Non-Coherent Capacity and Reliability of Sparse Multipath Channels in the Wideband Regime

Gautham Hariharan and Akbar M. Sayeed
Department of Electrical and Computer Engineering
University of Wisconsin-Madison
gauthamh@cae.wisc.edu, akbar@engr.wisc.edu

Abstract— In contrast to the prevalent assumption of rich multipath in information theoretic analysis of wireless channels, physical channels exhibit sparse multipath, especially at large bandwidths. We propose a model for sparse multipath fading channels and present results on the impact of sparsity on non-coherent capacity and reliability in the wideband regime. A key implication of sparsity is that the statistically independent degrees of freedom in the channel, that represent the delay-Doppler diversity afforded by multipath, scale at a *sub-linear* rate with the signal space dimension (time-bandwidth product). Our analysis is based on a training-based communication scheme that uses short-time Fourier (STF) signaling waveforms. Sparsity in delay-Doppler manifests itself as time-frequency coherence in the STF domain. From a capacity perspective, sparse channels are asymptotically coherent: the gap between coherent and non-coherent extremes vanishes in the limit of large signal space dimension without the need for peaky signaling. From a reliability viewpoint, there is a fundamental tradeoff between channel diversity and learnability that can be optimized to maximize the error exponent at any rate by appropriately choosing the signaling duration as a function of bandwidth.

I. INTRODUCTION

Recent advances in the emerging areas of ultra-wideband communication systems and wireless sensor networks have renewed the search for a complete understanding of the fundamental performance limits in the wideband/low SNR regime. The impact of multipath signal propagation, which leads to fading, on the capacity and reliability of wideband channels, depends critically on knowledge of the channel state information (CSI) at the receiver. The seminal work in [1] best illustrates this: with perfect CSI at the receiver, peak-power limited QPSK achieves second-order optimality, whereas with no receiver CSI, peaky signals are necessary to even achieve first-order optimality, although they fail to be second-order optimal. Motivated by the fact that channel learning can bridge this gap and the sharp cut-off, the authors in [2] study wideband capacity by assuming a coherence time scaling with SNR of the form

$$T_{coh} = \frac{k}{\mathsf{SNR}^\mu} \ , \quad \mu > 0 \tag{1}$$

However there is no explanation for why such scaling laws should hold in practice.

Accurate modeling of the channel characteristics in time and frequency, as a function of physical multipath characteristics, is critical in analyzing the performance of channel learning schemes and the impact of CSI on the performance limits. While most existing results assume rich multipath, there is growing experimental evidence (e.g. [3], [4]) that physical channels exhibit a sparse structure at wide bandwidths and when we code over long signaling durations. In this paper, we use a virtual representation [5] for physical multipath channels to present a framework for modeling sparsity. The virtual representation uniformly samples multipath in delay and Doppler at a resolution commensurate with the signaling bandwidth and signaling duration, respectively. Under this representation, the virtual channel

coefficients represent the DoF in delay and Doppler. Sparse channels correspond to a sparse set of virtual coefficients and a key implication is the sub-linear scaling of the number of independent degrees of freedom (DoF) with signal space dimensions. This is contrast to rich multipath, where the DoF scale linearly. We consider signaling over orthogonal short-time Fourier (STF) basis functions that serve as approximate eigenfunctions for underspread channels and provide a natural mechanism to relate sparsity in delay-Doppler to coherence in time-frequency.

With no receiver CSI *a priori*, we consider training-based communication schemes and investigate the wideband ergodic capacity under the assumption that the *time-frequency coherence dimension* N_c scales with SNR according to

$$N_c = \frac{k}{\mathsf{SNR}^\mu} \ , \quad \mu > 0 \tag{2}$$

It is observed that the coherence requirements for achieving capacity are shared between both time and frequency: the coherence bandwidth, W_{coh}, increases with bandwidth, W (due to sparsity in delay), and the coherence time, T_{coh}, increases with signaling duration T (due to sparsity in Doppler). As a result, the scaling requirements on T_{coh} with W needed in [2] for first- and second-order optimality are replaced by scaling requirements on $N_c = T_{coh}W_{coh}$. This leads to dramatically relaxed requirements on T_{coh} scaling with bandwidth/SNR compared to those assumed in [2]. In particular, sparse multipath channels are *asymptotically coherent*; that is, for a sufficiently large but fixed bandwidth, the conditions for first- and second-order optimality can be achieved by simply making the signaling duration sufficiently large according to

$$T \propto \frac{\left(T_m^{\delta_2} W_d^{\delta_1}\right)^{\frac{1}{1-\delta_1}} W^{\frac{\mu-1+\delta_2}{1-\delta_1}}}{P^{\frac{\mu}{1-\delta_1}}} \tag{3}$$

Equation (3) relates the signaling parameters (T,W,P), as a function of the channel parameters $(T_m,W_d,\delta_1,\delta_2)$ in order for the relationship (2) to hold between N_c and SNR at any desired value of μ (in particular, $\mu > 1$ for first-order optimality and $\mu > 3$ for second-order optimality). The asymptotic coherence of sparse channels also eliminates the need for peaky signaling that has been emphasized in existing results [6], [1] for increasing the spectral efficiency of non-coherent communication schemes.

