Capacity of Sparse Wideband Channels with Partial Channel Feedback

Gautham Hariharan, Vasanthan Raghavan and Akbar M. Sayeed

Abstract

This paper studies the ergodic capacity of wideband multipath channels with partial/limited channel state feedback. Our work builds on recent results that show a significant capacity gain in the wideband/low-SNR regime when there is perfect channel state information (CSI) at the transmitter and the receiver, relative to the case of perfect CSI at the receiver only. Furthermore, this benchmark capacity gain can be achieved with just one bit of feedback per channel coefficient. However, the input signals used in these works are peaky; that is, they have large peak-to-average power ratios. Signal peakiness requirement is related to channel coherence and many recent measurement campaigns show that, in contrast to previous assumptions, wideband channels exhibit a sparse multipath structure that naturally leads to coherence in time and frequency. In this work, we show that multipath sparsity significantly relaxes the requirement of peaky signaling in attaining the capacity gains with channel state feedback. First, we show that the benchmark capacity gain, with perfect CSI at the transmitter and the receiver, is achievable even under an instantaneous power constraint. In the more realistic non-coherent setting, we study the performance of a training-based signaling scheme with one bit of feedback per channel coefficient. We show that multipath sparsity can be leveraged to achieve the benchmark capacity gain under both average as well as instantaneous power constraints as long as the channel coherence scales at a sufficiently fast rate with the signal space dimension (time-bandwidth product). We also present guidelines for choosing signaling parameters as a function of the channel sparsity parameters to maximally exploit channel state feedback for capacity gains.

This work was supported in part by the NSF through grant #CCF-0431088. G. Hariharan is with the Qualcomm Inc., San Diego, CA 92121, USA (gharihar@qualcomm.com). V. Raghavan is with the Coordinated Science Laboratory and the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA (vasanthan_raghavan@ieee.org). A. M. Sayeed is with the Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53706, USA (akbar@engr.wisc.edu).

I. INTRODUCTION

Recent research on the fundamental limits of wideband/low-SNR communications has focused on the non-coherent regime where the impact of channel state information (CSI) on the achievable rates is critical. From a capacity perspective, spreading signals has been shown to be sub-optimal [1] and peaky or flash signaling schemes are necessary [2], [3] to achieve the non-coherent wideband capacity. Recent work by Zheng *et al.* [4] has emphasized the crucial role of channel coherence in the low-SNR regime and the importance of implicit/explicit channel learning schemes that can bridge the gap between the coherent and the non-coherent extremes. However, these results have been derived based on an implicit assumption of rich multipath where the independent degrees of freedom (DoF) in the delay domain scale linearly with bandwidth.

Recent measurement campaigns in the case of ultrawideband systems show that the number of independent DoF in the delay domain do not scale linearly with bandwidth [5]–[11]. In fact, the physical layer channel model proposed by the IEEE 802.15 working group for ultrawideband communication systems exhibits sparsity in the delay domain (see for example, the measurement data in [12, p. 15]). Motivated by these works, we introduced the notion of *multipath sparsity* in [13] as a source of channel coherence and proposed a channel modeling framework to capture the impact of sparsity in delay and Doppler on achievable rates. Specifically, the channel DoF scale *sub-linearly* with the signal space dimension (time-bandwidth product) and the analysis in [13] shows that multipath sparsity can help reduce or eliminate the need for peaky signaling in achieving wideband capacity.

Building on the results in [13], we study the impact of *channel state feedback* on achievable rates in sparse wideband channels. Although earlier works (for example [14]–[16] and references therein) have explored capacity with transmitter CSI, it is only recently [2], [17], [18] that the impact of feedback in the low-SNR, non-coherent regime has received attention. In particular, in the low-SNR regime, it is shown in [2], [17] that with an average power constraint, the capacity gain with perfect transmitter and receiver CSI (over the case when there is only receiver CSI) is $\log\left(\frac{1}{\mathsf{SNR}}\right)$. More interestingly, it is shown that a *partial/limited feedback* scheme where only one bit per independent DoF is available at the transmitter can also achieve the benchmark gain of $\log\left(\frac{1}{\mathsf{SNR}}\right)$ [2], [17]. However, for both the optimal waterfilling scheme [14], [19] as well as the one bit limited feedback scheme, the input signal tends to be peaky (or bursty) in time, leading to a high peak-to-average power ratio, and difficulties from an implementation standpoint. The need to reliably estimate the channel at the receiver leads to the use of peaky training followed by communication in [17]. Similar results have also been reported in [18] where the authors study the optimization of the training length, average training power and spreading bandwidth in a wideband setting.

The focus of this work is on leveraging multipath sparsity to overcome or reduce the need for peaky signaling schemes. We work towards this goal by providing a concise description of the sparse channel model [13] in Sec. II. We consider a system that uses short-time-Fourier (STF) signaling waveforms [20] that serve as approximate eigenfunctions for underspread channels and naturally relate channel sparsity in delay-Doppler to channel coherence in time and frequency. In particular, channel coherence is now shared between time and frequency and the role of channel coherence time T_{coh} in existing works is taken over by the *channel coherence dimension* $N_c = T_{coh} W_{coh}$, where W_{coh} denotes the coherence bandwidth.

