energy per coherence subspace is given by $E_{tr} = \frac{\eta TP}{D} = \eta N_c \text{SNR}$ and the communication energy per coherence subspace is given by $(1-\eta)N_c\text{SNR}$. Thus the variance of the minimum mean squared error ($MMSE$) estimate of the channel coefficient in each coherence subspace is given by $E\left[|\widehat{h}_i|^2\right] = \frac{E_{tr}}{1+E_{tr}}$, $1 \leq i \leq D$, and the corresponding MSE is given by $E\left[|\widetilde{h}_i|^2\right] = \frac{1}{1+E_{tr}}$. The communication phase can be

$$E_r^{LB}(N,R) = \begin{cases} \frac{1}{N_c}\log\left(1+\frac{N_c\mathsf{SNR}\left(1-\mathsf{SNR}^{1-\epsilon}\right)}{2}\right) - R - o(1) & 0 \le R \le R_{cr} \\ \frac{1}{N_c}\log\left(1+\frac{N_c\mathsf{SNR}\left(1-\mathsf{SNR}^{1-\epsilon}\right)\rho^*}{1+\rho^*}\right) - \rho^* R - o(1) & R_{cr} \le R \le R_{max} \\ 0 & R > R_{max} \end{cases} \quad (18)$$

$$R_{cr} = \frac{\mathsf{SNR}\left(1-\mathsf{SNR}^{1-\epsilon}\right)}{2\left[2+N_c\mathsf{SNR}\left(1-\mathsf{SNR}^{1-\epsilon}\right)\right]} \quad (19)$$

$$R_{max} = \mathsf{SNR}\left(1-\mathsf{SNR}^{1-\epsilon}\right) \quad (20)$$

described by

$$\begin{aligned} \mathbf{y}_d = \mathbf{H}\mathbf{x}_d + \mathbf{w}_d &= \widehat{\mathbf{H}}\mathbf{x}_d + \Delta\mathbf{H}\mathbf{x}_d + \mathbf{w}_d \\ &= \widehat{\mathbf{H}}_n\mathbf{x}_d + \mathbf{n}_d \end{aligned} \quad (21)$$

where $\mathbf{H}$ is the $(N_c-1)D$-dimensional communication channel, $\widehat{\mathbf{H}}$ and $\Delta\mathbf{H}$ represent the estimated channel matrix and the estimation error matrix respectively. The data vector $\mathbf{x}_d$ is zero-mean i.i.d. Gaussian with covariance matrix $\mathbf{Q} = \left(\frac{(1-\eta)N_c\mathsf{SNR}}{N_c-1}\right)\mathbf{I}_{(N_c-1)D}$. Furthermore (21) has been normalized so that the diagonal entries in $\widehat{\mathbf{H}}_n$ have unit variance. The overall noise vector $\mathbf{n}_d$ is the sum of the additive white Gaussian noise and the noise due to estimation error. For our analysis, we will assume that $\mathbf{n}_d$ has i.i.d. zero mean Gaussian entries with covariance matrix $\mathbf{Q}_n = \left(\frac{1+\frac{(1-\eta)N_c\mathsf{SNR}}{(N_c-1)(1+\eta N_c\mathsf{SNR})}}{\frac{\eta N_c\mathsf{SNR}}{1+\eta N_c\mathsf{SNR}}}\right)\mathbf{I}_{(N_c-1)D}$ and as conjectured in [12], this should provide a lower bound to the error exponent.

With the details of the training scheme in place, we investigate the random coding exponent, the result of which is summarized in the following theorem.

Theorem 2. *The average probability of error for the training-based communication scheme is upper-bounded by,*

$$P_e \le e^{-N\left[E_r^{tr}(N,R)\right]}$$

where $E_r^{tr}(N,R)$ is given in (22)

Proof. Once the channel coefficients have been estimated with the aid of training signals, we are back to the coherent set-up of Sec. III, although with some errors due to estimation. Therefore, we can compute the random coding exponent, $E_o^{nc}(N,\rho)$ starting with the integral form in (11) as in the perfect CSIR case. Evaluating the integrals and simplifying using the channel structure and the i.i.d. property of the estimated channel coefficients as in (12) now yields

$$E_o^{nc}(N,\rho) = -\frac{1}{N_c}\log\ E_{\widehat{\mathbf{H}}_n}\left[\left(1+\frac{K}{1+\rho}|\widehat{h}|^2\right)^{-(N_c-1)\rho}\right] \quad (28)$$

where

$$K(N_c,\mathsf{SNR}) = \frac{\eta(1-\eta)\left(N_c\mathsf{SNR}\right)^2}{(N_c-1)(1+\eta N_c\mathsf{SNR}) + (1-\eta)N_c\mathsf{SNR}} \quad (29)$$