Our investigation of the reliability of sparse channels is through random coding error exponents [7]. For training-based communication schemes, our results reveal a fundamental *learnability versus diversity tradeoff* in sparse channels. At any transmission rate less than the coherent capacity, there is an optimal choice of signal parameters (as a function of channel parameters) that optimizes the tradeoff and yields the largest error exponent.

II. System Setup

A. Sparse Multipath Channel Modeling

A physical discrete multipath channel can be modeled as

$$
\begin{aligned}
h(\tau, \nu) &= \sum_n \beta_n \delta(\tau - \tau_n)\delta(\nu - \nu_n) \\
r(t) &= \sum_n \beta_n x(t - \tau_n)e^{j2\pi\nu_n t} + w(t) \qquad (4)
\end{aligned}
$$

where $h(\tau, \nu)$ is the delay-Doppler spreading function of the channel, 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. T_m and W_d are the delay and Doppler spreads respectively and $w(t)$ is additive white Gaussian noise (AWGN). We assume a sufficiently underspread channel, $T_m W_d \ll 1$. In this paper we use a *virtual representation* [5], [8] for time- and frequency-selective multipath channels that captures the channel characteristics in terms of *resolvable paths* and greatly facilitates system analysis from a communication-theoretic perspective. The virtual representation uniformly samples the multipath in delay and Doppler at a resolution commensurate with signaling bandwidth W and signaling duration T, respectively [5], [8]

$$
\begin{aligned}
y(t) &= \sum_{\ell=0}^{\lceil T_m W \rceil} \sum_{m=-\lceil TW_d/2 \rceil}^{\lceil TW_d/2 \rceil} h_{\ell,m} x(t - \ell/W)e^{j2\pi mt/T} \quad (5) \\
h_{\ell,m} &\approx \sum_{n \in S_{\tau,\ell} \cap S_{\nu,m}} \beta_n \qquad (6)
\end{aligned}
$$

The sampled representation (5) is linear and is characterized by the virtual delay-Doppler channel coefficients $\{h_{\ell,m}\}$. Each $h_{\ell,m}$ consists of the sum of gains of all paths whose delays and Doppler shifts lie within the (ℓ, m)-th delay-Doppler resolution bin as shown in Fig. 1(a). Distinct $h_{\ell,m}$'s correspond to approximately *disjoint* subsets of paths and are hence approximately statistically independent (due to independent path gains and phases). In this work, we assume that the channel coefficients $\{h_{\ell,m}\}$ are perfectly independent. We also assume Rayleigh fading in which $\{h_{\ell,m}\}$ are zero-mean Gaussian random variables and the channel statistics are thus characterized by the power in the virtual channel coefficients

$$
\Psi(\ell, m) = E[|h_{\ell,m}|^2] \qquad (7)
$$

which is a measure of the (sampled) delay-Doppler power spectrum.

Let D denote the number of dominant [1] non-zero channel coefficients. The parameter D reflects the statistically independent degrees of freedom (DoF) in the channel and also signifies the delay-Doppler diversity afforded by the channel. It can be bounded as

$$
\begin{aligned}
D &= D_T D_W \leq D_{\max} = D_{T,\max} D_{W,\max} \\
D_{T,\max} &= \lceil TW_d \rceil \;\;,\;\; D_{W,\max} = \lceil T_m W \rceil \qquad (8)
\end{aligned}
$$

where $D_{T,\max}$ denotes the maximum number of resolvable paths in Doppler (maximum Doppler or time diversity) and $D_{W,\max}$ denotes maximum number of resolvable paths in delay (maximum delay or frequency diversity). In rich multipath, $D_T = D_{T,\max}$ and $D_W = D_{W,\max}$ and each delay-Doppler resolution bin in Fig. 1(a) is populated by a path. In this case D scales linearly with the signal space dimensions, $N = TW$.