We study the performance in the case where the receiver has perfect CSI and the transmitter has one bit (per independent DoF) in Sec. III. In contrast to [2], [17], [18] which study the performance only under an *average* (or long-term) power constraint, we also consider an *instantaneous* (or short-term) power constraint. We restrict our attention to *causal* signaling schemes that can be realized in practice. We show that an optimal threshold of the form $\mathsf{h_t} = \lambda \log\left(\frac{1}{\mathsf{SNR}}\right)$ for any $\lambda \in (0,1)$ provides a measure of achievable rate[1] which behaves as $(1+\mathsf{h_t})\,\mathsf{SNR}$ in the wideband limit. Thus when λ approaches 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 $\mathbf{E}\left[\mathsf{D_{eff}}\right]$ – the average number of *active* channel DoF (the average 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 $\mathbf{E}\left[\mathsf{D_{eff}}\right] - \mathsf{h_t} \to \infty$ as $\mathsf{SNR} \to 0$. We discuss the feasibility of this condition when the channel is rich as well as sparse.

In Sec. IV, the focus is on the case where the receiver has no CSI *a priori* and a training-based signaling scheme is employed. Along the same lines as in [17], [18], we study the rates

[1] All logarithms will be assumed to be base e and we will use nats per channel use for all rate quantities in this work.

achievable with this scheme, albeit for sparse channels. 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 rate achievable with the training-based scheme converges to the capacity with perfect receiver CSI, 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/packet duration, bandwidth and transmit power) such that $\mu > 1$ is realized. The critical role of channel sparsity is further revealed when we impose an instantaneous power constraint. In contrast to peaky signaling that violates the finiteness constraint on the peak-to-average power, channel sparsity is necessary to realize the conditions required to approach the performance gain with an instantaneous power constraint: $\mu > 1$ and $\mathbf{E}\left[\mathsf{D}_{\mathsf{eff}}\right] - \mathsf{h}_{\mathsf{t}} \rightarrow \infty$. We summarize the paper in Sec. V by highlighting our contributions and placing them in the context of [2], [17], [18].

II. System Model

In this section, we elucidate the model developed in [13] for sparse multipath channels. Our results are based on an orthogonal short-time Fourier signaling framework [20], [21] that naturally relates multipath sparsity in delay-Doppler to coherence in time and frequency.

A. Sparse Multipath Channel Modeling

A discrete, physical multipath channel can be modeled as

$$y(t) = \int_0^{T_m} \int_{-\frac{W_d}{2}}^{\frac{W_d}{2}} h(\tau, \nu) x(t-\tau) e^{j2\pi\nu t} \, \mathrm{d}\nu \, \mathrm{d}\tau + w(t) \tag{1}$$

$$h(\tau, \nu) = \sum_n \beta_n \delta(\tau - \tau_n) \delta(\nu - \nu_n), \quad y(t) = \sum_n \beta_n x(t - \tau_n) e^{j2\pi\nu_n t} + w(t) \tag{2}$$

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 the 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 where $T_m W_d \ll 1$.

We use a *virtual representation* [22], [23] of the physical model in (2) 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. Thus, we have

$$y(t) = \sum_{\ell=0}^{L} \sum_{m=-M}^{M} h_{\ell,m} x(t - \ell/W) e^{j2\pi m t/T} + w(t) \quad (3)$$

$$h_{\ell,m} \approx \sum_{n \in S_{\tau,\ell} \cap S_{\nu,m}} \beta_n \quad (4)$$

where $L = \lceil T_m W \rceil$ and $M = \lceil T W_d/2 \rceil$. The sampled representation (3) is linear and is characterized by the virtual delay-Doppler channel coefficients $\{h_{\ell,m}\}$ in (4). Each $h_{\ell,m}$ consists of the sum of gains of all paths whose delay 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 illustrated 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[2] Rayleigh fading in which $\{h_{\ell,m}\}$ are zero-mean Gaussian random variables.

Let D denote the number of non-zero channel coefficients that reflects the (dominant) statistically independent DoF in the channel and also signifies the delay-Doppler diversity afforded by the channel [22]. We decompose D as $D = D_T D_W$ where D_T denotes the Doppler/time diversity and D_W denotes the frequency/delay diversity. The channel DoF or delay-Doppler diversity is bounded as

$$D = D_T D_W \leq D_{\max} \triangleq D_{T,\max} D_{W,\max} \quad (5)$$

$$D_{T,\max} = \lceil T W_d \rceil \;,\; D_{W,\max} = \lceil T_m W \rceil \quad (6)$$

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

[2]Note that the Rayleigh fading assumption is used only for mathematical tractability. The general theme of results will continue to hold as long as the fading distributions have an exponential tail. See [17] for details and [13] for a discussion on modeling issues.

However, there is growing experimental evidence [5]–[12] 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. $D \ll D_{\max}$.