The expression in (28) needs to be maximized over the fraction of energy used for training, η, to obtain the best exponent. It is easily seen that this is equivalent to maximizing K in (29) as a function of η. Performing this maximization yields K^* and the optimal η^* as in (26) and (27) (next page). It is interesting to note that the η^* maximizing the error exponent here equals the one that maximized a lower bound on the mutual information related our results on capacity in [7]. Next, instead of computing the expectation, we use the lower bound of Lemma 2 that leads to the following bound

$$E_o^{nc}(N,\rho) \ge \frac{1}{N_c}\log\left(1+\frac{\rho K^*\left(1-(K^*)^{1-\epsilon}\right)}{1+\rho}\right) \quad (30)$$

Using the lower bound and maximizing over ρ to obtain the best exponent completes the proof. □

V. Discussion of Results

A. Asymptotic Coherence of Sparse Channels

The performance of the channel estimation scheme is an indicator of the position in an intermediate regime, whether we are closer to the coherent or to the non-coherent extreme. The quality of channel estimation is measured using the following metrics: (i) MSE of channel estimates and (ii) optimal fraction of total energy used for estimation, η^*. The following theorem provides the condition in terms of the channel coherence to obtain efficient and consistent estimation.

Theorem 3. *In the limit of large signal space dimensions, the performance metrics of the training scheme satisfy*

$$MSE \to 0 \quad \textit{and} \quad \eta^* \to 0 \quad (31)$$

if and only if $N_c = \frac{k}{\mathsf{SNR}^{\mu}}$ and $\mu > 1$

Proof. See Appendix A □

The result says that sparse channels are *asymptotically coherent* provided the channel coherence dimension satisfies the condition specified in Theorem 2. Whether or not this condition is satisfied depends critically on the choice of signal parameters (T, W) and transmit power, P. Using (7) and (8) along with $T = W^{\alpha}$ it can be shown that

$$\begin{aligned} (1+\alpha)(1-\delta) &> 1 \\ \text{or} \quad \alpha &> \frac{\delta}{(1-\delta)} \end{aligned} \quad (32)$$

$$
\begin{aligned}
E_r^{tr}(N,R) &= \begin{cases} \frac{1}{N_c}\log\left(1+\frac{(N_c-1)K^*\left(1-(K^*)^{1-\epsilon}\right)}{2}\right) - R - o(1) & 0 \le R \le R_{cr} \\ \frac{1}{N_c}\log\left(1+\frac{(N_c-1)K^*\left(1-(K^*)^{1-\epsilon}\right)\rho^*}{1+\rho^*}\right) - \rho^* R - o(1) & R_{cr} \le R \le R_{max} \\ 0 & R > R_{max} \end{cases} && (22) \\
R_{cr} &= \left(\frac{N_c-1}{2N_c}\right)\frac{K^*\left(1-(K^*)^{1-\epsilon}\right)}{[2+(N_c-1)K^*(1-(K^*)^{1-\epsilon})]} && (23) \\
R_{max} &= \left(\frac{N_c-1}{N_c}\right)K^*\left(1-(K^*)^{1-\epsilon}\right) && (24) \\
R_o &= \frac{1}{N_c}\log\left(1+\frac{(N_c-1)K^*\left(1-(K^*)^{1-\epsilon}\right)}{2}\right) && (25) \\
K^* &= \frac{\eta^*(1-\eta^*)N_c\mathsf{SNR}^2}{(N_c-1)(1+\eta^* N_c\mathsf{SNR})+(1-\eta^*)N_c\mathsf{SNR}} && (26) \\
\eta^* &= \frac{N_c\mathsf{SNR}+N_c-1}{(N_c-2)N_c\mathsf{SNR}}\left[\sqrt{1+\frac{(N_c-2)N_c\mathsf{SNR}}{N_c\mathsf{SNR}+N_c-1}}-1\right] && (27)
\end{aligned}
$$