However, recent measurement campaigns [3], [4] for UWB channels show that dominant channel coefficients get sparser in the delay domain as the bandwidth increases. As we consider large bandwidths and/or long signaling durations, the resolution of paths in both delay

---

[1]For which $\Psi(\ell, m) > \gamma$ for some prescribed threshold $\gamma > 0$.
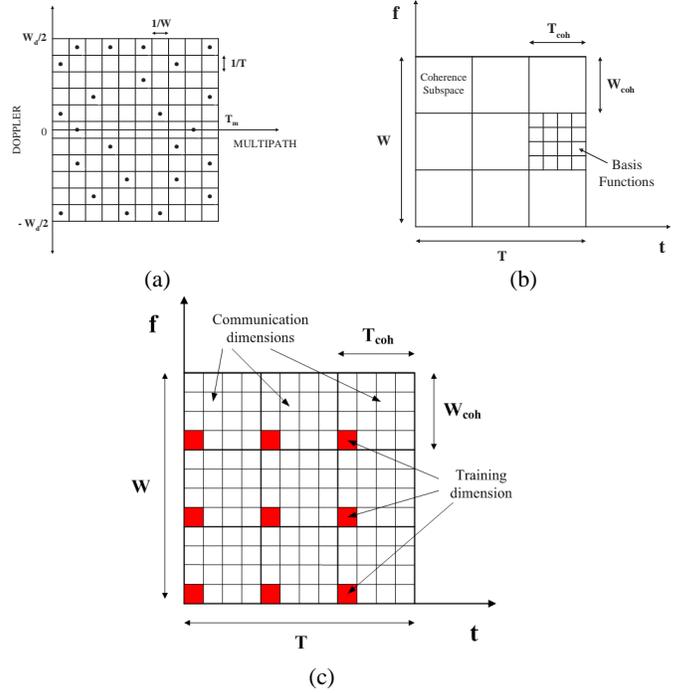

Fig. 1. *(a) Delay-doppler sampling commensurate with signaling duration and bandwidth. (b) Time-frequency coherence subspaces in STF signaling. (c) Illustration of the training-based communication scheme in the STF domain. One dimension in each coherence subspace (dark squares) represent the training dimension and the remaining dimensions are used for communication.*

and Doppler domains gets finer, leading to the scenario in Fig. 1(a) where the delay-Doppler resolution bins are sparsely populated with paths, i.e. $D < D_{max}$. Thus 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 behavior with a *sub-linear* scaling in D_T and D_W with T and W:

$$
D_T \sim (TW_d)^{\delta_1} \;,\; D_W \sim (T_m W)^{\delta_2} \;,\; \delta_1, \delta_2 \in [0, 1] \qquad (9)
$$

where $\{\delta_i\}$ represent channel sparsity; smaller the value of $\{\delta_i\}$, the slower (sparser) the growth in the resolvable paths in the corresponding domain. This implies that the delay-Doppler DoF, $D = D_T D_W$, scale sub-linearly with the number of signal space dimensions N. Note that with perfect CSI at the receiver, D reflects the delay-Doppler diversity afforded by the channel, whereas with no CSI, it reflects channel uncertainty.

B. Orthogonal Short-Time Fourier Signaling

We consider signaling using an orthonormal short-time Fourier (STF) basis [9], [10] that is a natural generalization of orthogonal frequency-division multiplexing (OFDM) for time-varying channels. An orthogonal STF basis for the signal space is generated from a fixed prototype waveform $g(t)$ via time and frequency shifts:

$$
\begin{aligned}
\phi_{\ell,m}(t) &= g(t - \ell T_o)e^{j2\pi W_o t} \quad \text{where} \quad T_o W_o = 1 \\
\ell &= 0, \ldots, N_T - 1, \quad m = 0, \ldots, N_W - 1 \\
N_T &= \frac{T}{T_o}, N_W = \frac{W}{W_o} \text{ and } N = N_T N_W \qquad (10)
\end{aligned}
$$

The N transmitted symbols $x_{\ell,m}$ are modulated onto the STF basis

$$
x(t) = \sum_{\ell,m} x_{\ell,m}\phi_{\ell,m}(t)
$$

For a signaling duration T and bandwidth W, the basis functions span the signal space with dimension equal to $N = TW$.

The received signal in (4) is projected onto the STF basis functions to yield the received symbols

$$r_{\ell,m} = \langle r, \phi_{\ell,m} \rangle = \sum_{\ell',m'} h_{\ell,m;\ell',m'} \; x_{\ell',m'} + w_{\ell,m} \qquad (11)$$

Equivalently, we can represent the system in STF-domain using an N-dimensional matrix system equation

$$\mathbf{r} = \sqrt{\mathsf{SNR}} \; \mathbf{Hx} + \mathbf{w} \qquad (12)$$

where $\mathbf{w}$ represents the additive noise vector whose entries are i.i.d. $\mathcal{CN}(0,1)$. The $N \times N$ matrix consists of the channel coefficients $\{h_{\ell,m;\ell',m'}\}$ in (11). The parameter SNR represents the transmit energy per modulated symbol and for a given transmit power P equals $\mathsf{SNR} = \frac{P}{W}$ ($\mathbf{E}[|\mathbf{x}|^2] = 1$). In this work, our focus is on the wideband regime, where $\mathsf{SNR} \to 0$.

