Non-Coherent Wideband Capacity of Sparse Multipath Channels with Limited Feedback

Gautham Hariharan and Akbar M. Sayeed

Abstract— **This paper investigates the ergodic capacity of multipath channels in the wideband regime with limited feedback. Our work builds on recent results that have established significant capacity gains achievable in the wideband/low-SNR regime, when there is perfect channel state information (CSI) at the transmitter. Furthermore, this benchmark gain can be obtained with just 1-bit of CSI at the transmitter about each channel coefficient. However, the capacity achieving signals are peaky; they have large instantaneous transmit power, especially in the non-coherent scenario. Signal peakiness is related to channel coherence and, in contrast to the prevalent assumption of rich multipath, we investigate the capacity of sparse multipath channels, motivated by recent experimental results. Sparsity naturally leads to coherence in time and frequency. With perfect receiver CSI, it is shown that limited feedback, *even* with an instantaneous power constraint, is sufficient to achieve the benchmark capacity. Our analysis reveals the benefits of channel sparsity in the non-coherent scenario, where we employ a training-based communication scheme. With an average power constraint, it is shown that the benchmark is achievable, provided the channel coherence scales at a sufficiently fast rate with signal space dimension. Furthermore, in contrast to peaky signaling schemes that violate instantaneous power constraints, we show that the benchmark is attainable in sparse channels with finite instantaneous transmit power. We present rules of thumb on choosing the signaling parameters as a function of the channel parameters so that the full benefits of sparsity can be realized.**

I. Introduction

Recent research on the fundamental performance limits of wideband/low-SNR communication has particularly focused on the impact of channel state information (CSI), more specifically the non-coherent regime, when there is no CSI at the receiver *a priori*. This is because, from a capacity perspective, spreading signals have been shown [1] to be sub-optimal and peaky or flash signaling schemes [2], [3] are necessary to achieve non-coherent wideband capacity. However these results have been derived based on an implicit assumption of rich multipath, in which the independent degrees of freedom (DoF) in the channel scale linearly with bandwidth. Recent work by Zheng *et al* [4] has emphasized the crucial role of channel coherence in the low SNR regime and the importance of channel learning schemes that can bridge the gap between the coherent and non-coherent extremes.

Motivated by these works as well as by recent measurement campaigns for UWB [5], [6], we recently introduced the notion of *multipath sparsity* as a physical source of channel coherence and proposed a channel modeling framework in [7] that captures the effect of sparsity in delay and Doppler. A key implication of sparsity is that the degrees of freedom (DoF) in the wideband channel scale sub-linearly with the signal space dimension (time-bandwidth product). Based on this model, we investigated the ergodic capacity of training-based communication schemes. The analysis in [7] reveals the impact of channel sparsity on channel coherence scaling and the role played by sparsity in reducing/eliminating peaky signaling to achieve wideband capacity.

Building on the results in [7], the focus of this paper is on the impact of *feedback* on the ergodic capacity of sparse wideband channels. Although earlier works (for example [8], [9], [10] and references therein) have explored capacity with transmitter CSI, it is only recently [11], [12] that the impact of feedback in the low-SNR, non-coherent regime has received attention. In particular, it is shown in [11] that with an average power constraint, the capacity gain with perfect transmitter and receiver CSI (over the case when there is only receiver CSI) equals $\log\left(\frac{1}{\mathsf{SNR}}\right)$ and is obtained with the well known water-filling solution [8]. More interestingly, it is shown that this gain can actually be achieved with limited feedback; when there is just 1 bit of CSI about each channel coefficient at the transmitter and on-off signaling is employed. However, for both water-filling and on-off signaling, the capacity achieving input tends to be peaky (or) bursty in time, leading to a high peak-to-average power ratio, and difficulties from an implementation standpoint. The peakiness aspect is much more relevant in the non-coherent scenario, for which [11] proposes peaky training and communication. The necessity of peaky training is tied in with the need to reliably estimate the channel at the receiver.

We analyze the capacity of sparse wideband channels with limited feedback as in [11], [12]. We start with a brief description in Section II of the sparse channel model [7]. Our main focus is on the case when there is no receiver CSI *a priori* and training-based communication is employed. The analysis is performed under an average or long-term power constraint as well as with an instantaneous or short-term constraint. We restrict our attention to *causal* signaling schemes that are realizable in practice. We first analyze in Section III, the perfect receiver CSI scenario. It is shown that an optimal threshold given by $\mathsf{h}_\mathsf{t} = \lambda \log\left(\frac{1}{\mathsf{SNR}}\right)$ for any $\lambda \in (0,1)$ directly provides a measure of capacity which behaves as $(1+\mathsf{h}_\mathsf{t})\,\mathsf{SNR}$ in the wideband limit. Thus with $\lambda \to 1$, we achieve the perfect transmitter CSI capacity, which is the benchmark for all limited feedback schemes. We derive a sufficient condition under which this benchmark can be approached even with an instantaneous power constraint. A key parameter that determines this condition is $\mathbf{E}\left[D_{\text{eff}}\right]$, the

This work was supported in part by the National Science Foundation through grant #CCF-0431088. The authors are with the University of Wisconsin-Madison, Madison WI 53706, USA. E-mail: gauthamh@cae.wisc.edu, akbar@engr.wisc.edu.

average number of "active" independent channel dimensions, the number of independent channel coefficients that exceed the threshold in the power allocation scheme. In particular, with an instantaneous power constraint, the benchmark capacity gain is achieved when $\mathbf{E}[D_{\text{eff}}] \to \infty$ as $\mathsf{SNR} \to 0$ and we discuss its feasibility when the channel is rich and when it is sparse.

In Section IV, we discuss the capacity of training-based communication schemes. With an average power constraint, it is shown that as long as the channel coherence dimension N_c scales with SNR as $N_c = \frac{1}{\mathsf{SNR}^\mu}$ for some $\mu > 1$, the capacity of the training-based scheme converges to the coherent capacity, the performance benchmark, in the wideband limit. Furthermore, this condition is achievable only when the channel is sparse and we provide guidelines on choosing the signal space parameters (signaling duration, bandwidth and transmit power) so that $\mu > 1$ is satisfied. The critical role of channel sparsity is further revealed when we impose an instantaneous power constraint. In contrast to peaky signaling that violates finite constraints on the instantaneous power, channel sparsity is sufficient to achieve both $\mu > 1$ *and* $\mathbf{E}[D_{\text{eff}}] \to \infty$ and helps attain the benchmark with both average and instantaneous power constraints.

II. System Model

In this section, we briefly summarize the model developed in [7] for sparse multipath channels. Our results are based on an orthogonal short-time Fourier (STF) signaling framework [13], [14] that naturally relates multipath sparsity in delay-Doppler to coherence in time and frequency.

A. Sparse Multipath Channel Modeling

A physical discrete multipath channel can be modeled as

$$\begin{aligned} h(\tau,\nu) &= \sum_n \beta_n \delta(\tau-\tau_n)\delta(\nu-\nu_n) \\ y(t) &= \sum_n \beta_n x(t-\tau_n) e^{j2\pi\nu_n t} + w(t) \end{aligned} \tag{1}$$

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 denote the delay and Doppler spreads, respectively. The quantities $x(t), y(t)$ and $w(t)$ denote the transmitted, received and additive white Gaussian noise waveforms, respectively. Throughout this paper, we assume an *underspread* channel: $T_m W_d \ll 1$.

We use a *virtual representation* [15], [16] of the physical model in (1) 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 [15], [16]:

$$y(t) = \sum_{\ell=0}^{L} \sum_{m=-M}^{M} h_{\ell,m} x(t-\ell/W) e^{j2\pi mt/T} + w(t) \tag{2}$$

$$h_{\ell,m} \approx \sum_{n \in S_{\tau,\ell} \cap S_{\nu,m}} \beta_n. \tag{3}$$

where $L = \lceil T_m W \rceil$ and $M = \lceil T W_d/2 \rceil$. The sampled representation (2) is linear and is characterized by the virtual delay-Doppler channel coefficients $\{h_{\ell,m}\}$ in (3). 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 $S_{\tau,\ell} \cap S_{\nu,m}$ of size $\Delta\tau \times \Delta\nu$, $\Delta\tau = \frac{1}{W}, \Delta\nu = \frac{1}{T}$ as shown in Fig. 1(a). Distinct $h_{\ell,m}$'s correspond to approximately *disjoint* subsets of paths and are hence approximately statistically independent. 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.

Let D denote the number of dominant non-zero channel coefficients; it reflects the (dominant) statistically independent degrees of freedom (DoF) in the channel and also signifies the delay-Doppler diversity afforded by the channel [15]. We decompose D as $D = D_T D_W$ where D_T denotes the Doppler/time diversity and D_W the frequency/delay diversity. The channel DoF or delay-Doppler diversity is bounded as:

$$\begin{aligned} D &= D_T D_W \le D_{\max} = D_{T,\max} D_{W,\max} \\ D_{T,\max} &= \lceil T W_d \rceil \ , \ D_{W,\max} = \lceil T_m W \rceil \end{aligned} \tag{4}$$

where $D_{T,\max}$ denotes the maximum Doppler diversity and $D_{W,\max}$ denotes maximum delay diversity. Note that $D_{T,\max}$ and $D_{W,\max}$ increase linearly with T and W, respectively, and represent a *rich multipath* environment in which each resolution bin in Fig. 1(a) corresponds to a dominant channel coefficient.