provides the desired conditions. Similarly, for a fixed T (delay-limited system), it can be shown that if $P \propto W^\beta$, then $\beta \in (\delta, 1)$ is necessary and sufficient to satisfy the condition in Theorem 3. The rate at which the MSE and η^* approach zero can also be derived and the dependance is given by $\eta \to 0$ as $\frac{1}{\sqrt{N_c\mathsf{SNR}}}$ and $MSE \to 0$ ($E_{tr} \to \infty$) as $\sqrt{N_c\mathsf{SNR}}$. On the other hand no such phenomenon occurs in the case of rich channels, where N_c is a constant and does not scale with SNR ($\mu = 0$). In fact the optimal training scheme in this case asymptotically uses half the total energy ($\eta^* \to \frac{1}{2}$) to estimate the channel coefficients and the MSE does not decay to zero as can be observed from the proof of Theorem 3.

B. *Learnability vs. Diversity Tradeoff*

In this sub-section, we present numerical results to illustrate the main result of this work, which is a fundamental tradeoff between channel learnability and diversity in sparse channels. Figure 4 illustrates the effect of sparsity and compares the error exponent for a family of sparse channels and for the rich channel using the training-based communication scheme. The corresponding curves with perfect CSIR (as in Fig. 2) are also shown for reference. For clarity, the exponent is drawn on a log scale. It is observed that for any fixed transmission rate, the error exponent increases as the channel gets sparser, peaks at a particular value of δ, and then decreases as we make δ smaller. The value of δ at which the maximum occurs is not the same for all rates and depends on the fraction of R_{max} at which we signal.

In Fig. 5 we fix the level of channel sparsity and investigate the effect of increasing signaling duration, T. Once again we observe a similar effect: for every transmission rate R, there is an optimal T that maximizes the error exponent. It is not beneficial to increase T indefinitely in sparse channels. As we traverse from R_{max} to smaller rates, this optimal T decreases until we hit a critical duration T_c, that maximizes the error exponent at $R = 0$.

We explore this further in Fig. 6 where the effect of channel spread factor and signal parameters on the error exponent is investigated by plotting the cut-off rate in (25). The cut-off rate is the error exponent at $R = 0$. For each scenario, the T at which the maximum is achieved defines T_c. We notice that smaller the $T_m W_d$, smaller is T_c. For fixed $T_m W_d$, T_c decreases as we decrease δ (sparser channels). The error exponent behavior in these plots is the result of a fundamental tradeoff between channel estimation performance and the inherent level diversity in the channel that manifests as we vary δ or T in all of the above plots.

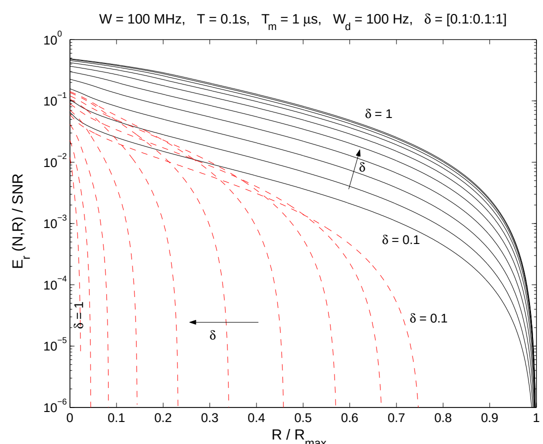

Fig. 4. *$E_r^{tr}(N,R)$ for a family of sparse channels; Solid lines denote the perfect CSIR case while dashed lines denote the training-based scheme*

It is clear from the plots that for fixed signaling dimensions there exists a $\delta \in (0,1)$ at which P_e is minimum. This is illustrated in Fig. 7 in which we plot P_e as a function of δ for the exact same parameters of Fig. 4 and for several bandwidths. Notice that P_e is minimized at a critical $\delta = \delta_c$ for each W (highlighted for $W = 1$GHz in Fig. 7). This point of minimum P_e (or maximum error exponent in Figs. 4, 5, 6) characterizes the tension between two competing effects (i) channel learnability, quantified in terms of the MSE of the channel estimates and (ii) the delay-Doppler diversity afforded by the channel. Wireless channels that are more