For sufficiently underspread channels, the parameters T_o and W_o can be matched to T_m and W_d so that the STF basis waveforms serve as approximate eigenfunctions of the channel [10], [9]. Thus the $N \times N$ channel matrix $\mathbf{H}$ is approximately diagonal. In this work, we will assume that $\mathbf{H}$ is exactly diagonal, that is,

$$\mathbf{H} = \mathrm{diag}\Big[\underbrace{h_{1,1} \cdots h_{1,N_c}}_{\text{Subspace } 1}, \; \underbrace{h_{2,1} \cdots h_{2,N_c}}_{\text{Subspace } 2} \; \cdots \; \underbrace{h_{D,1} \cdots h_{D,N_c}}_{\text{Subspace } D}\Big] . \qquad (13)$$

Furthermore, the diagonal entries of $\mathbf{H}$ in (13) admit an intuitive block fading interpretation in terms of *time-frequency coherence subspaces* [9] illustrated in Fig. 1(b). The signal space is partitioned as

$$N = TW = N_c D \qquad (14)$$

where D represents the number of statistically independent time-frequency coherence subspaces (delay-Doppler diversity), reflecting the DoF in the channel (see (8)), and N_c represents the dimension of each coherence subspace, which we will refer to as the **time-frequency coherence dimension**. In the block fading model in (13), the channel coefficients over the i-th coherence subspace $h_{i,1} \cdots h_{i,N_c}$ are assumed to be identical, h_i, whereas the coefficients across different coherence subspaces are independent and due to the stationarity of the channel statistics across time and frequency, identically distributed. Thus, the D distinct STF channel coefficients, $\{h_i\}$, are i.i.d. zero-mean Gaussian random variables (Rayleigh fading) with variance $\mathbf{E}[|h_i|^2] = 1$.

Using the DoF scaling for sparse channels in (9), the coherence dimension of each coherence subspace can be computed as

$$T_{coh} = \frac{T}{D_T} = \frac{T^{1-\delta_1}}{W_d^{\delta_1}} \; , \; W_{coh} = \frac{W}{D_W} = \frac{W^{1-\delta_2}}{T_m^{\delta_2}} \qquad (15)$$

$$N_c = T_{coh} W_{coh} = \frac{T^{1-\delta_1}}{W_d^{\delta_1}} \frac{W^{1-\delta_2}}{T_m^{\delta_2}} \geq \left\lceil \frac{1}{T_m W_d} \right\rceil = N_{c,\min} \qquad (16)$$

where T_{coh} is the *coherence time* and W_{coh} is the *coherence bandwidth* of the channel, as illustrated in Fig. 1(b). Note that $\delta_1 = \delta_2 = 1$ corresponds to a rich multipath channel in which $N_c = N_{c,\min} = 1/(T_m W_d)$ is *constant* and $D = D_{\max}$ increases *linearly* with $N = TW$. This is the assumption prevalent in existing works. In contrast, for sparse channels, $(\delta_1, \delta_2) \in (0,1)$, and both N_c and D increase *sub-linearly* with N. In terms of channel parameters, N_c increases with decreasing $T_m W_d$ as well as with smaller δ_i. In terms of signaling parameters, N_c can be increased by increasing T and/or W. On the other hand, when the channel is rich, N_c depends only on $T_m W_d$ and does not scale with T or W.

In this paper, our focus is on computing the sparse channel capacity and reliability and, as we will see later, both metrics turn out to be functions only of the parameters N_c and SNR. Furthermore, in the wideband limit they critically depend on the following relation between N_c and SNR

$$N_c = \frac{k}{\mathsf{SNR}^\mu} \; , \quad \mu > 0 \qquad (17)$$

where $k > 0$ is a constant.

C. Training-Based Communication Using STF Signaling

We use the block fading model induced by STF signaling to study the impact of time-frequency coherence on channel capacity and reliability in sparse multipath channels. Within the non-coherent regime, we focus our attention on a communication scheme in which the transmitted signals include training symbols to enable coherent detection. Although it is argued in [2] that training-based schemes are sub-optimal from a capacity point of view, the restriction to training schemes is motivated by practical considerations.

We provide an outline of the training-based communication scheme, adapted from [2], suitable to STF signaling (see [11] for details). The total energy available for training and communication is PT, of which a fraction η is used for training and the remaining fraction $(1 - \eta)$ is used for communication. Since the quality of the channel estimate over one coherence subspace depends only on the training energy and *not* on the number of training symbols [12], our scheme uses one signal space dimension in each coherence subspace for training and the remaining $(N_c - 1)$ for communication, as illustrated in Fig. 1(c). We consider minimum mean squared error (MMSE) estimation under which the channel estimation performance is measured in terms of the resulting mean squared error (MSE).