However, there is growing experimental evidence [5], [17], [6] that the dominant channel coefficients get sparser in delay as the bandwidth increases. Furthermore, we are also interested in modeling scenarios with Doppler effects, due to motion. In such cases, as we consider large bandwidths and/or long signaling durations, the resolution of paths in both delay 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. $D < D_{\max}$.

We model multipath sparsity with a *sub-linear* scaling in D_T and D_W with T and W:

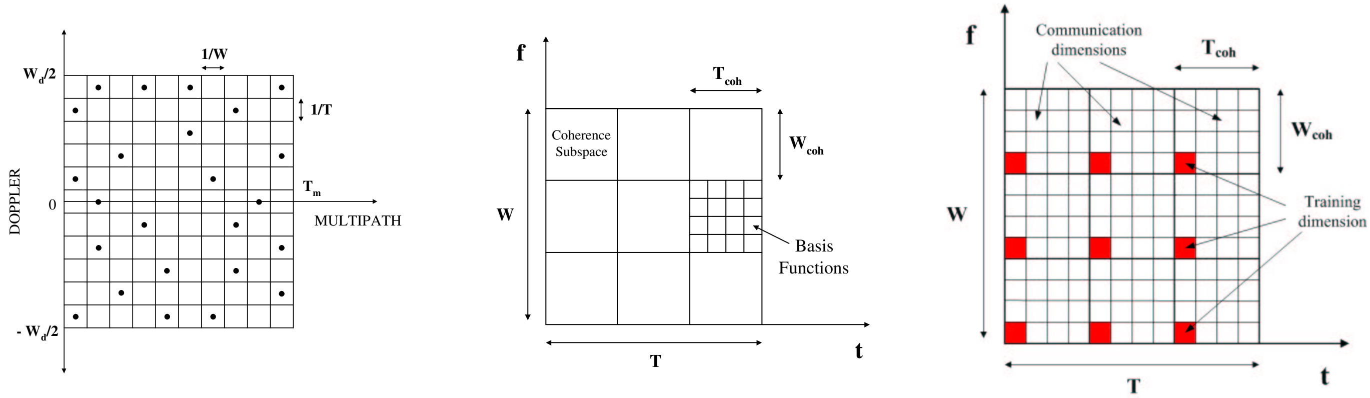
$$D_W \sim g_1(W) \ , \ D_T \sim g_2(T) \tag{5}$$

where g_1 and g_2 are *arbitrary* sub-linear functions. As a concrete example, we will focus on a specific power-law scaling for the rest of this paper:

$$D_T = \frac{T^{\delta_1}}{W_d^{\delta_1}}, \quad D_W = \frac{W^{\delta_2}}{T_m^{\delta_2}} \tag{6}$$

for $\delta_1, \delta_2 \in (0,1)$. But the results derived here hold true in general for any sub-linear scaling law. Note that (4) and (5) imply that the total number of delay-Doppler DoF, $D = D_T D_W$, scales *sub-linearly* with the signal space dimension $N = TW$ in sparse multipath.

Remark 1: With perfect CSI at the receiver, the parameter D denotes the delay-Doppler diversity afforded by the channel, whereas with no CSI, it reflects the level of channel uncertainty; the number of channel parameters that need to be estimated at the receiver for coherent processing.



(a) (b) (c)

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) represents the training dimension and the remaining dimensions are used for communication.*

B. Orthogonal Short-Time Fourier Signaling

We consider signaling using an orthonormal short-time Fourier (STF) basis [13], [14] that is a natural generalization[1] of orthogonal frequency-division multiplexing (OFDM) for time-varying channels. An orthogonal STF basis $\{\phi_{\ell m}(t)\}$ for the signal space is generated from a fixed prototype waveform $g(t)$ via time and frequency shifts: $\phi_{\ell m}(t) = g(t - \ell T_o)e^{j2\pi W_o t}$, where $T_o W_o = 1$, $\ell = 0, \cdots, N_T - 1$, $m = 0, \cdots, N_W - 1$ and $N = N_T N_W = TW$ with $N_T = T/T_o, N_W = W/W_o$. The transmitted signal can be represented as

$$x(t) = \sum_{\ell=0}^{N_T-1} \sum_{m=0}^{N_W-1} x_{\ell m}\phi_{\ell m}(t) \qquad 0 \le t \le T \tag{7}$$

where $\{x_{\ell m}\}$ represent the N transmitted symbols that are modulated onto the STF basis waveforms. The received signal is projected onto the STF basis waveforms to yield

$$y_{\ell m} = \langle y, \phi_{\ell m}\rangle = \sum_{\ell', m'} h_{\ell m, \ell' m'} \;\; x_{\ell' m'} + w_{\ell m}. \tag{8}$$

We can represent the system using an N-dimensional matrix equation

$$\mathbf{y} = \;\mathbf{H}\mathbf{x} + \mathbf{w} \tag{9}$$

where $\mathbf{w}$ represents the additive noise vector whose entries are i.i.d. $CN(0, 1)$. The $N \times N$ matrix $\mathbf{H}$ consists of the channel coefficients $\{h_{\ell m, \ell' m'}\}$ in (8). We assume that the input symbols that form the transmit codeword $\mathbf{x}$ satisfy an average power constraint

$$\frac{1}{T} \cdot \mathbf{E}\left[\|\mathbf{x}\|^2\right] \le P \tag{10}$$

Since there are $N = TW$ symbols per codeword, we define the parameter SNR (transmit energy per modulated symbol) for a given average transmit power P as $\mathsf{SNR} = \frac{TP}{TW} = \frac{P}{W}$. In this work, the focus is on the wideband regime where $\mathsf{SNR} \to 0$ as $W \to \infty$ for a fixed P.

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 [14], [13]; that is, (8) simplifies to[2] $y_{\ell m} \approx h_{\ell m} x_{\ell m} + w_{\ell m}$. Thus the channel matrix $\mathbf{H}$ is approximately diagonal. In this work, we assume that $\mathbf{H}$ is exactly diagonal; that is,

$$\mathbf{H} = \text{diag}\Big[\underbrace{h_{11}\cdots h_{1N_c}}_{\text{Subspace } 1},\; \underbrace{h_{21}\cdots h_{2N_c}}_{\text{Subspace } 2}\;\cdots\;\underbrace{h_{D1}\cdots h_{DN_c}}_{\text{Subspace } D}\Big]. \tag{11}$$

The diagonal entries of $\mathbf{H}$ in (11) admit an intuitive block fading interpretation in terms of *time-frequency coherence subspaces* [13] illustrated in Fig. 1(b). The signal space is partitioned as $N = TW = N_c D$ where D represents the number of statistically independent time-frequency coherence subspaces, reflecting the DoF in the channel, and N_c represents the dimension of each coherence subspace, which we refer to as the **coherence dimension**. In the block fading model in (11), the channel coefficients over the i-th coherence subspace $h_{i1}, \cdots, h_{iN_c}$ are assumed to be identical (denoted by h_i), whereas the coefficients across different coherence subspaces are independent and identically distributed. Thus, the channel is characterized by the D distinct STF channel coefficients, $\{h_i\}$, that are i.i.d. zero-mean Gaussian random variables (Rayleigh fading) with (normalized) variance equal to $\mathbf{E}[|h_i|^2] = \sum_n \mathbf{E}[|\beta_n|^2] = 1$ [13].

Using the DoF scaling for sparse channels in (5), the scaling behavior for the coherence dimension can be computed as

$$\begin{aligned} W_{coh} &= \frac{W}{D_W} \sim f_1(W), \quad T_{coh} = \frac{T}{D_T} \sim f_2(T) \\ N_c &= W_{coh}T_{coh} \sim f_1(W)f_2(T) \end{aligned} \tag{12}$$

where T_{coh} is the *coherence time* and W_{coh} is the *coherence bandwidth* of the channel, as illustrated in Fig. 1(b). As a consequence of the sub-linearity of g_1 and g_2 in (5), f_1 and f_2 are also sub-linear. In particular, corresponding to the power-

[1] STF signaling can be considered as OFDM signaling over a block of OFDM symbol periods and with an appropriately chosen OFDM symbol duration.

[2] The STF channel coefficients are different from the delay-Doppler coefficients, even though we are using the same symbols.

law scaling in (6), we obtain

$$T_{coh} = \frac{T^{1-\delta_1}}{W_d^{\delta_2}}, \quad W_{coh} = \frac{W^{1-\delta_2}}{T_m^{\delta_1}} \tag{13}$$

***Remark** 2:* Note that when the channel is sparse, both N_c and D increase sub-linearly with N, whereas when the channel is rich, D scales linearly with N, while N_c is fixed.

In this work, our focus is on computing non-coherent channel capacity with feedback and as we will see later in Sections III and IV, capacity turns out to be a function only of the parameters N_c and SNR. Thus, in order to analyze the low SNR asymptotics, the following relation between N_c and SNR $(= P/W)$ plays a key role:

$$N_c = \frac{1}{\mathsf{SNR}^{\mu}} \ , \quad \mu > 0 \tag{14}$$

where the parameter μ reflects the level of channel coherence. We will revisit (14) and discuss its achievability and implications in Section IV.