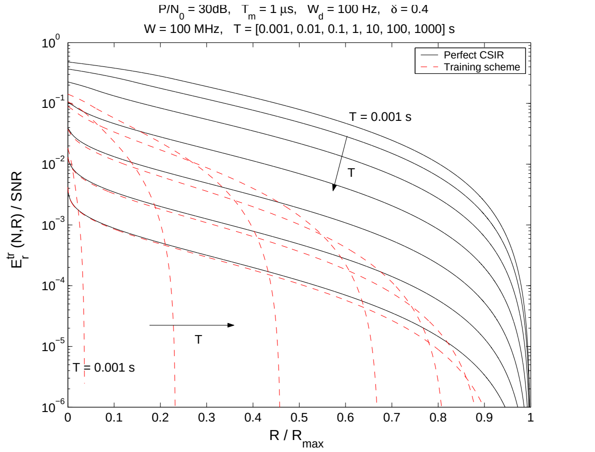

Fig. 5. *Effect of changing signaling duration T on $E_r^{tr}(N, R)$ for a sparse channel with $\delta = 0.4$*

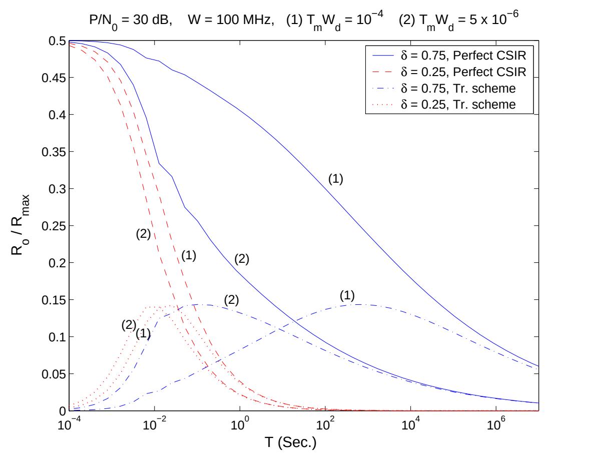

Fig. 6. *Cut-off rate vs. T for perfect CSIR and training-based communication schemes*

sparse ($\delta < \delta_c$) can be estimated better (smaller MSE) than those that are less sparse ($\delta > \delta_c$). However, when there is perfect CSIR, the latter class of channels are more reliable to communicate over due to the larger number of DoF (as was illustrated earlier in Fig. 2). The best P_e is obtained when the two effects are balanced at the optimal value $\delta = \delta_c$. This value of δ_c can be derived in terms of the signal parameters as

$$\delta_c = \frac{\log(TP/k)}{\log(TWT_m W_d)} \tag{33}$$

Furthermore, the condition on channel coherence dimension that defines the optimal tradeoff point can be derived as summarized in the following theorem.

Theorem 4. *For the family of underspread sparse channels, a necessary and sufficient condition that minimizes the probability of decoding error for the training-based communication scheme is*