III. ERGODIC CAPACITY OF THE TRAINING-BASED COMMUNICATION SCHEME

We first characterize the coherent capacity of the wideband channel with perfect CSI at the receiver. The coherent capacity per dimension (in bps/Hz) is defined as

$$C_{coh}(\mathsf{SNR}) = \sup_{\mathbf{Q}: \; \mathrm{Tr}(\mathbf{Q}) \leq TP} \frac{\mathbf{E}\left[\log_2 \det\left(\mathbf{I}_N + \mathbf{HQH}^H\right)\right]}{N}$$

where P denotes transmit power and $\mathbf{H}$ is the diagonal channel matrix in (13) with the diagonal elements following the block-fading structure. Due to the diagonal nature of $\mathbf{H}$, the optimal $\mathbf{Q}$ is also diagonal. In particular, uniform power allocation $\mathbf{Q} = \frac{TP}{N}\mathbf{I}_N = \mathsf{SNR}\,\mathbf{I}_N$ achieves capacity and we have

$$\begin{aligned} C_{coh}(\mathsf{SNR}) &= \frac{\sum_{i=1}^{D} \mathbf{E}\left[\log_2\left(1 + \frac{TP}{N}|h_i|^2\right)\right]}{D} \\ &\overset{(a)}{=} \mathbf{E}\left[\log_2\left(1 + \mathsf{SNR}|h|^2\right)\right] \end{aligned} \qquad (18)$$

where (a) follows since $\{h_i\}$ are i.i.d. with h representing a generic random variable, $N = TW$ and $\mathsf{SNR} = \frac{P}{W} = \frac{TP}{N}$.

The next proposition provides a lower bound to the coherent capacity in the low SNR regime [11].

Proposition 1: The coherent capacity, $C_{coh}(\mathsf{SNR})$, satisfies

$$C_{coh} \geq \log_2(e)\left(\mathsf{SNR} - \mathsf{SNR}^2\right) \qquad (19)$$

Moreover the capacity converges to the lower bound in the limit of $\mathsf{SNR} \to 0$.

The lower bound in Proposition 1 shows that the minimum energy per bit necessary for reliable communication is given by $\frac{E_b}{N_o}_{\min} = \log_e(2)$ and the wideband slope $S_0 = 1$, the two fundamental metrics of spectral efficiency in the wideband regime defined in [1].

In terms of the scaling law, $N_c = \frac{k}{\mathsf{SNR}^\mu}, \mu > 0$, as defined in (17), we are interested in computing the value of μ such that the training-based communication scheme achieves first- and second-order optimality. The result is summarized in the following theorem.

Theorem 1: The average mutual information of the training-based scheme, with the scaling law $N_c = \frac{k}{\mathsf{SNR}^\mu}$, satisfies

$$I_{tr} \geq \log_2(e) \cdot \left[\mathsf{SNR} - \mathcal{O}\left(\mathsf{SNR}^{\frac{1+\mu}{2}}\right)\right] \tag{20}$$

In particular, the first- and second-order optimality conditions are met if and only if $\mu > 1$ and $\mu > 3$, respectively.

Proof: Omitted for this version. See [11] for details. ∎

It can be seen that the coherence dimension N_c plays a critical role in determining the capacity of the training and communication scheme. Both channel and signaling parameters impact N_c in sparse channels and for a given $T_m W_d$ and $\{\delta_i\}$, the signal space parameters T and W can be suitably chosen to obtain any desired value for N_c. Recalling the expression for W_{coh} in (15), we note that

$$W_{coh} = \frac{W^{1-\delta_2}}{(T_m)^{\delta_2}} = \frac{P^{1-\delta_2}}{(T_m)^{\delta_2}\mathsf{SNR}^{1-\delta_2}} \tag{21}$$

and thus W_{coh} *naturally* scales with SNR. Using (21) the expression for N_c in (16) becomes

$$N_c = \frac{T^{1-\delta_1}}{(W_d)^{\delta_1}} \frac{P^{1-\delta_2}}{(T_m)^{\delta_2}\mathsf{SNR}^{1-\delta_2}} \tag{22}$$

Equating (17) with (22) leads to the following canonical relationship

$$T = \frac{(k)^{\frac{1}{1-\delta_1}} \left(T_m^{\delta_2} W_d^{\delta_1}\right)^{\frac{1}{1-\delta_1}} W^{\frac{\mu-1+\delta_2}{1-\delta_1}}}{P^{\frac{\mu}{1-\delta_1}}} \tag{23}$$

that relates the signaling parameters (T,W,P), as a function of the channel parameters $(T_m,W_d,\delta_1,\delta_2)$ in order for the relationship (17) to hold between N_c and $\mathsf{SNR} = P/W$. Equations (17) and (23) are the two key equations that capture the essence of the results in this paper.