III. Capacity with Perfect Receiver CSI and Limited Feedback

In this section, we study the scenario when there is perfect CSI at the receiver. We assume throughout this paper that both the transmitter and the receiver have statistical CSI - knowledge of T_m, W_d, g_1, g_2, f_1 and f_2 so that the scaling in D and N_c is known. On the one extreme, with perfect receiver CSI and no transmitter CSI (no feedback), the coherent capacity per dimension (in b/s/Hz) equals

$$C_{\mathsf{coh},0}(\mathsf{SNR}) = \sup_{\mathbf{Q}:\ \mathrm{Tr}(\mathbf{Q})\ \leq\ TP} \frac{\mathbf{E}\left[\log_2 \det\left(\mathbf{I}_{N_c D} + \mathbf{H}\mathbf{Q}\mathbf{H}^H\right)\right]}{N_c D} \tag{15}$$

The optimization is over the set of $N_c D$-dimensional positive definite input covariance matrices $\mathbf{Q} = \mathbf{E}\left[\mathbf{x}\mathbf{x}^H\right]$ satisfying the average power constraint in (10). Due to the diagonal nature of $\mathbf{H}$ in (11), the optimal $\mathbf{Q}$ is also diagonal. Furthermore, with no transmitter CSI, the uniform power allocation $\mathbf{Q} = \frac{TP}{N_c D}\mathbf{I}_{N_c D} = \mathsf{SNR}\,\mathbf{I}_{N_c D}$ achieves this optimum. The corresponding capacity in the limit of low SNR equals [7]

$$C_{\mathsf{coh},0}(\mathsf{SNR}) = \log_2(e) \cdot \left[\mathsf{SNR} - \mathsf{SNR}^2\right] \tag{16}$$

On the other extreme is the case of perfect receiver and transmitter CSI, where the receiver instantaneously feeds back all the channel coefficients, $\{h_i\}_{i=1}^D$, to the transmitter, corresponding to the D independent coherence subspaces. The optimum transmitter power allocation in this case is water-filling [8] over the different coherence subspaces. In the low SNR extreme, it is shown in [11] that the capacity with perfect transmitter CSI scales as $\log\left(\frac{1}{\mathsf{SNR}}\right)\mathsf{SNR}$. Thus the capacity gain compared to receiver CSI only is directly proportional to the water-filling threshold, $h_w \sim \log\left(\frac{1}{\mathsf{SNR}}\right)$, and this gain serves as a benchmark for all limited feedback schemes. More interestingly, it is shown in [11] that this maximum capacity gain can be achieved with just one bit of feedback per channel coefficient.

In this case of limited feedback, both the transmitter and the receiver have *a priori* knowledge of a common threshold, $\mathsf{h_t}$. The receiver compares the channel strength ($|h_i|^2$, $i = 1, 2, \cdots, D$) in each coherence subspace with $\mathsf{h_t}$, and feeds back

$$b_i = \begin{cases} 1 & \text{if} \quad |h_i|^2 \geq \mathsf{h_t} \\ 0 & \text{if} \quad |h_i|^2 < \mathsf{h_t}. \end{cases} \tag{17}$$

At the transmitter, power allocation is uniform across the coherence subspaces for which $b_i = 1$ and no power is allocated to those subspaces for which $b_i = 0$. Conditioned on the $\{b_i\}_{i=1}^D$, the input power allocation, which we still denote by $\mathbf{Q}$ with a little abuse of notation, takes the form

$$\begin{aligned} \mathbf{Q} &= \mathrm{diag}\left(|x_1|^2, |x_2|^2, \cdots, |x_N|^2\right) \\ &= \mathrm{diag}\big(\underbrace{q_1, \cdots, q_1}_{N_c}, \underbrace{q_2, \cdots, q_2}_{N_c}, \cdots, \underbrace{q_D, \cdots, q_D}_{N_c}\big) \\ q_i &= P_{\mathrm{o}} \cdot \chi(|h_i|^2 \geq \mathsf{h_t}). \end{aligned} \tag{18}$$

The choice of P_{o} depends on the type of power constraint and also on the nature of feedback. To explore this further, let D_{eff} denote the number of active subspaces, those which exceed the threshold $\mathsf{h_t}$. We have

$$D_{\mathrm{eff}} = \sum_{i=1}^{D} \chi(|h_i|^2 \geq \mathsf{h_t}) \tag{19}$$

$$\mathbf{E}\left[D_{\mathrm{eff}}\right] \overset{(a)}{=} D\mathbf{E}\left[\chi(|h|^2 \geq \mathsf{h_t})\right] \overset{(b)}{=} De^{-\mathsf{h_t}} \tag{20}$$

where (a) is due to the fact that $\{h_i\}_{i=1}^D$ are i.i.d. and (b) is because for a standard Gaussian h_i, $\mathbf{E}\left[\chi(|h_i|^2 \geq \mathsf{h_t})\right] = \Pr\left(|h_i|^2 \geq \mathsf{h_t}\right) = e^{-\mathsf{h_t}}$.

If we assume knowledge of all $\{b_i\}_{i=1}^D$ at the beginning of each codeword, then we can uniformly divide power among the active subspaces. That is

$$P_{\mathrm{o},\mathsf{nc}} = \frac{TP}{N_c D_{\mathrm{eff}}}. \tag{21}$$

Using (15) and (18), the capacity with this power allocation equals

$$C_{\mathsf{coh},1,LT}(\mathsf{SNR}) = \max_{\mathsf{h_t}} \frac{1}{D}\sum_{i=1}^{D} \mathbf{E}\left[\log_2\left(1 + \frac{TP}{N_c D_{\mathrm{eff}}} \cdot |h_i|^2\right)\chi\left(|h_i|^2 \geq \mathsf{h_t}\right)\right]. \tag{22}$$

The power allocation in (21) satisfies the power constraint instantaneously as well as on average. To see this, note that

$$\begin{aligned} P_{\mathsf{inst,nc}} = \frac{1}{T}\|\mathbf{x}\|^2 &= \frac{N_c}{T}\sum_{i=1}^{D} q_i \\ &= \frac{N_c}{T}\sum_{i=1}^{D} \frac{TP}{N_c D_{\mathrm{eff}}}\chi\left(|h_i|^2 \geq \mathsf{h_t}\right) = P \end{aligned} \tag{23}$$

and clearly $\mathbf{E}\left[P_{\mathsf{inst,nc}}\right] \leq P$ as well. However, the above scheme is not realizable in practice since it is not causal. This is especially relevant in the more practical scenario when the receiver estimates the channel coefficients $\{h_i\}_{i=1}^D$ and feeds back $\{b_i\}_{i=1}^D$ based on these estimates. This motivates us to instead consider a causal power allocation scheme, one in which for all $i = 1, \cdots, D$, q_i in (18) depends only on b_i and $P_{\rm o}$ is independent of $\{b_i\}_{i=1}^{D}$. From (18), we have

$$\mathbf{E}\left[\|\mathbf{x}\|^2\right] = N_c \sum_{i=1}^{D} \mathbf{E}\left[q_i\right] = N_c \sum_{i=1}^{D} P_{\rm o} \cdot \mathbf{E}\left[\chi(|h_i|^2 \geq \mathsf{h_t})\right] \stackrel{(a)}{=} N_c P_{\rm o} \mathbf{E}\left[D_{\rm eff}\right] \tag{24}$$

where (a) follows from (20). Thus to satisfy $\mathbf{E}\left[\|\mathbf{x}\|^2\right] \leq TP$, the power allocation for this causal scheme is given by

$$P_{\rm o,c} = \frac{TP}{N_c \mathbf{E}\left[D_{\rm eff}\right]} = \frac{TP}{N_c D e^{-\mathsf{h_t}}} \tag{25}$$

and the corresponding capacity is given by

$$\widehat{C}_{\mathsf{coh},1,LT}(\mathsf{SNR}) = \max_{\mathsf{h_t}} \frac{1}{D} \sum_{i=1}^{D} \mathbf{E}\left[\log_2\left(1 + \frac{TP}{N_c D e^{-\mathsf{h_t}}}|h_i|^2\right)\chi(|h_i|^2 \geq \mathsf{h_t})\right]. \tag{26}$$

The causal power allocation policy in (25) satisfies the average power constraint but can have a large instantaneous power. We have

$$P_{\mathsf{inst,c}} = \frac{N_c}{T} \sum_{i=1}^{D} \frac{TP}{N_c D e^{-\mathsf{h_t}}} \chi\left(|h_i|^2 \geq \mathsf{h_t}\right) = \left(\frac{D_{\rm eff}}{D e^{-\mathsf{h_t}}}\right) P. \tag{27}$$

Thus $\mathbf{E}\left[P_{\mathsf{inst,c}}\right] \leq P$, but unlike (23), $P_{\mathsf{inst,c}} \in [0, \infty)$. We will address this important issue in Section III-B, but first, we solve the capacity problem in (26), considering only an average power constraint.

A. Capacity with Average Power Constraint

The following theorem establishes that a threshold of the form $\mathsf{h_t} \sim \lambda \log\left(\frac{1}{\mathsf{SNR}}\right)$ for some $\lambda \in (0,1)$ provides the solution to (26).

Theorem 1: Given any $\lambda \in (0,1)$, a causal signaling scheme satisfying the average power constraint achieves $\widehat{C}_{LB} \leq \widehat{C}_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR}) \leq \widehat{C}_{UB}$ as specified in (29) and (30) (on page (6)) using an optimal threshold satisfying

$$\lim_{\mathsf{SNR}\to 0} \frac{\mathsf{h_t}}{\lambda \log\left(\frac{1}{\mathsf{SNR}}\right)} = 1. \tag{28}$$

Proof: We start with (26)

$$\widehat{C}_{\mathsf{coh},1,LT}(\mathsf{SNR}) = \max_{\mathsf{h_t}} \frac{1}{D} \sum_{i=1}^{D} \mathbf{E}\left[\log_2\left(1 + \frac{TP}{N_c D e^{-\mathsf{h_t}}}|h_i|^2\right)\chi(|h_i|^2 \geq \mathsf{h_t})\right] \stackrel{(a)}{=} \mathbf{E}\left[\log_2\left(1 + \mathsf{SNR}\, e^{\mathsf{h_t}}\,|h|^2\right)\chi(|h|^2 \geq \mathsf{h_t})\right] \tag{31}$$

where (a) follows from the fact that $\{h_i\}$ are i.i.d. $CN(0,1)$.