$$N_c = \frac{k}{\mathsf{SNR}} \tag{34}$$

Proof. See Appendix B □

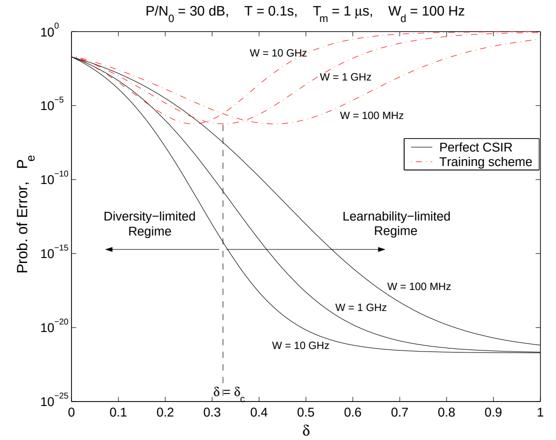

Fig. 7. *P_e vs. δ that illustrates the learnability-diversity tradeoff*

Theorem 4 provides guidelines for choosing the signaling parameters that match the behavior of the sparse channel in order to obtain the best P_e. In particular, the result says that given a δ, the best P_e is obtained by choosing signal parameters $T \propto W^{\frac{\delta}{1-\delta}}$. The result also captures the fundamental difference between the requirements on channel coherence for minimum probability of error as opposed to those for maximizing mutual information [6], [7]. From a capacity point of view, channel uncertainty is the only limiting factor. This was quantified in [7, Th. 1] where we showed that with $N_c = \frac{k}{\mathsf{SNR}^{\mu}}$, first order optimality is obtained for $\mu > 1$ and second order optimality is obtained in addition if $\mu > 3$. The choice of signal parameters is driven by the requirement to make the cost of channel estimation as small as possible and the quality of the channel estimates near-perfect (the conditions of Theorem 3). However, as far as P_e goes, we have two competing effects. Given a fixed amount of resources at the transmitter (energy), it is not necessarily better to aim for arbitrarily good channel estimation. Peaky signaling schemes of the kind proposed in [6], [12], although good from a capacity point of view, certainly do not provide the best P_e. In the context of our framework, such schemes implicitly fall into the learnability-enhanced but diversity-limited regime of Fig. 7. However, as we have seen, the best error performance is obtained by finding the right balance between learnability and diversity.

C. Energy-per-nat

Most practical communication systems have certain quality of service (QoS) requirements. The primary requirements are transmission at a specified rate, say r nats/s and with probability of decoding error guaranteed to be less than a threshold value, $P_e \leq e^{-\gamma}$ ($\gamma-$reliability). Furthermore, delay requirements usually impose a hard constraint on the codeword duration T. We would like to know what is the minimum energy-per-nat to satisfy the requirements, given the channel spread parameters and sparsity $\delta = \delta_0$. In other words, what is the minimum transmit power P required to ensure $\gamma-$reliability, given other system constraints. In the framework of our analysis, it turns out that minimum energy is expended when we jointly choose the power-bandwidth pair (P, W) so that the system operates at the optimal tradeoff point. The result is summarized in the following corollary.

Corollary 1. *For communication at r nats/s and a fixed T, the choice of the power-bandwidth tuple $(P,\ W)$ to achieve $\gamma-$reliability with the least amount of energy satisfies*

$$P = P_{min} = \frac{k}{\log(1+\frac{k}{2})}\ \left[\frac{\gamma}{T} + \rho^* r\right] \quad (35)$$

$$W = W_{crit} = \left(\frac{1}{T_m W_d}\right)\ \left(\frac{P_{min}}{k}\right)^{\frac{1}{\delta_0}}\ T^{\frac{1-\delta_0}{\delta_0}} \quad (36)$$

The corresponding minimum energy-per-nat equals

$$\left(\frac{E_n}{N_o}\right)^{th} = \frac{k}{\log(1+\frac{k}{2})}\ \left[\frac{\gamma}{Tr} + \rho^*\right] \quad (37)$$

Proof. We provide a sketch of the proof. The result follows by combining the achievable upper bound to P_e in (22) along with the requirement for threshold $\gamma-$reliability

$$P_e \leq e^{-TW\left[E_r^{tr}(N,R)\right]} \text{ and we want } P_e \leq e^{-\gamma}$$

Therefore the choice of P and W should satisfy

$$TW\left[E_r^{tr}(N,R)\right] \geq \gamma \quad (38)$$

Given the values of T_m, W_d, δ_0 and T, and for any fixed power P, E_r^{tr} is maximized under the condition of Theorem 4. This means that we can guarantee $\gamma-$reliability with the least amount of power when we choose (P, W) jointly satisfying both (38) and (34) and this leads directly to the result. The corresponding energy-per-nat is obtained by using $\left(\frac{E_n}{N_o}\right) = \frac{P_{min}}{N_0 r}$. For the low rate case of $R < R_{cr}$ ($\rho^* = 1$), the solution gets simplified since equations (35) and (36) can be easily decoupled. □

Communication with reliability greater than or equal to γ is possible with the least energy-per-nat if we not only choose $P = P_{min}$ as in (35) but in addition also use a unique $W = W_{crit}$ according to (36). If we communicate at $W < W_{crit}$, we operate in the diversity limited regime at $\delta = \delta_0$ and cannot achieve $\gamma-$reliability. Similar behavior is observed when $W > W_{crit}$, this time the system operates in the learnability limited regime. In either case, we require $P > P_{min}$ at $\delta = \delta_0$ and signaling takes place with $\left(\frac{E_n}{N_o}\right) > \left(\frac{E_n}{N_o}\right)^{th}$.