A. Discussion of Results on Ergodic Capacity

In the context of existing results in [2] that assume rich multipath ($\delta_1 = \delta_2 = 1$), Theorem 1 shows that the requirement on T_{coh} is now the requirement on the coherence dimension $N_c = T_{coh}W_{coh}$. Thus, the coherence cost is shared in both time and frequency resulting in *significantly weakened* scaling requirements for T_{coh}. If we have $W_{coh} = \mathcal{O}\left(W^{1-\delta_2}\right)$, then the T_{coh} scaling requirement reduces to

$$T_{coh} = N_c/W_{coh} = \mathcal{O}\left(W^{2+\delta_2}\right) \tag{24}$$

to achieve second-order optimality. This is significantly less stringent than the $T_{coh} = \mathcal{O}\left(W^3\right)$ required in the framework of [2].

Combining Theorem 1 with (23) lead to scaling rules for the locus of points (T,W,P) in order to achieve a desired value of μ (Recall $\mu > 1$ for first-order optimality and $\mu > 3$ for second-order optimality). Specifically,

$$\log(T) = \frac{1}{1-\delta_1}\log\left(W_d^{\delta_1}T_m^{\delta_2}\right) + \left(\frac{\mu+\delta_2-1}{1-\delta_1}\right)\log(W) - \left(\frac{\mu}{1-\delta_1}\right)\log(P) . \tag{25}$$

It is observed that smaller δ_i's imply a slower scaling of T with W. Conversely, for any system operating at a particular T and W, (25) can be used to determine the effective value of μ as

$$\mu_{\text{eff}} = \frac{(1-\delta_1)\log(T/c) + (1-\delta_2)\log(P)}{\log(W/P)} + (1-\delta_2) \tag{26}$$

where $c = \left(T_m^{\delta_2} W_d^{\delta_1}\right)^{\frac{1}{1-\delta_1}}$.

Note that $\mu_{\text{eff}} \to \infty$ as $T \to \infty$ for sparse channels, which implies that first- and second-order optimality can be achieved by simply increasing T. This is due to the impact of sparsity in Doppler and in direct contrast to the case of rich multipath where the coherence requirement is independent of signaling duration. We provide numerical illustration of the results by considering the low SNR asymptote of the coherent capacity in (19). The coefficients of the first- and second-order terms are $\lambda_1 = \log_2(e)$ and $\lambda_2 = -\log_2(e)$, respectively. In Fig. 2, we plot the numerically estimated values c_1 and c_2 of λ_1 and λ_2, respectively, for the training-based communication scheme, which are estimated using Monte-Carlo simulations. We observe that for a large enough T such that $\mu_{eff} > 3$, the second-order constant $c_2 \to \lambda_2 = -\log_2(e)$. Also shown in the figure is the behavior of the first-order constant and it is seen that $c_1 \to \lambda_1$ for a much smaller value of T since all we need is $\mu > 1$.

Contrary to the traditional emphasis on peaky signaling to improve the spectral efficiency of non-coherent communication, our results imply that delay-Doppler sparsity, along with a suitable choice of T, W and P as in (25) is sufficient to achieve a desired level of coherence with *non-peaky* signaling schemes. It is shown in [11] that the T_{coh} requirements for non-peaky signals under our framework are still better than those for peaky training-based communication schemes proposed in [2] based on a rich multipath assumption.
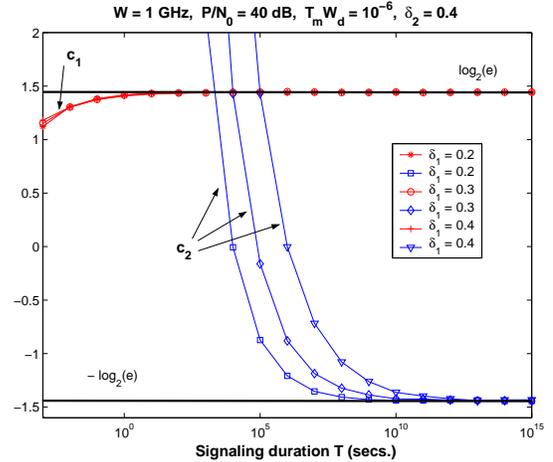

Fig. 2. *Convergence of the coefficients of the* SNR *and* SNR^2 *terms in capacity as a function of* T.

IV. RELIABILITY OF SPARSE MULTIPATH CHANNELS

The reliability function of the channel is defined as [7]

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

where $P_e(N,R)$ is the average probability of error over an ensemble of codes (random coding) in which each codeword spans the signal space dimensions $N = TW$ and communication takes place at transmission rate R. For any finite N, while the random coding exponent $E_r(N,R)$ provides a lower bound to $E(R)$, the sphere-packing exponent $E_{sp}(N,R)$ is an upper bound to $E(R)$ [7]. We recall the random coding upper bound on P_e [7] given by

$$P_e \leq e^{-N[E_r(N,R)]} \tag{33}$$

$$E_r(N,R) = \max_{0 \leq \rho \leq 1} \max_Q [E_o(N,\rho,Q) - \rho R] \tag{34}$$