In this work, we model multipath sparsity by a *sub-linear scaling* of D_T and D_W with T and W, respectively:

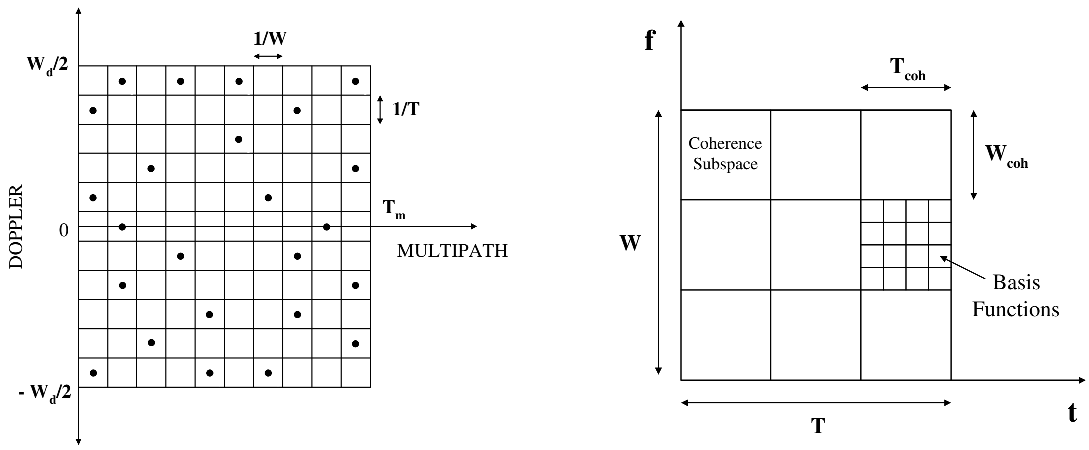
$$D_W \sim g_1(W) \ , \ D_T \sim g_2(T) \tag{7}$$

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

$$D_T = (TW_d)^{\delta_1} \ , \quad D_W = (WT_m)^{\delta_2} \tag{8}$$

for some $\delta_1, \delta_2 \in (0, 1)$. But the results derived here hold true for any general sub-linear scaling law. Note that (6) and (7) imply that in sparse multipath, the total number of delay-Doppler DoF, $D = D_T D_W$, scales *sub-linearly* with the signal space dimension $N = TW$.

***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 learned at the receiver for coherent processing.



(a) (b)

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

B. Orthogonal Short-Time Fourier Signaling

We consider signaling using an orthonormal short-time Fourier (STF) basis [20], [21] that is a natural generalization[3] 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

$$x(t) = \sum_{\ell=0}^{N_T-1} \sum_{m=0}^{N_W-1} x_{\ell m} \phi_{\ell m}(t) \qquad 0 \le t \le T \tag{9}$$

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

$$y_{\ell m} = \langle y, \phi_{\ell m} \rangle = \sum_{\ell', m'} h_{\ell m, \ell' m'} \;\; x_{\ell' m'} + w_{\ell m}. \tag{10}$$

We can represent the system using an N-dimensional matrix equation [20], [21]

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \tag{11}$$

where $\mathbf{w}$ is the additive noise vector whose entries are i.i.d. $\mathcal{CN}(0,1)$. The $N \times N$ matrix $\mathbf{H}$ consists of the channel coefficients $\{h_{\ell m, \ell' m'}\}$ in (10). We assume that the input symbols that form the transmit codeword $\mathbf{x}$ satisfy an average power constraint

$$\frac{1}{T} \cdot \mathbf{E}\left[\|\mathbf{x}\|^2\right] \le P. \tag{12}$$

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 [20], [21]; that is, (10) simplifies to[4] $y_{\ell m} \approx h_{\ell m} x_{\ell m} + w_{\ell m}$. Thus the channel matrix $\mathbf{H}$ is approximately

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

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

diagonal. In this work, we assume that $\mathbf{H}$ is exactly diagonal; that is,

$$\mathbf{H} = \text{diag}\Big[\underbrace{h_{11}\cdots h_{1N_c}}_{\text{Subspace 1}},\ \underbrace{h_{21}\cdots h_{2N_c}}_{\text{Subspace 2}}\ \cdots\ \underbrace{h_{D1}\cdots h_{DN_c}}_{\text{Subspace } D}\Big]. \tag{13}$$

The diagonal entries of $\mathbf{H}$ in (13) admit an intuitive block fading interpretation in terms of *time-frequency coherence subspaces* [20] 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 (13), 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 $\mathbf{E}[|h_i|^2] = \sum_n \mathbf{E}[|\beta_n|^2] = 1$ [20].

Using the DoF scaling for sparse channels in (7), the scaling behavior for the coherence dimension can be computed as

$$W_{coh} = \frac{W}{D_W} \sim f_1(W), \quad T_{coh} = \frac{T}{D_T} \sim f_2(T) \tag{14}$$

$$N_c = W_{coh}T_{coh} \sim f_1(W)f_2(T) \tag{15}$$

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 (7), f_1 and f_2 are also sub-linear. In particular, corresponding to the power-law scaling in (8), we obtain

$$T_{coh} = \frac{T^{1-\delta_1}}{W_d^{\delta_1}}, \quad W_{coh} = \frac{W^{1-\delta_2}}{T_m^{\delta_2}}. \tag{16}$$

***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, the focus is on computing achievable rates in the non-coherent setting with feedback and as we will see in Sec. III and IV, the rates turn 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:

$$N_c = \frac{1}{\mathsf{SNR}^{\mu}}, \quad \mu > 0 \tag{17}$$

where the parameter μ reflects the level of channel coherence. We will revisit (17) and discuss its achievability and implications in Sec. IV.