Computing the expectation in (31) using [18, 4.337(1), p. 574] and defining $\alpha = \frac{1 + \mathsf{SNR}\,\mathsf{h_t}\,e^{\mathsf{h_t}}}{\mathsf{SNR}\,e^{\mathsf{h_t}}}$ for convenience, we have

$$\widehat{C}_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR}) = e^{-\mathsf{h_t}} \cdot \left[\log_2\left(1 + \mathsf{SNR}\,\mathsf{h_t}\,e^{\mathsf{h_t}}\right) + \log_2(e) \cdot \exp(\alpha) \int_{\alpha}^{\infty} \frac{e^{-t}}{t}\mathrm{d}t\right]$$

$$= e^{-\mathsf{h_t}} \cdot \left[\log_2\left(1 + \mathsf{SNR}\,\mathsf{h_t}\,e^{\mathsf{h_t}}\right) + \log_2(e) \cdot \nu_\alpha\right] \tag{32}$$

where we define

$$\nu_\alpha = \exp(\alpha) \int_{\alpha}^{\infty} \frac{e^{-t}}{t}\mathrm{d}t. \tag{33}$$

Furthermore, in the limit of $\alpha \to \infty$, we have the following bounds to ν_α [19, 5.1.20, p. 229]:

$$\frac{1}{2}\log_e\left(1 + \frac{2}{\alpha}\right) \leq \nu_\alpha \leq \log_e\left(1 + \frac{1}{\alpha}\right). \tag{34}$$

The choice of $\mathsf{h_t}$ that maximizes (32) is obtained by setting its derivative to zero and satisfies

$$\Delta \triangleq 1 - \log_e\left(1 + \mathsf{SNR}\mathsf{h_t}e^{\mathsf{h_t}}\right) - \frac{1}{\mathsf{SNR}e^{\mathsf{h_t}}} \cdot \nu_\alpha = 0. \tag{35}$$

Now if $\mathsf{h_t}$ is such that $\lim_{\mathsf{SNR}\to 0} \frac{\mathsf{h_t}}{\lambda \log\left(\frac{1}{\mathsf{SNR}}\right)} = 1$ for some $\lambda \in (0,1)$, then as $\mathsf{SNR} \to 0$, we have $\mathsf{SNR}\,\mathsf{h_t}\,e^{\mathsf{h_t}} \to 0$ and $\alpha \to \infty$. Thus using (34), $\nu_\alpha \approx \frac{1}{\alpha}$. Using this in (35), we have $\frac{1}{\mathsf{SNR}e^{\mathsf{h_t}}} \cdot \nu_\alpha \approx \frac{1}{1+\mathsf{SNR}\mathsf{h_t}e^{\mathsf{h_t}}} \to 1$. Therefore, with the choice of $\mathsf{h_t}$ as in (28), it follows that as $\mathsf{SNR} \to 0$, $\Delta \to 0$. Substituting this choice of $\mathsf{h_t}$ in (32) and using the upper and lower bounds on ν_α in (34), we obtain the bounds in (29) and (30). ■

It can also be shown that the rate achievable with this causal scheme is asymptotically (in low SNR) the same as the non-causal capacity in (22). That is, $\widehat{C}_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR})$ is a tight bound to $C_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR})$ and for all $\lambda \in (0,1)$, we have

$$\lim_{\mathsf{SNR}\to 0} \frac{\left|C_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR}) - \widehat{C}_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR})\right|}{C_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR})} = 0. \tag{36}$$

We omit the details here for lack of space but a rigorous derivation can be found in [20, Appendix A].

Corollary 1: The capacity gain for the D-bit feedback, causal power allocation scheme over the receiver CSI only capacity in (16) satisfies

$$\lim_{\mathsf{SNR}\to 0} \frac{\widehat{C}_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR})}{C_{\mathsf{coh},0}(\mathsf{SNR})} = (1 + \mathsf{h_t}) = \left(1 + \lambda \log\left(\frac{1}{\mathsf{SNR}}\right)\right). \tag{37}$$

Proof: By performing a Taylor series expansion of the upper and lower bounds in (29) and (30), we note that they are equal upto a first-order and obtain, $\widehat{C}_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR}) = \left[1 + \lambda \log\left(\frac{1}{\mathsf{SNR}}\right)\right]\mathsf{SNR} = (1 + \mathsf{h_t})\mathsf{SNR}$. On the other hand, with only receiver CSI, we have from (16), $C_{\mathsf{coh},0}(\mathsf{SNR}) = \mathsf{SNR}$. Thus the desired result follows. ■

Remark 3: The capacity gain due to feedback is directly proportional to $\mathsf{h_t}$ and the highest gain is obtained by choosing $\lambda \to 1$, and equals the perfect CSI benchmark.

We now revert our attention back to the instantaneous transmit power described in (27). Note that as $D \to \infty$, $P_{\mathsf{inst,c}} \to P$ as a consequence of the law of large numbers. However, for any large but finite D, $P_{\mathsf{inst,c}}$ may be much larger than P. This is a serious issue in practical systems that typically operate with peak power limitations. Thus it is important to analyze the impact of constraints on the instantaneous power in (27), as discussed next.

$$\widehat{C}_{UB} = \mathsf{SNR}^{\lambda} \cdot \left[\log_2\left(1+\lambda\mathsf{SNR}^{1-\lambda}\log\left(\tfrac{1}{\mathsf{SNR}}\right)\right) + \log_2\left(1+\frac{\mathsf{SNR}^{1-\lambda}}{1+\lambda\mathsf{SNR}^{1-\lambda}\log\left(\frac{1}{\mathsf{SNR}}\right)}\right)\right] \quad (29)$$

$$\widehat{C}_{LB} = \mathsf{SNR}^{\lambda} \cdot \left[\log_2\left(1+\lambda\mathsf{SNR}^{1-\lambda}\log\left(\tfrac{1}{\mathsf{SNR}}\right)\right) + \frac{1}{2}\log_2\left(1+\frac{2\mathsf{SNR}^{1-\lambda}}{1+\lambda\mathsf{SNR}^{1-\lambda}\log\left(\frac{1}{\mathsf{SNR}}\right)}\right)\right] \quad (30)$$

B. Capacity with Instantaneous Power Constraint

In addition to the average power constraint, we impose a constraint on the instantaneous transmit power of the form

$$P_{\mathsf{inst,c}} \overset{a.s.}{\leq} AP \quad (38)$$

where $A > 1$ and finite. With this short-term constraint, we calculate the capacity, $\widehat{C}_{\mathsf{coh},1,\mathsf{ST}}(\mathsf{SNR})$, of the causal signaling scheme. We are particularly interested in exploring conditions under which we do not take a hit in capacity and $\widehat{C}_{\mathsf{coh},1,\mathsf{ST}}(\mathsf{SNR}) \approx \widehat{C}_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR})$. To this end, we employ the following power allocation

$$\mathbf{Q} = \mathrm{diag}\big(\underbrace{q_1,\cdots,q_1}_{N_c}, \underbrace{q_2,\cdots,q_2}_{N_c}, \cdots, \underbrace{q_D,\cdots,q_D}_{N_c}\big)$$

$$q_i = P_{\mathsf{o,c}}\, \chi(|h_i|^2 \geq \mathsf{h_t})\, \chi\left(\sum_{j=1}^{i} \chi(|h_j|^2 \geq \mathsf{h_t}) \leq ADe^{-\mathsf{h_t}}\right). \quad (39)$$

The second indicator function in (39) checks for the constraint in (38) causally, during each time-frequency coherence slot, and allocates power only if this constraint is satisfied. The capacity of this scheme can be computed as shown in the set of equations (40) (on page (7)) where $\widehat{C}_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR})$ is the capacity of the causal signaling scheme in (26), with only an average power constraint, and (a) follows from the fact that $\{h_i\}$ are i.i.d. and $p_i \triangleq \Pr\left(\sum_{j=1}^{i} \chi(|h_j|^2 \geq \mathsf{h_t}) \leq ADe^{-\mathsf{h_t}}\right)$. Thus, characterizing $\widehat{C}_{\mathsf{coh},1,\mathsf{ST}}(\mathsf{SNR})$ is equivalent to characterizing p_i. In particular, under what condition does $\frac{\sum_{i=1}^{D} p_i}{D} \to 1$? This is discussed in the following proposition.

Proposition 1: With $\mathsf{h_t} \sim \lambda\log\left(\frac{1}{\mathsf{SNR}}\right)$ as in (28), $\frac{\sum_{i=1}^{D} p_i}{D} \geq L$, where for $1 < A < 2$, L satisfies

$$L \approx 1 - \frac{4}{\mathsf{SNR}^{\lambda}\left(1+\mathsf{SNR}^{\lambda}/4\right)^{\frac{AD}{2}-1}} - \frac{D(1-A/2)}{\left(1+\mathsf{SNR}^{\lambda}/4\right)^{D(A-1)^2}} \quad (41)$$

and for $A > 2$, we have

$$L \approx 1 - \frac{4}{\mathsf{SNR}^{\lambda}\left(1+\mathsf{SNR}^{\lambda}/4\right)^{D(A-1)}} \quad (42)$$

In particular, if

$$\lim_{\mathsf{SNR}\to 0} \mathbf{E}\left[D_{\mathrm{eff}}\right] = D\mathsf{SNR}^{\lambda} = \infty \quad (43)$$

then $L \to 1$ for all $A > 1$ and $\widehat{C}_{\mathsf{coh},1,\mathsf{ST}}(\mathsf{SNR}) \to \widehat{C}_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR})$.