$$E_r^{tr}(N,R) = \begin{cases} \frac{1}{N_c}\log\left(1 + \frac{(N_c-1)K^*\left(1-(K^*)^{1-\epsilon}\right)}{2}\right) - R - o(1) & 0 \le R \le R_{cr} \\ \frac{1}{N_c}\log\left(1 + \frac{(N_c-1)K^*\left(1-(K^*)^{1-\epsilon}\right)\rho^*}{1+\rho^*}\right) - \rho^* R - o(1) & R_{cr} \le R \le R_{max} \\ 0 & R > R_{max} \end{cases} \tag{27}$$

$$R_{cr} = \left(\frac{N_c-1}{2N_c}\right)\frac{K^*\left(1-(K^*)^{1-\epsilon}\right)}{\left[2+(N_c-1)K^*\left(1-(K^*)^{1-\epsilon}\right)\right]} \tag{28}$$

$$R_{max} = \left(\frac{N_c-1}{N_c}\right)K^*\left(1-(K^*)^{1-\epsilon}\right) \tag{29}$$

$$K^* = \frac{\eta^*\left(1-\eta^*\right)(N_c\mathsf{SNR})^2}{(N_c-1)(1+\eta^*N_c\mathsf{SNR})+(1-\eta^*)N_c\mathsf{SNR}} \tag{30}$$

$$\eta^* = \frac{N_c\mathsf{SNR}+N_c-1}{(N_c-2)N_c\mathsf{SNR}}\left[\sqrt{1+\frac{(N_c-2)N_c\mathsf{SNR}}{N_c\mathsf{SNR}+N_c-1}}-1\right] \tag{31}$$

$$\rho^* = \frac{-(2+k_1)+\sqrt{(2+k_1)^2+4(1+k_1)(\frac{k_1}{N_cR}-1)}}{2(1+k_1)} \qquad k_1 = (N_c-1)K^*\left(1-(K^*)^{1-\epsilon}\right) \tag{32}$$

$$E_o(N,\rho,Q) =$$
$$-\frac{1}{N}\log\left(E_\mathbf{H}\left[\int_\mathbf{y}\left[\int_\mathbf{x} q(\mathbf{x})p(\mathbf{y}|\mathbf{x},\mathbf{H})^{\frac{1}{1+\rho}}dx\right]^{1+\rho}dy\right]\right) \tag{35}$$

We compute the random coding error exponent in (34) for the training and communication scheme described in Sec. II-C. The result 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 (27) on the next page. R_{max} in (29) defines the maximum rate until which we have a non-zero error exponent (decaying P_e). The critical rate, R_{cr} in (28) delineates the regime of the optimal parameter ρ^* that maximizes the exponent. $\rho^* = 1$ for $0 < R < R_{cr}$ and ρ^* is given in (32) for $R_{cr} < R < R_{max}$. The constant $\epsilon > 0$ and is chosen very small ($\epsilon \to 0$) so that the $o(1)$ terms are negligible. See [13] for more details.

Note that the error exponent of the training and communication scheme in (27) depends *only* on SNR and $N_c = \frac{k}{\mathsf{SNR}^\mu}$.

A. Discussion of Results on Reliability

We investigate the behavior of the random coding exponent for different values of μ as illustrated in Fig. 3 for the given channel parameter set. It is observed that for any transmission rate R, there exists an optimum value of $\mu = \mu_{opt}(R)$ for which the error exponent in (27) is maximum. For any N, we formally define

$$\mu_{opt}(N,R) = \arg\max_\mu\left[E_r(R,N,\mu)\right] \tag{36}$$

where we have written $E_r(R,N,\mu) = \left[E_r^{tr}(N,R)\right]$ in (27) explicitly as a function of μ to emphasize its dependance. As we traverse from $R = 0$ to $R = C_{coh}$ (as in (18)), the optimal operating point at each rate is dictated by the value of μ_{opt} in (36) and can be *achieved* by choosing T, W and P as in (23). Furthermore, the optimizing μ_{opt} increases monotonically as we consider larger transmission rates. In fact, using the results on capacity from Theorem 1, it follows that with $\mu = 1$, we only obtain first-order optimality and therefore the error exponent in Fig. 3 is non-zero only for a fraction of the coherent capacity, C_{coh}. On the other hand, at $R = C_{coh}$, we would require $\mu_{opt} > 3$ (second-order optimal) in order to achieve a positive error exponent.