III. ACHIEVABLE RATES WITH PERFECT RECEIVER CSI AND LIMITED CHANNEL STATE 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 are known. On one extreme, with perfect receiver CSI and no transmitter CSI (no feedback), the coherent capacity per dimension (in nats/s/Hz) equals

$$C_{\mathsf{coh},0}(\mathsf{SNR}) = \sup_{\mathbf{Q}:\mathrm{Tr}(\mathbf{Q})\leq TP} \frac{\mathbf{E}\left[\log\det\left(\mathbf{I}_{N_cD} + \mathbf{H}\mathbf{Q}\mathbf{H}^H\right)\right]}{N_cD}. \tag{18}$$

The optimal input $\mathbf{x}$ is zero-mean Gaussian and the optimization is over the set of N_cD-dimensional positive definite input covariance matrices $\mathbf{Q} = \mathbf{E}\left[\mathbf{x}\mathbf{x}^H\right]$ satisfying the average power constraint in (12). Due to the diagonal nature of $\mathbf{H}$ in (13), the optimal $\mathbf{Q}$ is also diagonal. Furthermore, with no transmitter CSI, the uniform power allocation $\mathbf{Q} = \frac{TP}{N_cD}\mathbf{I}_{N_cD} = \mathsf{SNR}\cdot\mathbf{I}_{N_cD}$ achieves the optimum; that is, the elements of $\mathbf{x}$ are i.i.d. zero-mean Gaussian with variance SNR. Using this optimum $\mathbf{Q}$ and the block-fading structure of $\mathbf{H}$, the expression for coherent capacity in (18) reduces to

$$C_{\mathsf{coh},0}(\mathsf{SNR}) = \frac{1}{D}\sum_{i=1}^{D}\mathbf{E}\left[\log\left(1 + \frac{PT}{N_cD}|h_i|^2\right)\right] \tag{19}$$

which in the limit of low-SNR is [2], [4]

$$C_{\mathsf{coh},0}(\mathsf{SNR}) \approx \mathsf{SNR} - \mathsf{SNR}^2. \tag{20}$$

On the other extreme is the case of perfect receiver and transmitter CSI, where the receiver instantaneously feeds back all the channel coefficients, $\{h_i\}_{i=1}^{D}$, corresponding to the D independent coherence subspaces to the transmitter. The optimum transmitter power allocation in this case is waterfilling [14], [19] over the different coherence subspaces. In the low-SNR extreme, it is shown in [2], [17] that the capacity with perfect transmitter CSI scales as $\log\left(\frac{1}{\mathsf{SNR}}\right)\mathsf{SNR}$. That is, the capacity gain (compared with the receiver CSI only case) is directly proportional to the waterfilling 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 [2], [17] that this benchmark (maximum) capacity gain can be achieved with just one bit of feedback per channel coefficient.

In the case of limited feedback, both the transmitter and the receiver have *a priori* knowledge of a common threshold denoted by $\mathsf{h_t}$. 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

$$b_i = \chi(|h_i|^2 \geq \mathsf{h_t}) = \begin{cases} 1 & \text{if} \quad |h_i|^2 \geq \mathsf{h_t} \\ 0 & \text{if} \quad |h_i|^2 < \mathsf{h_t}. \end{cases} \tag{21}$$

At the transmitter, conditioned on the limited feedback about the CSI, the input symbols $\mathbf{x}$ are still zero-mean independent Gaussian and the power allocation is uniform across the coherence subspaces for which $b_i = 1$ and no power is allocated to those subspaces for which $b_i = 0$. Thus, conditioned on $\{b_i\}_{i=1}^{D}$, the input covariance matrix is given by

$$\mathbf{Q}(\{b_i\}) = \text{diag}\big(\underbrace{q_1, \cdots, q_1}_{N_c}, \underbrace{q_2, \cdots, q_2}_{N_c}, \cdots, \underbrace{q_D, \cdots, q_D}_{N_c}\big) \tag{22}$$

$$q_i = P_{\text{o}} b_i = P_{\text{o}} \cdot \chi(|h_i|^2 \geq \mathsf{h_t}) \tag{23}$$

where q_i denotes the power of the Gaussian input in the i-th coherence subspace. The choice of P_{o} depends on the type of power constraint and also on the choice of threshold $\mathsf{h_t}$. To explore this further, let $\mathsf{D_{eff}}$ denote the number of *active* subspaces, those which exceed the threshold $\mathsf{h_t}$. We have

$$\mathsf{D_{eff}} = \sum_{i=1}^{D} b_i = \sum_{i=1}^{D} \chi(|h_i|^2 \geq \mathsf{h_t}) \tag{24}$$

$$\mathbf{E}\left[\mathsf{D_{eff}}\right] \stackrel{(a)}{=} D\mathbf{E}\left[\chi(|h|^2 \geq \mathsf{h_t})\right] \stackrel{(b)}{=} De^{-\mathsf{h_t}} \tag{25}$$

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

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

$$P_{\text{o},\mathsf{nc}} = \frac{TP}{N_c \mathsf{D_{eff}}} \tag{26}$$

and the maximum rate achievable with this power allocation, denoted by $C_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR})$, is

$$C_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR}) = \max_{\mathsf{h_t}} \frac{1}{D} \sum_{i=1}^{D} \mathbf{E}\left[\log\left(1 + \frac{TP}{N_c \mathsf{D_{eff}}} \cdot |h_i|^2\right) \chi\left(|h_i|^2 \geq \mathsf{h_t}\right)\right]. \tag{27}$$