Proof: See Appendix A. ■

C. Discussion: Rich vs. Sparse Multipath

The result of Theorem 1 implies that the capacity of the D-bit feedback scheme approaches the benchmark ($\lambda \to 1$), the capacity with perfect transmitter and receiver CSI. Furthermore, this benchmark can be attained in the wideband limit, *even* when there is an instantaneous power constraint. As described in Proposition 1, $\mathbf{E}\left[D_{\mathrm{eff}}\right] \to \infty$ provides a sufficient condition. We now discuss the feasibility of satisfying these conditions when the channel is rich and when it is sparse. As described below, analyzing the behavior of $\mathbf{E}\left[D_{\mathrm{eff}}\right]$, the average number of active coherence subspaces, provides the key insights in this regard. Note that with $\mathsf{h_t} \sim \lambda\log\left(\frac{1}{\mathsf{SNR}}\right)$ as in (28), we have, using (20)

$$\mathbf{E}\left[D_{\mathrm{eff}}\right] = De^{-\mathsf{h_t}} = D\,\mathsf{SNR}^{\lambda} \quad (44)$$

A1) Rich multipath: For a rich channel, we note from (4) that D scales linearly with T and W. Therefore, for a fixed T, $D \sim \mathsf{SNR}^{-1}$ (since $\mathsf{SNR} = \frac{P}{W}$). Thus $\mathbf{E}\left[D_{\mathrm{eff}}\right] = D\,\mathsf{SNR}^{\lambda} = \mathsf{SNR}^{\lambda-1} \to \infty$ for $0 < \lambda < 1$. We conclude that for rich multipath the benchmark is trivially attained with both average and instantaneous power constraints.

A2) Sparse multipath: From the power-law scaling in (6), we have (ignoring constants) $D \sim T^{\delta_1}W^{\delta_2}$ and therefore

$$\mathbf{E}\left[D_{\mathrm{eff}}\right] = D\,\mathsf{SNR}^{\lambda} \sim T^{\delta_1}\mathsf{SNR}^{\lambda-\delta_2} \quad (45)$$

For a fixed T, we have, in the limit of $\mathsf{SNR} \to 0$:

$$\mathbf{E}\left[D_{\mathrm{eff}}\right] \to \begin{cases} 0 & 1 > \lambda > \delta_2 \\ \text{constant} & \lambda = \delta_2 \\ \infty & 0 < \lambda < \delta_2 \end{cases} \quad (46)$$

Thus although we can approach the benchmark for average power constraint, (46) suggests a cap ($\lambda \to \delta_2$ $(0 < \delta_2 < 1)$) on the highest achievable gain with an instantaneous power constraint.

We propose the following solution to get around this restriction. Instead of signaling with a fixed T, suppose we maintain a scaling relationship for T as a function of W, that is, $T \sim W^{\rho}$ for some $\rho > 0$. Consequently, $D \sim T^{\delta_1}W^{\delta_2} \sim W^{\delta_2+\rho\delta_1}$ and we have

$$\mathbf{E}\left[D_{\mathrm{eff}}\right] = D\,\mathsf{SNR}^{\lambda} \sim \mathsf{SNR}^{\lambda-\delta_2-\rho\delta_1} \quad (47)$$

In the limit of $\mathsf{SNR} \to 0$, the asymptotic behavior is modified as

$$\mathbf{E}\left[D_{\mathrm{eff}}\right] \to \begin{cases} 0 & 1 > \lambda > \delta_2+\rho\delta_1 \\ \text{constant} & \lambda = \delta_2+\rho\delta_1 \\ \infty & 0 < \lambda < \delta_2+\rho\delta_1 \end{cases} \quad (48)$$

By choosing $\rho \geq \frac{1-\delta_2}{\delta_1}$, that is, $\delta_2 + \rho\delta_1 \geq 1$, it follows that for all $\lambda \in (0,1)$, $\mathbf{E}\left[D_{\mathrm{eff}}\right] \to \infty$. Thus the benchmark capacity is achieved with both average and instantaneous power constraints.

$$
\begin{aligned}
&\widehat{C}_{\mathsf{coh},1,\mathsf{ST}}(\mathsf{SNR}) \\
&= \frac{1}{D}\mathbf{E}\left[\sum_{i=1}^{D}\log_2\left(1+\frac{TP}{N_c}|h_i|^2\frac{\chi(|h_i|^2\geq \mathsf{h_t})}{De^{-\mathsf{h_t}}}\chi\left(\sum_{j=1}^{i}\chi(|h_j|^2\geq \mathsf{h_t})\leq ADe^{-\mathsf{h_t}}\right)\right)\right] \\
&= \frac{1}{D}\sum_{i=1}^{D}\mathbf{E}\left[\log_2\left(1+\mathsf{SNR}\cdot e^{\mathsf{h_t}}\cdot|h_i|^2\chi(|h_i|^2\geq \mathsf{h_t})\right)\chi\left(\sum_{j=1}^{i}\chi(|h_j|^2\geq \mathsf{h_t})\leq ADe^{-\mathsf{h_t}}\right)\right] \\
&= \frac{1}{D}\sum_{i=1}^{D}\Pr\left(\sum_{j=1}^{i}\chi(|h_j|^2\geq \mathsf{h_t})\leq ADe^{-\mathsf{h_t}}\right)\cdot\mathbf{E}\left[\log_2\left(1+\mathsf{SNR}\cdot e^{\mathsf{h_t}}\cdot|h_i|^2\chi(|h_i|^2\geq \mathsf{h_t})\right)\right] \\
&\overset{(a)}{=} \mathbf{E}\left[\log_2\left(1+\mathsf{SNR}\cdot e^{\mathsf{h_t}}\cdot|h|^2\chi(|h|^2\geq \mathsf{h_t})\right)\right]\cdot\frac{\sum_{i=1}^{D}\Pr\left(\sum_{j=1}^{i}\chi(|h_j|^2\geq \mathsf{h_t})\leq ADe^{-\mathsf{h_t}}\right)}{D} = \widehat{C}_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR})\cdot\frac{\sum_{i=1}^{D}p_i}{D}
\end{aligned}
\tag{40}
$$

IV. Feedback Capacity with Channel Estimation at the Receiver

The focus of this section is on the more realistic scenario when there is no CSI at the receiver *a priori*. We first consider only an average power constraint and show that the first-order term of coherent capacity can be achieved if the channel is sparse and the channel coherence dimension N_c scales with SNR at an appropriate rate, allowing the receiver to learn the channel reliably. We also show that this is infeasible when the channel is rich, due to poor channel estimation. Within the non-coherent regime, we focus on training-based communication schemes.

We consider a communication scheme where the transmitted signals include training symbols to enable channel estimation and coherent detection. The restriction to training schemes is motivated by their practical feasibility. 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. Due to the block fading model, 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) channel estimation in this work. The reader is referred to [7, Sec. II(c)] for a more detailed description of the training scheme.

A. Capacity of the Training-based Communication Scheme with Average Power Constraint

Let $\widehat{C}_{\mathsf{train},1,\mathsf{LT}}(\mathsf{SNR})$ denote the average mutual information achievable (per-dimension) with the causal training and communication scheme satisfying the average power constraint. To characterize $\widehat{C}_{\mathsf{train},1,\mathsf{LT}}(\mathsf{SNR})$, we proceed on the same lines as the case with no feedback [7, Lemma 1]. Let $\mathbf{H}$ be the actual channel, $\widehat{\mathbf{H}}$ be the estimated channel and $\mathbf{\Delta}=\mathbf{H}-\widehat{\mathbf{H}}$ denotes the estimation error matrix. We begin with the following well-known lower-bound [21] to $\widehat{C}_{\mathsf{train},1,\mathsf{LT}}(\mathsf{SNR})$:

$$
\begin{aligned}
&\widehat{C}_{\mathsf{train},1,\mathsf{LT}}(\mathsf{SNR}) \\
&\geq \sup_{\mathbf{Q}}\frac{\mathbf{E}\left[\log_2\det\left(\mathbf{I}_{(N_c-1)D}+\widehat{\mathbf{H}}\mathbf{Q}\widehat{\mathbf{H}}^H\left(\mathbf{I}+\Sigma_{\mathbf{\Delta x}}\right)^{-1}\right)\right]}{N_cD}.
\end{aligned}
$$

where the supremum is over $\{\mathbf{Q}:\ \mathrm{Tr}(\mathbf{Q})\ \leq\ (1-\eta)TP\}$. As before, the optimal $\mathbf{Q}$ is diagonal and analogous to (18), equals

$$
\begin{aligned}
\mathbf{Q} &= \mathrm{diag}\big(\underbrace{q_1,\cdots,q_1}_{N_c-1},\underbrace{q_2,\cdots,q_2}_{N_c-1},\cdots,\underbrace{q_D,\cdots,q_D}_{N_c-1}\big) \\
q_i &= \frac{(1-\eta)TP}{(N_c-1)D}\cdot\frac{\chi\left(|\widehat{h}_i|^2\geq \mathsf{h}_{\mathsf{t}}^{\mathsf{train}}\right)}{\mathbf{E}\left[\chi\left(|\widehat{h}|^2\geq \mathsf{h}_{\mathsf{t}}^{\mathsf{train}}\right)\right]}.
\end{aligned}
\tag{49}
$$

The following theorem describes the conditions under which the capacity of the training-based communication scheme converges to the coherent capacity.