In Fig. 4, we plot the error exponent in (27) as a function of the parameter μ for two different transmission rates. For each scenario, we observe that the error exponent is concave as a function of

μ and is maximized at $\mu = \mu_{opt}$. Also illustrated in the figure is the error exponent with perfect CSI at the receiver that is an upper bound to $E_r^{tr}(N,R)$ in (27) and decreases monotonically with μ. These plots reveal a fundamental *learnability versus diversity tradeoff* in sparse channels. For any rate R, when $\mu < \mu_{opt}(R)$ (too little coherence), the system is in the learnability-limited regime and the error exponent of the training-based communication scheme is smaller due to poor channel estimation performance. On the other hand when $\mu > \mu_{opt}(R)$ (too much coherence), we are in the diversity-limited regime and the hit taken by the error exponent here is due to the inherent reduction in the degrees of freedom (DoF) (or delay-Doppler diversity, D). The best exponent is obtained at $\mu = \mu_{opt}(R)$, which demarcates the two regimes and describes the optimal tradeoff.
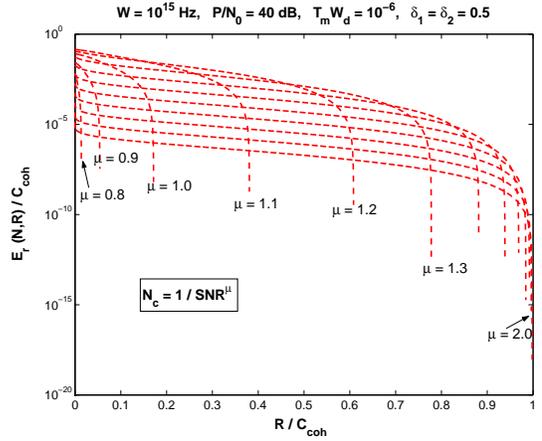

Fig. 3. *Random coding error exponent versus rate for a sparse channel and for the training-based communication scheme. Different curves correspond to different values of μ in the key relationship $N_c = \frac{1}{\mathsf{SNR}^\mu}$.*

References

[1] S. Verdú, "Spectral Efficiency in the Wideband Regime," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.

[2] L. Zheng, M. Medard, and D. N. C. Tse, "Channel Coherence in the Low SNR Regime," *Submitted, IEEE Trans. Inform. Theory*, 2005.

[3] A. F. Molisch, "Ultrawideband Propagation Channels - Theory, Measurement and Modeling," *IEEE Trans. Veh. Tech.*, vol. 54, no. 5, pp. 1528–1545, Sept. 2005.
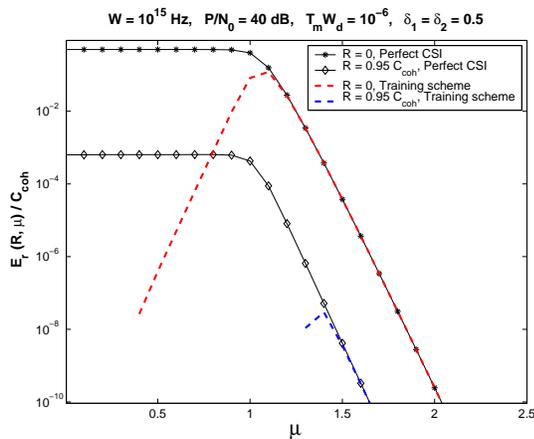
Fig. 4. *Illustration of the learnability versus diversity tradeoff for sparse channels. The value of μ at which the maximum is attained in each case defines $\mu_{opt}(R)$ at the corresponding transmission rate R.*

[4] R. Saadane, D. Aboutajdine, A. M. Hayar, and R. Knopp, "On the Estimation of the Degrees of Freedom of Indoor UWB Channel," *IEEE Veh. Tech. Conf. (Spring)*, 2005.

[5] A. M. Sayeed and B. Aazhang, "Joint Multipath-Doppler Diversity in Mobile Wireless Communications," *IEEE Trans. Commun.*, pp. 123–132, Jan. 1999.

[6] R. S. Kennedy, *Fading Dispersive Communication Channels*, Wiley InterScience, NY, 1969.

[7] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley and Sons Inc., 1968.

[8] A. M. Sayeed and V. V. Veeravalli, "Essential Degrees of Freedom in Space-Time Fading Channels," *Proc. 13th IEEE Intern. Symp. Pers. Indoor, Mobile Radio Commun.*, vol. 4, pp. 1512–1516, Sept. 2002.

[9] K. Liu, T. Kadous, and A. M. Sayeed, "Orthogonal Time-Frequency Signaling over Doubly Dispersive Channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 11, pp. 2583–2603, Nov. 2004.

[10] 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, pp. 323-352, 1997.

[11] V. Raghavan, G. Hariharan, and A. M. Sayeed, "Capacity of Sparse Multipath Channels in the Ultra-Wideband Regime," *Submitted, IEEE Journal of Selected Topics in Signal Proc.*, Dec 2006.

[12] B. Hassibi and B. Hochwald, "How Much Training is Needed in a Multiple Antenna Wireless Link?," *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951–964, Apr. 2003.

[13] G. Hariharan and A. M. Sayeed, "Minimum Probability of Error in Sparse Wideband Channels," *44th Annual Allerton Conference on Communication, Control and Computing*, Sep. 2006.