The power allocation in (26) satisfies the power constraint instantaneously as well as on average. To see this, note that

$$P_{\mathsf{inst,nc}} = \frac{1}{T}\|\mathbf{x}\|^2 = \frac{N_c}{T}\sum_{i=1}^{D} q_i = \frac{N_c}{T}\sum_{i=1}^{D}\frac{TP}{N_c\mathsf{D}_{\mathsf{eff}}}\chi\left(|h_i|^2 \geq \mathsf{h_t}\right) = P \tag{28}$$

and clearly $\mathbf{E}\left[P_{\mathsf{inst,nc}}\right] = P$ as well. The non-causality of the scheme is more relevant in the 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, with an average power constraint, in which $P_{\rm o}$ in (23) depends on the *average* number of active coherence subspaces, $\mathbf{E}[\mathsf{D}_{\mathsf{eff}}]$, rather than the instantaneous value of $\mathsf{D}_{\mathsf{eff}}$ as in (26). From (23), we have

$$\mathbf{E}\left[\|\mathbf{x}\|^2\right] = N_c\sum_{i=1}^{D}\mathbf{E}\left[q_i\right] = N_c\sum_{i=1}^{D}P_{\rm o}\cdot\mathbf{E}\left[\chi(|h_i|^2 \geq \mathsf{h_t})\right] = N_cP_{\rm o}\mathbf{E}\left[\mathsf{D}_{\mathsf{eff}}\right]. \tag{29}$$

Thus to satisfy $\mathbf{E}\left[\|\mathbf{x}\|^2\right] \leq TP$, the power allocation for the causal scheme is given by

$$P_{\mathsf{o,c}} = \frac{TP}{N_c\mathbf{E}\left[\mathsf{D}_{\mathsf{eff}}\right]} = \frac{TP}{N_cDe^{-\mathsf{h_t}}} \tag{30}$$

and the corresponding maximum achievable rate, $\widehat{C}_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR})$, is given by

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

Note that the causal average power allocation policy in (30) satisfies the average power constraint but can have a large instantaneous power. This is because

$$P_{\mathsf{inst,c}} = \frac{N_c}{T}\sum_{i=1}^{D}\frac{TP}{N_cDe^{-\mathsf{h_t}}}\chi\left(|h_i|^2 \geq \mathsf{h_t}\right) = \left(\frac{\mathsf{D}_{\mathsf{eff}}}{De^{-\mathsf{h_t}}}\right)P. \tag{32}$$

Thus $\mathbf{E}\left[P_{\mathsf{inst,c}}\right] \leq P$, but unlike (28), $P_{\mathsf{inst,c}} \in [0,\infty)$ depending on the choice of $\mathsf{h_t}$. We will address this issue in Sec. III-B, but first we consider the causal scheme, that satisfies the average power constraint, more carefully.

A. Achievable Rates under 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 (31).

Theorem 1: Given any $\lambda \in (0,1)$, a causal on-off signaling scheme under an average power constraint achieves $\widehat{C}_{\mathsf{LB}} \leq \widehat{C}_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR}) \leq \widehat{C}_{\mathsf{UB}}$ with an optimal threshold that satisfies

$$\lim_{\mathsf{SNR}\to 0} \frac{\mathsf{h_t}}{\lambda \log\left(\frac{1}{\mathsf{SNR}}\right)} = 1 \tag{33}$$

where

$$\widehat{C}_{\mathsf{UB}} = \mathsf{SNR}^{\lambda} \cdot \left[\log\left(1+\lambda\mathsf{SNR}^{1-\lambda}\log\left(\tfrac{1}{\mathsf{SNR}}\right)\right) + \log\left(1+\frac{\mathsf{SNR}^{1-\lambda}}{1+\lambda\mathsf{SNR}^{1-\lambda}\log\left(\frac{1}{\mathsf{SNR}}\right)}\right)\right] \tag{34}$$

$$\widehat{C}_{\mathsf{LB}} = \mathsf{SNR}^{\lambda} \cdot \left[\log\left(1+\lambda\mathsf{SNR}^{1-\lambda}\log\left(\tfrac{1}{\mathsf{SNR}}\right)\right) + \frac{1}{2}\log\left(1+\frac{2\mathsf{SNR}^{1-\lambda}}{1+\lambda\mathsf{SNR}^{1-\lambda}\log\left(\frac{1}{\mathsf{SNR}}\right)}\right)\right]. \tag{35}$$

Proof: Starting from (31), we have

$$\widehat{C}_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR}) = \max_{\mathsf{h_t}} \frac{1}{D}\sum_{i=1}^{D} \mathbf{E}\left[\log\left(1+\frac{TP}{N_c D e^{-\mathsf{h_t}}}|h_i|^2\right)\chi(|h_i|^2 \geq \mathsf{h_t})\right] \tag{36}$$

$$\stackrel{(a)}{=} \mathbf{E}\left[\log\left(1+\mathsf{SNR}\, e^{\mathsf{h_t}}\,|h|^2\right)\chi(|h|^2 \geq \mathsf{h_t})\right] \tag{37}$$

where (a) follows from the fact that $\{h_i\}$ are i.i.d. $\mathcal{CN}(0,1)$ and h is a generic i.i.d. $\mathcal{CN}(0,1)$ random variable. The expectation in (37) can be computed using [24, 4.337(1), p. 574]. With $\alpha \triangleq \frac{1+\mathsf{SNR}\,\mathsf{h_t}\,e^{\mathsf{h_t}}}{\mathsf{SNR}\,e^{\mathsf{h_t}}}$, we have