VI. Concluding Remarks

In this paper we used random coding error exponents to investigate the error probability of multipath fading channels in the wideband regime. We introduced a sparse multipath model in which the number of resolvable paths due to multipath and Doppler shifts scales sub-linearly with the signal space dimensions, as motivated by recent measurement campaigns. We used orthogonal STF signaling waveforms for communication that maps sparsity in delay-Doppler to channel coherence in time-frequency. The main result of our work is the illustration of a fundamental tradeoff between channel learnability and diversity in sparse wideband channels. For communication systems with practical delay and energy constraints, we provided guidelines for choosing signal parameters to match the channel behavior to obtain the best tradeoff. Code construction using practical signaling alphabets and designs based on the insights obtained here would be a natural extension to pursue.

An assumption inherent in our model is the linear scaling of channel power with signal space dimensions. We are currently exploring the effect of sub-linear channel power growth on capacity and reliability in the light of recent results [18] that demonstrate the profound impact of channel power on capacity scaling in MIMO channels. We believe that a thorough investigation of these factors will ultimately lead to a better understanding of finite-energy communication [19] over multipath fading channels.

Appendix

A. Proof of Theorem 3

To prove the result we will follow the approach of [7, Th. 1]. Using $N_c = \frac{k}{\mathsf{SNR}^{\mu}}$ we rewrite

$$\eta^* = K_1 K_2, \qquad K_1 = \frac{\mathsf{SNR}^{\mu-1}(k\mathsf{SNR}+k-\mathsf{SNR}^{\mu})}{k(k-2\mathsf{SNR}^{\mu})}$$

$$K_2 = \left[\sqrt{1+\frac{k\mathsf{SNR}^{1-\mu}\ (k-2\mathsf{SNR}^{\mu})}{k\mathsf{SNR}+k-\mathsf{SNR}^{\mu}}} - 1\right] \quad (39)$$

In order to track the behavior of $MSE = \frac{1}{1+E_{tr}}$ we study $E_{tr} = \eta^* N_c \mathsf{SNR} = k\ \mathsf{SNR}^{1-\mu} K_1 K_2$. We consider the asymptotics in either of the following two scenarios: (i) fixed μ and $\mathsf{SNR} \to 0$ (as would be the case if we increase W and scale T appropriately) (ii) fixed SNR and increasing μ (for a system with large but fixed W and increasing T). The analysis is done over the following three cases: Case 1: $\mu < 1$, Case 2: $\mu = 1$ and Case 3: $\mu > 1$.

Case 1: When $\mu < 1$ we have

$$K_1 = \frac{\mathsf{SNR}^{\mu-1}}{k} \sum_{i=\{0,1\}} \sum_{j=0}^{j=\infty} \mathcal{O}\left(\mathsf{SNR}^{i+j\mu}\right)$$

$$K_2 = \left(\frac{k}{2}\ \mathsf{SNR}^{1-\mu}\ \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \mathcal{O}\left(\mathsf{SNR}^{i+j\mu}\right)\right) \quad (40)$$

This leads to

$$\eta^* = \frac{1}{2} + \sum_{i=0}^{\infty}\sum_{j=2}^{\infty} \mathcal{O}\left(\mathsf{SNR}^{i+j\mu}\right)$$

$$E_{tr} = \frac{k}{2}\ \mathsf{SNR}^{1-\mu} + \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \mathcal{O}\left(\mathsf{SNR}^{i+j\mu}\right) \quad (41)$$

which implies that $\eta^* \to \frac{1}{2}$ and $MSE \to 1$.

Case 2: When $\mu = 1$

$$K_1 = \frac{1}{k} + o\left(\mathsf{SNR}\right)$$

$$K_2 = \left(\sqrt{1+k+o\left(\mathsf{SNR}\right)} - 1\right) \quad (42)$$

The above relationships imply that

$$\eta^* \to \left(\frac{\sqrt{1+k}-1}{k}\right)$$

$$E_{tr} \to \left(\sqrt{1+k}-1\right) \quad \text{or} \quad MSE \to \frac{1}{\sqrt{1+k}} \quad (43)$$

Case 3: For $\mu > 1$, K_1 is the same as in (40) but the asymptotic expansion for K_2 is

$$K_2 = \sqrt{k}\, \mathsf{SNR}^{\frac{1-\mu}{2}} - 1 + o(1) \tag{44}$$

It is easy to see in this case that $\eta^* \to 0$. Similarly it follows that $E_{tr} \to \infty$ and so $MSE \to 0$.