Theorem 2: If $N_c=\frac{1}{\mathsf{SNR}^{\mu}}$ for some $\mu>1$, then

$$
\lim_{\mathsf{SNR}\to 0}\frac{\widehat{C}_{\mathsf{train},1,\mathsf{LT}}(\mathsf{SNR})}{\widehat{C}_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR})}=1. \tag{50}
$$

Proof: Using the choice of $\mathbf{Q}$ from (49) in (49) and proceeding on the lines of (40), we obtain the following simplification:

$$
\begin{aligned}
&\widehat{C}_{\mathsf{train},1,\mathsf{LT}}\left(\mathsf{h}_{\mathsf{t}}^{\mathsf{train}},\eta,N_c,\mathsf{SNR}\right) \\
&= \kappa_1\cdot\left[\log_2\left(1+\frac{(1-\eta)(1+\eta N_c\mathsf{SNR})\mathsf{h}_{\mathsf{t}}^{\mathsf{train}}\mathsf{SNR}}{(1-\eta)\mathsf{SNR}+\kappa_1\kappa_2}\right)\right. \\
&\left.+\ \nu_{\frac{(1-\eta)(1+\eta N_c\mathsf{SNR})\mathsf{h}_{\mathsf{t}}^{\mathsf{train}}\mathsf{SNR}+(1-\eta)\mathsf{SNR}+\kappa_1\kappa_2}{\eta(1-\eta)N_c\mathsf{SNR}^2}}\right], \\
&\kappa_1=e^{-\frac{\mathsf{h}_{\mathsf{t}}^{\mathsf{train}}(1+\eta N_c\mathsf{SNR})}{\eta N_c\mathsf{SNR}}},\ \ \kappa_2=\eta(N_c-1)\mathsf{SNR}+\left(1-\frac{1}{N_c}\right)
\end{aligned}
\tag{51}
$$

where ν_x is as defined in (33). The tightest lower bound to (51) is obtained by maximizing $\widehat{C}_{\mathsf{train},1,\mathsf{LT}}\left(\mathsf{h}_{\mathsf{t}}^{\mathsf{train}},\eta,N_c,\mathsf{SNR}\right)$ over the fraction of energy spent on training, η, and over $\mathsf{h}_{\mathsf{t}}^{\mathsf{train}}$:

$$
C^{*}_{\mathsf{train},1,\mathsf{LT}} = \max_{\mathsf{h}_{\mathsf{t}}^{\mathsf{train}}}\left[\max_{\eta}\widehat{C}_{\mathsf{train},1}\left(\mathsf{h}_{\mathsf{t}}^{\mathsf{train}},\eta,N_c,\mathsf{SNR}\right)\right] \tag{52}
$$

As such, performing the double optimization as formulated above seems difficult. However, motivated by our study in Section III, we can assume a specific form for the threshold, given by $\mathsf{h}_{\mathsf{t}}^{\mathsf{train}}=\epsilon\log\left(\frac{1}{\mathsf{SNR}}\right)$. We omit the details here but it is rigorously shown in [20, Appendix C] that with this choice of the threshold, the optimal choice of η and N_c can be obtained in closed form and the desired result is established.

Instead, we demonstrate here that a sub-optimal, yet simpler approach actually suffices to obtain the desired result. We consider a sub-optimal choice for η^* in (52) that optimizes the average mutual information in the no feedback case [7, Lemma 2] and is given by

$$\eta^* \quad = \quad \frac{N_c\mathsf{SNR}+N_c-1}{(N_c-2)N_c\mathsf{SNR}} \cdot \left[\sqrt{1+\frac{N_c\mathsf{SNR}(N_c-2)}{N_c\mathsf{SNR}+N_c-1}}-1\right]. \quad (53)$$

With this choice, we next show that choosing a threshold of the form

$$\mathsf{h}_{\mathsf{t}}^{\mathsf{train}} = \frac{\eta^* N_c \mathsf{SNR}}{1+\eta^* N_c \mathsf{SNR}} \mathsf{h}_{\mathsf{t}} \quad (54)$$

where $\mathsf{h}_{\mathsf{t}} \sim \lambda \log\left(\frac{1}{\mathsf{SNR}}\right)$ from (28), is sufficient to obtain the first-order term of coherent capacity. To this end, let

$$\begin{aligned} A_1 &= \frac{(1-\eta^*)(1+\eta^* N_c\mathsf{SNR})\mathsf{h}_{\mathsf{t}}^{\mathsf{train}}\mathsf{SNR}}{(1-\eta^*)\mathsf{SNR}+\kappa_1\kappa_2} \\ A_2 &= \frac{(1-\eta^*)(1+\eta^* N_c\mathsf{SNR})\mathsf{h}_{\mathsf{t}}^{\mathsf{train}}\mathsf{SNR}+(1-\eta^*)\mathsf{SNR}+\kappa_1\kappa_2}{\eta^*(1-\eta^*)N_c\mathsf{SNR}^2}. \quad (55) \end{aligned}$$

It can be shown that

$$\lim_{\mathsf{SNR}\to 0} A_1 = 0 \qquad \lim_{\mathsf{SNR}\to 0} \frac{1}{A_2} = 0 \quad (56)$$

for any $\mu > 0$ and as a consequence, we have, using (51)

$$\begin{aligned} \widehat{C}_{\mathsf{train},1,\mathsf{LT}} &\geq \kappa_1 \cdot \left[\log_2(1+A_1) + \nu_{A_2}\right] \\ &\overset{(a)}{\geq} \kappa_1 \cdot \left[\log_2(1+A_1) + \frac{1}{2}\log\left(1+\frac{2}{A_2}\right)\right] \\ &\overset{(b)}{\approx} \kappa_1 \cdot \left[A_1 + \frac{1}{A_2}\right] \quad (57) \end{aligned}$$

where (a) follows from the inequality $\nu_\beta \geq \frac{1}{2}\log\left(1+\frac{2}{\beta}\right)$ and (b) is due to the low SNR approximation, the Taylor series for the $\log$ considering only the first-order term.

Substituting for $\mathsf{h}_{\mathsf{t}}^{\mathsf{train}}$ from (54) and simplifying (ignoring higher order terms) we can reduce the lower bound in (57) to

$$\begin{aligned} &\widehat{C}_{\mathsf{train},1,\mathsf{LT}}(\mathsf{SNR}) \\ &\geq (1-\eta^*)\left(\frac{N_c}{N_c-1}\right)\left(\frac{\eta^* N_c\mathsf{SNR}}{1+\eta^* N_c\mathsf{SNR}}\right)[1+\mathsf{h}_{\mathsf{t}}]\,\mathsf{SNR} \end{aligned} \quad (58)$$

Substituting for η^* from (53) and $N_c = \frac{1}{\mathsf{SNR}^\mu}$, it can be checked that when $\mu > 1$ the leading term equals $[1+\mathsf{h}_{\mathsf{t}}]\,\mathsf{SNR}$ which equals the first-order term of the coherent capacity as described in Corollary 1. On the other hand when $\mu < 1$, the leading term takes the form $\mathcal{O}\left(\mathsf{SNR}^{\frac{3-\mu}{2}}\right)$. ∎

Having established the result with average power constraint, we turn our attention to the scenario when there is an instantaneous power constraint.

B. Capacity of Training-based Scheme with Instantaneous Power Constraint

We impose a finite constraint on the instantaneous transmit power as in (38) for the communication phase of the channel learning scheme. With this constraint, we explore an achievable lower bound for the capacity. We consider exactly the same power allocation scheme as in (39) (Section III-B) and proceed on the same lines to compute the capacity. Thus, analogous to (40), we obtain

$$\begin{aligned} &\widehat{C}_{\mathsf{train},1,\mathsf{ST}}(\mathsf{SNR}) \\ &= \left(1-\frac{1}{N_c}\right)\frac{1}{D}\sum_{i=1}^{D}\mathbf{E}\left[\log_2\left(1+\frac{|\widehat{h}_i|^2 q_i(1+E_{tr})}{1+q_i+E_{tr}}\right.\right. \times \\ &\left.\left.\chi\left(\sum_{j=1}^{i}\chi(|\widehat{h}_j|^2 \geq \mathsf{h}_{\mathsf{t}}^{\mathsf{train}}) \leq \frac{ADe^{-\frac{\mathsf{h}_{\mathsf{t}}^{\mathsf{train}}(1+\eta N_c\mathsf{SNR})}{\eta N_c\mathsf{SNR}}}}{(1-\eta)}\right)\right)\right] \\ &= \widehat{C}_{\mathsf{train},1,\mathsf{LT}}(\mathsf{SNR}) \cdot \frac{\sum_{i=1}^{D} p_i^{\mathsf{train}}}{D} \end{aligned}$$

where $p_i^{\mathsf{train}} = \Pr\left(\sum_{j=1}^{i}\chi(|\widehat{h}_j|^2 \geq \mathsf{h}_{\mathsf{t}}^{\mathsf{train}}) \leq \frac{ADe^{-\frac{\mathsf{h}_{\mathsf{t}}^{\mathsf{train}}(1+\eta N_c\mathsf{SNR})}{\eta N_c\mathsf{SNR}}}}{(1-\eta)}\right)$.