$$\widehat{C}_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR}) = e^{-\mathsf{h_t}} \cdot \left[\log\left(1+\mathsf{SNR}\,\mathsf{h_t}\,e^{\mathsf{h_t}}\right) + \exp(\alpha)\int_{\alpha}^{\infty}\frac{e^{-t}}{t}\,\mathrm{d}t\right] \tag{38}$$

$$= e^{-\mathsf{h_t}} \cdot \left[\log\left(1+\mathsf{SNR}\,\mathsf{h_t}\,e^{\mathsf{h_t}}\right) + \nu_\alpha\right] \tag{39}$$

where $\nu_\alpha \triangleq \exp(\alpha)\int_{\alpha}^{\infty}\frac{e^{-t}}{t}\,\mathrm{d}t$. It can be checked that the choice of $\mathsf{h_t}$ maximizing (39) is obtained by setting its derivative with respect to $\mathsf{h_t}$ to zero and satisfies

$$\Delta \triangleq 1 - \log\left(1+\mathsf{SNR}\mathsf{h_t}e^{\mathsf{h_t}}\right) - \frac{1}{\mathsf{SNR}e^{\mathsf{h_t}}}\cdot\nu_\alpha = 0. \tag{40}$$

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 $\alpha \to \infty$. Note that as $\alpha \to \infty$, the following bounds hold for ν_α [25, 5.1.20, p. 229]:

$$\frac{1}{2}\log\left(1+\frac{2}{\alpha}\right) \leq \nu_\alpha \leq \log\left(1+\frac{1}{\alpha}\right). \tag{41}$$

Thus we can approximate ν_α as $\nu_\alpha \approx \frac{1}{\alpha}$. With this approximation in (40), we have $\frac{1}{\mathsf{SNR}e^{\mathsf{h_t}}}\cdot\nu_\alpha \approx \frac{1}{1+\mathsf{SNR}\mathsf{h_t}e^{\mathsf{h_t}}} \to 1$. Using the choice of $\mathsf{h_t}$ as in (33), it follows that as $\mathsf{SNR} \to 0$, $\Delta \to 0$. Substituting this choice of $\mathsf{h_t}$ in (39) and using the upper and lower bounds on ν_α in (41), we obtain the bounds in (34) and (35). ■

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 (27). That is, $\widehat{C}_{\text{coh},1,\text{LT}}(\text{SNR})$ is a tight bound to $C_{\text{coh},1,\text{LT}}(\text{SNR})$ and for all $\lambda \in (0,1)$, we have

$$\lim_{\text{SNR}\to 0} \frac{\left|C_{\text{coh},1,\text{LT}}(\text{SNR}) - \widehat{C}_{\text{coh},1,\text{LT}}(\text{SNR})\right|}{C_{\text{coh},1,\text{LT}}(\text{SNR})} \quad = \quad 0. \tag{42}$$

The proof of the above statement can be found in Appendix A.

***Corollary** 1:* The capacity gain for the D-bit channel state feedback, causal power allocation scheme over the capacity with only receiver CSI in (20) is

$$\lim_{\text{SNR}\to 0} \frac{\widehat{C}_{\text{coh},1,\text{LT}}(\text{SNR})}{C_{\text{coh},0}(\text{SNR})} = (1 + \text{h}_\text{t}) = 1 + \lambda \log\left(\frac{1}{\text{SNR}}\right). \tag{43}$$

Proof: A Taylor series expansion of the upper and lower bounds in (34) and (35) shows that they are equal up to first-order. This common term is such that

$$\widehat{C}_{\text{coh},1,\text{LT}}(\text{SNR}) = \text{SNR}\left(1 + \lambda \log\left(\frac{1}{\text{SNR}}\right)\right) = (1 + \text{h}_\text{t})\text{SNR}. \tag{44}$$

On the other hand, with CSI at the receiver alone, we have from (20), $\frac{C_{\text{coh},0}(\text{SNR})}{\text{SNR}} = (1 + \text{o}(1))$. Thus the desired result follows. ∎

***Remark** 3:* The capacity gain due to feedback is directly proportional to h_t and the highest gain is obtained by choosing $\lambda \to 1$, and equals the benchmark where perfect CSI is available at both the ends [17]. Statements analogous to those in Theorem 1 and Corollary 1 are well-known from prior work; see [2], [17], [18] for details.

We now revert our attention back to the instantaneous transmit power described in (32). Note that as $D \to \infty$, $P_{\text{inst,c}} \to P$ as a consequence of the law of large numbers. However, for any finite D, $P_{\text{inst,c}}$ may be much larger than P. This is an important 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 (32), as discussed next.

B. Achievable Rates under Instantaneous Power Constraint

In addition to the average power constraint in (29) and (30), let us impose a constraint on the instantaneous transmit power of the form

$$P_{\text{inst,c}} \overset{a.s.}{\leq} AP \tag{45}$$

where $A > 1$ is finite. With this short-term constraint, we now compute the rate, $\widehat{C}_{\mathsf{coh},1,\mathsf{ST}}(\mathsf{SNR})$, achievable with the causal signaling scheme. We are particularly interested in exploring conditions under which $\widehat{C}_{\mathsf{coh},1,\mathsf{ST}}(\mathsf{SNR}) \approx \widehat{C}_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR})$. To this end, we employ the same power allocation as in (22) but the q_i in (23) are now given by

$$q_i = P_{\mathsf{o,c}}\, \chi(|h_i|^2 \geq \mathsf{h_t})\, \chi\left(\sum_{j=1}^{i} \chi(|h_j|^2 \geq \mathsf{h_t}) \leq ADe^{-\mathsf{h_t}} \right) \quad . \tag{46}$$