B. Proof of Theorem 4

We prove the result for the simple case of low rates ($R < R_{cr}$) for which $\rho^* = 1$. The proof for the general case follows by similar arguments. For fixed T and W, finding the minimum P_e is equivalent to maximizing the exponent in the P_e upper bound of Theorem 2 which equals $TWE_r^{tr}(N,R) = TWE_o^{tr}(N,\rho) - TWR$. Since we consider transmission at a fixed rate, the maximization is just over the first term, which using (30) equals

$$\begin{aligned} T_1 &= \frac{TP}{N_c \mathsf{SNR}} \log\left(1 + \frac{(N_c-1)\,K^*\left(1-(K^*)^{1-\epsilon}\right)}{2}\right) \\ &= \frac{TP}{N_c \mathsf{SNR}} \log(1+L) \end{aligned} \tag{45}$$

where we have used $\mathsf{SNR} = \frac{P}{W}$ and we define $L = \frac{(N_c-1)K^*\left(1-(K^*)^{1-\epsilon}\right)}{2}$. First we focus on the term K^*. Using $N_c = \frac{k}{\mathsf{SNR}^{\mu}}$, we get the following asymptotic expansion that follows directly from [7, Th.1]

$$K^* = \mathsf{SNR}^{\max(1,2-\mu)} + o\left(\mathsf{SNR}^{\max(1,2-\mu)}\right) \tag{46}$$

Now we consider the low SNR asymptotics of T_1 in the following cases: Case 1: $\mu < 1$, Case 2: $\mu = 1$ and Case 3: $\mu > 1$.

Case 1: When $\mu < 1, K^* = \mathsf{SNR}^{(2-\mu)} + o\left(\mathsf{SNR}^{(2-\mu)}\right)$ which implies

$$\begin{aligned} L =& \tfrac{k}{2}\left(1 - \tfrac{\mathsf{SNR}^{\mu}}{k}\right)\mathsf{SNR}^{-\mu}\left[\mathsf{SNR}^{(2-\mu)} + o\left(\mathsf{SNR}^{(2-\mu)}\right)\right] \\ &\times\left[1 - \mathsf{SNR}^{(2-\mu)(1-\epsilon)} - o\left(\mathsf{SNR}^{(2-\mu)(1-\epsilon)}\right)\right] \\ =& \tfrac{k}{2}\,\mathsf{SNR}^{(2-2\mu)}\,(1+o(1))\left(1 - \tfrac{\mathsf{SNR}^{\mu}}{k}\right) \\ &\times\left[1 - \mathsf{SNR}^{(2-\mu)(1-\epsilon)} - o\left(\mathsf{SNR}^{(2-\mu)(1-\epsilon)}\right)\right] \end{aligned}$$

Using the approximation $\log(1+L) \approx L$ since $L \to 0$, we get

$$\begin{aligned} T_1 \approx& \tfrac{TP}{2}\,\mathsf{SNR}^{(1-\mu)}\,(1+o(1))\left(1 - \tfrac{\mathsf{SNR}^{\mu}}{k}\right) \\ &\times\left[1 - \mathsf{SNR}^{(2-\mu)(1-\epsilon)} - o\left(\mathsf{SNR}^{(2-\mu)(1-\epsilon)}\right)\right] \end{aligned} \tag{47}$$

Case 2: In this case $K^* = \mathsf{SNR} + o(\mathsf{SNR})$ which leads to

$$\begin{aligned} L =& \tfrac{k}{2}\,(1+o(1))\left(1 - \tfrac{\mathsf{SNR}}{k}\right) \\ &\times\left[1 - \mathsf{SNR}^{(1-\epsilon)} - o\left(\mathsf{SNR}^{(1-\epsilon)}\right)\right] \\ =& \frac{k}{2}\,(1+o(1)) \end{aligned}$$