Once again the problem reduces to analyzing the sum of the probabilities $\{p_i^{\mathsf{train}}\}_{i=1}^{D}$ and we desire $\frac{\sum_{i=1}^{D} p_i^{\mathsf{train}}}{D} \to 1$. Taking recourse to the analysis in Proposition 1 and by using a threshold of the form $\mathsf{h}_{\mathsf{t}}^{\mathsf{train}} = \frac{\eta^* N_c\mathsf{SNR}}{1+\eta^* N_c\mathsf{SNR}}\mathsf{h}_{\mathsf{t}}$ for the training-based communication scheme with $\mathsf{h}_{\mathsf{t}} \sim \lambda\log\left(\frac{1}{\mathsf{SNR}}\right)$ as in (28), it can be shown that the $\frac{\sum_{i=1}^{D} p_i^{\mathsf{train}}}{D}$ is lower bounded by exactly the same expression as in (41),(42). More specifically, we conclude that if $\mathbf{E}[D_{\text{eff}}] \to \infty$, then $\widehat{C}_{\mathsf{train},1,\mathsf{ST}}(\mathsf{SNR}) \to \widehat{C}_{\mathsf{train},1,\mathsf{LT}}(\mathsf{SNR})$.

C. Discussion of Results

The analysis in Sections IV-A and IV-B reveals the following two conditions that are critical in evaluating the performance of the D-bit feedback scheme in the non-coherent scenario.

C1) The channel coherence dimension, N_c, scales with SNR according to $N_c \sim \frac{1}{\mathsf{SNR}^\mu}$ with $\mu > 1$, and

C2) The independent degrees of freedom (DoF), D, in the channel scales with SNR such that $\mathbf{E}[D_{\text{eff}}] = D\mathsf{SNR}^\lambda \to \infty$ as $\mathsf{SNR} \to 0$.

With only an average power constraint, C1 is necessary and sufficient so that $\widehat{C}_{\mathsf{train},1,\mathsf{LT}}(\mathsf{SNR}) \to \widehat{C}_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR})$. In particular, with $\lambda \to 1$, we approach the perfect CSI capacity, the benchmark for all limited feedback schemes. When there is an instantaneous power constraint, we need to satisfy *both* C1 and C2 so that the benchmark can be attained.

What do these two conditions mean? Note that C1 predicates a certain minimum channel coherence level to ensure the fidelity of the training performance. On the other hand, C2 describes the required growth rate in the DoF so that $\mathbf{E}[D_{\text{eff}}] = D\mathsf{SNR}^\lambda \to \infty$ and the instantaneous power constraint is satisfied, without any loss in capacity. It is clear that the two conditions are somewhat conflicting in nature since for a richer channel, it is easier to increase D but a lot tougher to increase N_c, while for a sparser channel, it is vice versa. Therefore a natural question is if they can be satisfied simultaneously.

In this quest, we first analyze the achievability of C1. What are the conditions on the channel parameters (T_m, W_d, δ_1 and δ_2) and how do they interact with the signal space parameters (T, W and P) so that $\mu > 1$ is feasible? As we discuss next, by leveraging delay and Doppler sparsity and using peaky signaling (when necessary), $\mu > 1$ is achievable.

B1) Rich multipath: When the channel is rich in both delay and Doppler, $N_c = \frac{1}{T_m W_d}$ is fixed and does not scale with SNR. Thus we can never maintain the scaling relationship in N_c as in Theorem 2 and C1 can never be satisfied. Therefore, we cannot attain the benchmark even with only an average power constraint.

B2) Doppler sparsity only: In this case $W_{coh} = \frac{1}{T_m}$ is fixed and the scaling in N_c is only through $T_{coh} \sim f_2(T)$ (see (12)). Therefore, by scaling T with W according to $T \sim {f_2}^{-1}(W^\mu)$ and choosing $\mu > 1$, we have $N_c \sim T_{coh} \sim f_2\left({f_2}^{-1}(W^\mu)\right) \sim \frac{1}{\mathsf{SNR}^\mu}$. For the power-law scaling in (13), we obtain

$$T \sim W^{\frac{\mu}{1-\delta_1}} \tag{59}$$

B3) Delay sparsity only: In this case, $T_{coh} = \frac{1}{W_d}$ and $N_c = W_{coh} T_{coh}$ scales with SNR only through $W_{coh} \sim f_1\left(\frac{1}{\mathsf{SNR}}\right)$. Therefore, for any sub-linear f_1, we cannot satisfy $\mu > 1$. A solution to this is to use peaky signaling where training and communication is performed only on a subset of the D coherence subspaces. We model peakiness similar to [4], [7] and define $\zeta = \mathsf{SNR}^\gamma,\ \gamma > 0$ as the fraction of D over which signaling is performed. It can be shown in this scenario [7, Lemma 3] that the condition for asymptotic coherence gets relaxed to $N_c = \frac{1}{\mathsf{SNR}^{\mu_{\mathrm{peaky}}}}$ from the original $N_c = \frac{1}{\mathsf{SNR}^\mu}$ where $\mu_{\mathrm{peaky}} = \mu + \gamma$. Thus now we require $\mu_{\mathrm{peaky}} > 1$, that is $\mu > 1 - \gamma$. For the power-law scaling in (13), we have $N_c \sim f_1(W) = W^{1-\delta_2} \sim \frac{1}{\mathsf{SNR}^{1-\delta_2}}$. Thus with $\gamma > \delta_2$, we satisfy the desired condition.

B4) Delay and Doppler sparsity: Using (12), we have $W_{coh} \sim f_1(W)$ and $T_{coh} \sim f_2(T)$. Therefore, if we scale T with W according to

$$T \sim f_3(W) \quad \text{with} \quad f_3(x) = {f_2}^{-1}\left(\frac{x^\mu}{f_1(x)}\right) \tag{60}$$

so that $N_c = W_{coh} T_{coh} \sim f_1(W) f_2(f_3(W)) = f_1(W) f_2\left({f_2}^{-1}\left(\frac{W^\mu}{f_1(W)}\right)\right) \sim \frac{1}{\mathsf{SNR}^\mu}$. Thus with $\mu > 1$ in (60), we attain the desired scaling of N_c with SNR. For the power-law scaling in (13), the desired scaling in N_c can be obtained by choosing T, W and P satisfying the following canonical relationship that is obtained using (13) in (60)

$$T = \frac{\left(T_m^{\delta_2} W_d^{\delta_1}\right)^{\frac{1}{1-\delta_1}} W^{\frac{\mu-1+\delta_2}{1-\delta_1}}}{P^{\frac{\mu}{1-\delta_1}}}. \tag{61}$$

From the above discussion, it is clear that channel sparsity is necessary and in addition we also require a specific scaling relationship between T and W as defined in (61). How does this scaling law impact the scaling of D with SNR? This is critical in determining the achievability of C2, which we discussed next. We recall that by definition

$$D = \frac{TW}{N_c} = TW\,\mathsf{SNR}^\mu \tag{62}$$

Using (61) in (62) and simplifying, we obtain the induced scaling behavior on D as

$$D \sim \mathsf{SNR}^{\frac{\delta_1(1-\mu)-\delta_2}{1-\delta_1}} \tag{63}$$

Therefore, we have

$$\mathbf{E}\left[D_{\mathrm{eff}}\right] = D\,\mathsf{SNR}^\lambda = \mathsf{SNR}^{\lambda + \frac{\delta_1(1-\mu)-\delta_2}{1-\delta_1}} \tag{64}$$

and consequently

$$\mathbf{E}\left[D_{\mathrm{eff}}\right] \rightarrow \begin{cases} 0 & 1 > \lambda > \frac{\delta_2 + (\mu-1)\delta_1}{1-\delta_1} \\ \text{constant} & \lambda = \frac{\delta_2 + (\mu-1)\delta_1}{1-\delta_1} \\ \infty & 0 < \lambda < \frac{\delta_2 + (\mu-1)\delta_1}{1-\delta_1} \end{cases} \tag{65}$$

It is easily seen that if $\frac{\delta_2 + (\mu-1)\delta_1}{1-\delta_1} > 1$, that is, $\mu > \frac{1-\delta_2}{\delta_1}$, $\mathbf{E}\left[D_{\mathrm{eff}}\right] \rightarrow \infty$ for all $\lambda \in (0,1)$ and C2 is satisfied. The special cases of delay sparsity only and Doppler sparsity only (as in B2 and B3) are simple extensions and naturally follow.

In summary, the capacity of the training-based communication scheme converges to the coherent capacity and achieves the benchmark provided (i) the channel is sparse and (ii) the canonical scaling law in (61) relating the signal and channel parameters is obeyed. With only an average power constraint, we require $\mu > 1$, whereas with an instantaneous power constraint, we require

$$\mu > \max\left(1, \frac{1-\delta_2}{\delta_1}\right) \tag{66}$$

V. Concluding Remarks

We contrast the results of this work with recently made observations in [11], [12]. The focus in [11] is on training-based communication schemes and on scenarios when T_{coh} increases as SNR decreases, although there is no mention of how such scaling laws would hold in general. In particular, the authors show that capacity scales as $\log(T_{coh})\,\mathsf{SNR}$ if $\log(T_{coh}) \preceq \log\left(\frac{1}{\mathsf{SNR}}\right)$ and equals the coherent capacity, $\log\left(\frac{1}{\mathsf{SNR}}\right)\mathsf{SNR}$ when $\log(T_{coh}) \succeq \log\left(\frac{1}{\mathsf{SNR}}\right)$. On the other hand, we have shown that when the channel is sparse, channel coherence scales naturally with T and W and the benchmark gain, $\log\left(\frac{1}{\mathsf{SNR}}\right)$ can always be attained by appropriately choosing T and W. Furthermore, while [11], [12] considered only an average power constraint, we have established achievability under both average and instantaneous power constraints. Also, peaky training schemes are necessary in the framework of [11] to achieve perfect training performance. Such schemes would violate any finite instantaneous power constraint. Our findings here reveal that channel sparsity is a new degree of freedom that can be exploited in obtaining near-coherent performance with non-peaky training-based communication schemes.