The second indicator function in (46) checks for the constraint in (45) causally in each time-frequency coherence subspace and allocates power only if this constraint is met. For a given threshold $\mathsf{h_t}$, the achievable rate with this power allocation scheme is given by

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

where $\widehat{C}_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR})$ is the rate achievable with only an average power constraint, and (a) follows from the fact that $\{h_i\}$ are i.i.d. and

$$p_i \triangleq \Pr\left(\sum_{j=1}^{i} \chi(|h_j|^2 \geq \mathsf{h_t}) \leq ADe^{-\mathsf{h_t}} \right). \tag{47}$$

Thus, characterizing $\widehat{C}_{\mathsf{coh},1,\mathsf{ST}}(\mathsf{SNR})$ is equivalent to computing p_i. In particular, under what condition does $\frac{\sum_{i=1}^{D} p_i}{D} \rightarrow 1$? This is discussed in the following proposition.

***Proposition** 1:* With $\mathsf{h_t} \sim \lambda \log\left(\frac{1}{\mathsf{SNR}}\right)$ as in (33), we have $\frac{\sum_{i=1}^{D} p_i}{D} \geq L$ where

$$L \approx 1 - \frac{4}{\mathsf{SNR}^{\lambda}\left(1+\mathsf{SNR}^{\lambda}/4\right)^{\frac{AD}{2}-1}} - \frac{D(1-A/2)}{\left(1+\mathsf{SNR}^{\lambda}/4\right)^{D(A-1)^2}} \tag{48}$$

if $1 < A < 2$, and if $A > 2$, we have

$$L \approx 1 - \frac{4}{\mathsf{SNR}^{\lambda}\left(1+\mathsf{SNR}^{\lambda}/4\right)^{D(A-1)}}. \tag{49}$$

In particular, if

$$\mathbf{E}\left[\mathsf{D}_{\mathsf{eff}}\right] - \mathsf{h}_{\mathsf{t}} = De^{-\mathsf{h}_{\mathsf{t}}} - \mathsf{h}_{\mathsf{t}} \sim D\mathsf{SNR}^{\lambda} + \lambda\log(\mathsf{SNR}) \rightarrow \infty \text{ as } \mathsf{SNR} \rightarrow 0, \tag{50}$$

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

Proof: See Appendix B. ■

C. Discussion: Rich vs. Sparse Multipath

Theorem 1 states that the rate achievable with the D-bit channel state feedback scheme approaches the benchmark, the perfect transmitter CSI capacity, when $\lambda \rightarrow 1$. Furthermore, this benchmark can be attained in the wideband limit, *even* when there is an instantaneous power constraint. As described in Prop. 1, $\mathbf{E}\left[\mathsf{D}_{\mathsf{eff}}\right] - \mathsf{h}_{\mathsf{t}} \rightarrow \infty$ provides a sufficient condition. We now discuss the feasibility of satisfying these conditions when the channel is rich and when it is sparse. The behavior of $\mathbf{E}\left[\mathsf{D}_{\mathsf{eff}}\right]$ provides key insights in this regard.

A1) Rich multipath: For a rich channel, from (6) we note that D scales linearly with T and W. For a fixed T, $D \sim \mathsf{SNR}^{-1}$ (since $\mathsf{SNR} = \frac{P}{W}$). That is, $\mathbf{E}\left[\mathsf{D}_{\mathsf{eff}}\right] - \mathsf{h}_{\mathsf{t}} = D\mathsf{SNR}^{\lambda} + \lambda\log(\mathsf{SNR}) \rightarrow \infty$ for $0 < \lambda < 1$. We can thus conclude that for rich multipath the perfect CSI benchmark is attained trivially with both average and instantaneous power constraints.

A2) Sparse multipath: From the power-law scaling in (8), ignoring the constant factors, we have $D \sim T^{\delta_1}W^{\delta_2}$ and therefore

$$\mathbf{E}\left[\mathsf{D}_{\mathsf{eff}}\right] - \mathsf{h}_{\mathsf{t}} \sim T^{\delta_1}\mathsf{SNR}^{\lambda-\delta_2} + \lambda\log(\mathsf{SNR}). \tag{51}$$

For a fixed T, as $\mathsf{SNR} \rightarrow 0$, we have

$$\mathbf{E}\left[\mathsf{D}_{\mathsf{eff}}\right] - \mathsf{h}_{\mathsf{t}} \rightarrow \begin{cases} \infty & \text{if } 0 < \lambda < \delta_2 \\ -\infty & \text{if } \delta_2 \leq \lambda < 1. \end{cases} \tag{52}$$

While we can approach the benchmark capacity with an average power constraint, (52) suggests a cap on λ, the highest achievable gain with an instantaneous power constraint.