Thus we have

$$T_1 = \frac{TP}{k}\log\left(1 + \frac{k}{2}\,(1+o(1))\right) \tag{48}$$

Case 3: Once again $K^* = \mathsf{SNR} + o(\mathsf{SNR})$ which implies

$$\begin{aligned} L =& \tfrac{k}{2}\,\mathsf{SNR}^{(1-\mu)}\,(1+o(1))\left(1 - \tfrac{\mathsf{SNR}^{\mu}}{k}\right) \\ &\times\left[1 - \mathsf{SNR}^{(1-\epsilon)} - o\left(\mathsf{SNR}^{(1-\epsilon)}\right)\right] \\ =& \frac{k}{2}\,\mathsf{SNR}^{(1-\mu)}\,(1+o(1)) \end{aligned}$$

Therefore

$$T_1 = \frac{TP}{k}\,\mathsf{SNR}^{(\mu-1)}\log\left(1 + \frac{k}{2}\,\mathsf{SNR}^{(1-\mu)}\,(1+o(1))\right) \tag{49}$$

It can be seen from (47),(48) and (49) that in the limit, either as $\mathsf{SNR} \to 0$ for fixed μ (or) for fixed but small SNR and increasing μ, the maximum value for T_1 is obtained when $\mu = 1$. This concludes the proof.

References

[1] R. G. Gallager "Information Theory and Reliable Communication," Wiley & Sons, 1968.

[2] M. Medard and R. G. Gallager, "Bandwidth scaling for fading multipath channels," *IEEE Trans. Inform. Theory*, April 2002.

[3] E. Telatar and D. Tse, "Capacity and mutual information of wideband multipath fading channels," *IEEE Trans. Inform. Theory*, July 2000.

[4] S. Verdu, "Spectral Efficiency in the Wideband Regime," *IEEE Trans. Inform. Theory*, June 2002

[5] D. Porrat, D. Tse, "Bandwidth Scaling in Ultrawideband Communication," *Proc. 41st Allerton Annual Conference*, Oct. 2003

[6] L. Zheng, M. Medard, D. Tse, "Channel Coherence in the Low SNR Regime," *submitted to IEEE Trans. Inform. Theory*.

[7] V. Raghavan, G. Hariharan, A. Sayeed, "Exploiting Time-Frequency Coherence to Achieve Coherent Capacity in Wideband Wireless Channels," *Proc. of 43rd Allerton Annual Conference*, Sep. 2005

[8] W. K. M. Ahmed, "Information Theoretic Reliability Function for Flat Fading Channels," *Ph.D. Thesis*, Dept. of ECE, Queen's University, Sep 1997.

[9] G. Kaplan, S. Shamai, "Error Probabilities for the Block Fading Gaussian Channel," *AEU*, vol. 49, pp. 192-205, July 1995.

[10] E. Malkamaki and H. Leib, "Coded Diversity on Block-Fading Channels," *IEEE Trans. on Inform. Theory*, Mar. 1999.

[11] X. Wu, R. Srikant, "Near-Optimal Signaling for Wideband Fading Channels," *Proc. of ISIT* July 2004.

[12] S. Ray, M. Medard, L. Zheng, "On Non-coherent MIMO Channels in the Wideband Regime: Capacity and Reliability," *submitted to IEEE Trans. Inform. Theory*.

[13] A. Molisch, "Ultrawideband Propagation Channels - Theory, Measurement and Modeling," *IEEE Trans. Veh. Tech.,* Sep. 2005.

[14] A. Sayeed, B. Aazhang, "Joint Multipath-Doppler Diversity in Mobile Wireless Communications," *IEEE Trans. Comm.*, Jan. 1999.

[15] K. Liu, T. Kadous, A. M. Sayeed, "Orthogonal Time-Frequency Signaling over Doubly Dispersive Channels, *IEEE Trans. Inform. Theory*, Nov. 2004.

[16] 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.

[17] I. S. Gradshteyn, I. M. Ryzhik, A. Jeffrey, D. Zwillinger, "Table of Integrals, Series, and Products," *Academic Press*, 6th edition, 2000

[18] A. M. Sayeed, V. Raghavan, J. Kotecha, "Capacity of Space-Time Wireless Channels: A Physical Perspective," *Proc. of ITW*, San Antonio, Oct. 2004.

[19] R. G. Gallager, "Energy Limited Channels: Coding, Multiaccess and Spread spectrum," *LIDS Tech. Report*, Nov. 1987.