Finally, the results obtained here closely parallel our earlier work [7], where we analyzed non-coherent capacity *without* feedback and showed that when $N_c = \frac{1}{\mathsf{SNR}^\mu}$ with $\mu > 1$, the channel is *asymptotically coherent*; channel estimation performance is near-perfect at a vanishing energy cost. Here we have shown, analogous to [7], that under the assumption of an error-free D-bit feedback link, the capacity of the training-based scheme converges to the benchmark. Furthermore, the cost of feedback, measured in terms of the number of feedback bits per dimension $\left(\frac{D}{N}\right)$ goes to zero asymptotically when the channel is sparse.

Appendix

A. Proof of Proposition 1

To compute $p_i \triangleq \Pr\left(\sum_{j=1}^{i} \chi(|h_j|^2 \geq \mathsf{h_t}) \leq ADe^{-\mathsf{h_t}}\right)$, we need the following result [22, Theorem 2.8, p. 57] on the tail probability of a sum of independent random variables.

Lemma 1: Let $\mathbf{X}_i, i = 1, \cdots, n$ be independent random variables with $\mathbf{E}[\mathbf{X}_i] = 0$ and $\mathbf{E}[\mathbf{X}_i^2] = \sigma_i^2$. Define $B_n = \sum_{i=1}^{n} \sigma_i^2$. If there exists a positive constant H such that

$$\mathbf{E}[\mathbf{X}_i^m] \leq \frac{1}{2} m! \sigma_i^2 H^{m-2} \tag{67}$$

for all i and $x \geq \frac{B_n}{H}$, then we have $\Pr\left(\sum_{i=1}^{n} \mathbf{X}_i > x\right) \leq \exp\left(-\frac{x}{4H}\right)$. If $x \leq \frac{B_n}{H}$, then we have $\Pr\left(\sum_{i=1}^{n} \mathbf{X}_i > x\right) \leq \exp\left(-\frac{x^2}{4B_n}\right)$. ■

To apply Lemma 1, we set $n = i$ and $\mathbf{X}_j = \chi(|h_j|^2 \geq \mathsf{h_t}) - \mathbf{E}\left[\chi(|h_j|^2 \geq \mathsf{h_t})\right] = \chi(|h_j|^2 \geq \mathsf{h_t}) - e^{-\mathsf{h_t}} = \chi_j - \mathsf{E}$ for $j = 1, \cdots, i$. Then, a simple computation of the higher moments of $\mathbf{X}_j$ implies that $\mathbf{E}[\mathbf{X}_j^2] = \sigma_j^2 = \mathsf{E}(1-\mathsf{E})$, $B_i = i\mathsf{E}(1-\mathsf{E})$, $\mathbf{E}[\mathbf{X}_j^m] = \mathsf{E}(1-\mathsf{E}) \cdot ((1-\mathsf{E})^{m-1} + (-1)^m \mathsf{E}^{m-1})$. It can be checked that $H = (1-\mathsf{E})$ is sufficient to satisfy the conditions of Lemma 1. With this setting, we have

$$\begin{aligned}\Pr&\left(\sum_{j=1}^{i} \chi(|h_j|^2 \geq \mathsf{h_t}) - i\mathsf{E} > (AD-i)\mathsf{E}\right) \\ &\leq \begin{cases} \exp\left(-\frac{(AD-i)\mathsf{E}}{4(1-\mathsf{E})}\right) & \text{if } i \leq \lfloor \frac{AD}{2} \rfloor, \\ \exp\left(-\frac{(AD-i)^2\mathsf{E}}{4i(1-\mathsf{E})}\right) & \text{if } i \geq \lfloor \frac{AD}{2} \rfloor + 1. \end{cases}\end{aligned} \tag{68}$$

If $1 < A < 2$, with $\kappa = \frac{\mathsf{E}}{4(1-\mathsf{E})}$ using (68), the following lower bound, L, holds for $\frac{\sum_{i=1}^{D} p_i}{D}$:

$$\begin{aligned} L &= 1 - \left[e^{-AD\kappa} \sum_{i \leq \lfloor \frac{AD}{2} \rfloor} e^{i\kappa} + \sum_{i \geq \lfloor \frac{AD}{2} \rfloor + 1} e^{-\frac{(AD-i)^2\kappa}{i}}\right] \\ &\overset{(a)}{=} 1 - \left[\frac{e^{-\kappa(AD-1)} \cdot (e^{\kappa \lfloor \frac{AD}{2} \rfloor} - 1)}{e^{\kappa} - 1}\right. \\ &\qquad\qquad \left. + \left(D - \left\lfloor \frac{AD}{2} \right\rfloor\right) e^{-(A-1)^2 D\kappa}\right] \\ &\geq 1 - \left[\frac{1}{e^{\kappa} - 1} \cdot e^{-\kappa\left(\frac{AD}{2} - 1\right)}\right. \\ &\qquad\qquad \left. + (1 + D(1 - A/2))\, e^{-(A-1)^2 D\kappa}\right] \end{aligned} \tag{69}$$

where (a) follows from first using $\frac{(AD-i)^2}{i} \geq (A-1)^2 D$ for all $1 \leq i \leq D$ and then upon further simplification using the sum of a geometric series.

If $A \geq 2$, we have the following lower bound to $\frac{\sum_{i=1}^{D} p_i}{D}$:

$$L = 1 - \exp(-AD\kappa) \sum_{1 \leq i \leq D} e^{i\kappa} \approx 1 - e^{-\kappa(D(A-1)-1)} \cdot \frac{1}{e^{\kappa} - 1}. \tag{70}$$

With $\mathsf{h_t} = \lambda \log\left(\frac{1}{\mathsf{SNR}}\right)$ as in (28), the dominant term of E is SNR^{λ} and hence in κ is $\frac{\mathsf{SNR}^{\lambda}}{4}$. With this choice of $\mathsf{h_t}$ in (69) and (70) and simplifying, we obtain the desired bounds in (41) and (42). It is also straightforward to see that when D satisfies $D\,\mathsf{SNR}^{-\lambda} \to \infty$ as $\mathsf{SNR} \to 0$, $L \to 1$ in both the cases. ■

References

[1] M. Medard and R. G. Gallager, "Bandwidth Scaling for Fading Multipath Channels," *IEEE Trans. Inform. Theory*, vol. 48, no. 4, pp. 840–852, Apr. 2002.

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

[3] Í. E. Telatar and D. N. C. Tse, "Capacity and Mutual Information of Wideband Multipath Fading Channels," *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1384–1400, July 2000.

[4] L. Zheng, D. N. C. Tse, and M. Medard, "Channel Coherence in the Low-SNR Regime," *IEEE Trans. Inform. Theory*, vol. 53, no. 3, pp. 976–997, March 2007.

[5] A. F. Molisch, "Ultrawideband Propagation Channels - Theory, Measurement and Modeling," *IEEE Trans. Veh. Tech.*, vol. 54, no. 5, pp. 1528–1545, Sept. 2005.

[6] 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.

[7] V. Raghavan, G. Hariharan, and A. M. Sayeed, "Capacity of Sparse Multipath Channels in the Ultra-Wideband Regime," *to appear, IEEE Journal on Selected Topics in Signal Proc.*, Oct. 2007. Preprint available at http://homepages.cae.wisc.edu/∼gauthamh.

[8] A. Goldsmith and P. Varaiya, "Capacity of Fading Channels with Channel Side Information," *IEEE Trans. Inform. Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.

[9] G. Caire, G. Tarrico, and E. Biglieri, "Optimum Power Control Over Fading Channels," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1468–1489, July 1999.

[10] E. Biglieri, J. Proakis, and S. Shamai, "Fading Channels: Information-Theoretic and Communications Aspects," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[11] S. Borade and L. Zheng, "Wideband Fading Channels with Feedback," *Allerton Conf. Commun. Cont. and Comp.*, 2004.

[12] M. Agarwal and M. Honig, "Wideband Channel Capacity with Training and Partial Feedback," *Allerton Conf. Commun. Cont. and Comp.*, 2005.

[13] 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.

[14] 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.

[15] A. M. Sayeed and B. Aazhang, "Joint Multipath-Doppler Diversity in Mobile Wireless Communications," *IEEE Trans. Commun.*, vol. 47, no. 1, pp. 123–132, Jan. 1999.

[16] 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.

[17] C. C. Chong and S. K. Yong, "A Generic Statistical-Based UWB Channel Model for High-Rise Apartments," *IEEE Trans. Antennas Propagat.*, vol. 53, no. 8, pp. 2389–2399, Aug. 2005.

[18] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, NY, 4th edition, 1980.

[19] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, National Bureau of Standards, USA, 10th edition, 1972.

[20] G. Hariharan, V. Raghavan, and A. Sayeed, "Capacity of sparse wideband channels with partial channel feedback," *submitted, European Trans. on Telecom.*, June 2007. preprint available at http://homepages.cae.wisc.edu/∼gauthamh.

[21] M. Medard, "The Effect Upon Channel Capacity in Wireless Communications of Perfect and Imperfect Knowledge of the Channel," *IEEE Trans. Inform. Theory*, vol. 46, no. 3, pp. 935–946, May 2000.

[22] V. V. Petrov, *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*, Springer, Berlin, 1975.