D. Capacity Optimal Packet Configurations

From (52), we see that the perfect CSI gain is not always achievable when there is an instantaneous power constraint. However, we note that (52) is derived assuming a *fixed* choice of T, while we know that sparsity in Doppler facilitates any desired scaling in the DoF with increasing T. Leveraging both delay and Doppler sparsities, we propose the following solution to get around the restriction in **A2**. Instead of signaling with a fixed duration T, let us suppose that we maintain a scaling relationship for T as a function of W. For example, let $T \sim W^{\rho}$ for some $\rho > 0$. Consequently, $D \sim T^{\delta_1} W^{\delta_2} \sim W^{\delta_2 + \rho\delta_1}$ and we have

$$\mathbf{E}\left[\mathsf{D}_{\mathsf{eff}}\right] - \mathsf{h}_{\mathsf{t}} \sim \mathsf{SNR}^{\lambda - \delta_2 - \rho\delta_1} + \lambda \log(\mathsf{SNR}). \tag{53}$$

Thus in the limit as $\mathsf{SNR} \to 0$, the asymptotic behavior of $\mathbf{E}\left[\mathsf{D}_{\mathsf{eff}}\right] - \mathsf{h}_{\mathsf{t}}$ is given by

$$\mathbf{E}\left[\mathsf{D}_{\mathsf{eff}}\right] - \mathsf{h}_{\mathsf{t}} \to \begin{cases} \infty & \text{if } 0 < \lambda < \delta_2 + \rho\delta_1 \\ -\infty & \text{if } \delta_2 + \rho\delta_1 \geq \lambda < 1. \end{cases} \tag{54}$$

Note that in (54), we have

$$\delta_2 + \rho\delta_1 \geq 1 \iff \rho \geq \frac{1 - \delta_2}{\delta_1} \tag{55}$$

which leads to the desired result that $\mathbf{E}\left[\mathsf{D}_{\mathsf{eff}}\right] - \mathsf{h}_{\mathsf{t}} \to \infty$ for all $\lambda \in (0, 1)$ and thus the benchmark capacity is achievable even under an instantaneous power as long as ρ satisfies (55) and $T \sim W^{\rho}$.

To further illustrate this idea, we present an example when channel sparsity follows the power-law scaling in (8). For simplicity, let us assume that $\delta_1 = \delta_2 = \delta$. From (55), we require $T \sim W^{\rho}$ with $\rho \geq \frac{1-\delta}{\delta}$ to achieve the benchmark performance. With $N = TW$, the capacity optimal (T, W) packet configuration is then given by

$$T \sim N^{\frac{\rho}{1+\rho}}, \quad W \sim N^{\frac{1}{1+\rho}}. \tag{56}$$

Fig. 2 illustrates the optimal packet configuration for a rich multipath channel ($\delta \to 1$), for a medium sparse channel ($\delta = 0.5$), and for a very sparse channel ($\delta \to 0$). As evident, for rich channels, the optimal packet configuration is narrow in time relative to bandwidth, whereas in extremely sparse channels the optimal packet configuration is narrow in bandwidth relative to time. For medium sparse channels, both T and W scale at the same rate with N. These guidelines can be easily extended to generic sub-linear scaling laws.

IV. Achievable Rates with Channel Estimation at the Receiver

We now consider the more realistic case where no CSI is available *a priori* at the receiver. We first consider only an average power constraint and show that the first-order term of the benchmark capacity can be achieved if the channel is sparse and the channel coherence dimension,

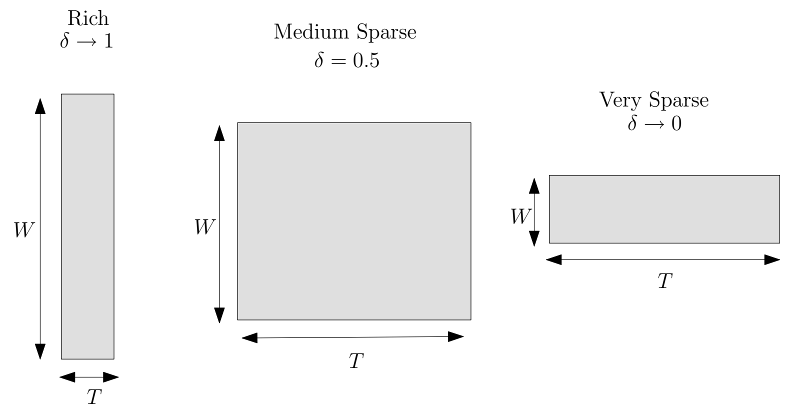


Fig. 2. *Optimal packet configurations with perfect receiver CSI and limited feedback as a function of richness of the channel. Three cases are illustrated here: Rich multipath* $(\delta \to 1)$*, medium sparsity* $(\delta = 0.5)$ *and very high sparsity* $(\delta \to 0)$.

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.

More specifically, we consider a training-based signaling scheme where the transmitted signals include training symbols to enable channel estimation and coherent detection at the receiver. The restriction to training schemes is motivated by their easy realizability. 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 in communication. With the block fading model, this means that one signal space dimension in each coherence subspace is used for training and the remaining $(N_c - 1)$ are used in communication. This is pictorially illustrated in Fig. 3. In each coherence subspace, the receiver estimates the channel coefficient using the training symbol which is then used to coherently detect the $N_c - 1$ communication symbols. We consider minimum mean-squared error (MMSE) channel estimation and the reader is referred to [13, Sec. II(c)] for more details on the training scheme.

A. Achievable Rates under Average Power Constraint

Let $\widehat{C}_{\mathsf{train},1,\mathsf{LT}}(\mathsf{SNR})$ denote the average mutual information (per dimension) achievable with the causal training scheme under the average power constraint. We note that the power constraint in this case is a slight modification of (30) to account for the fraction of energy used in training. We proceed along the same lines as the no-feedback case [13, Lemma 1] to characterize

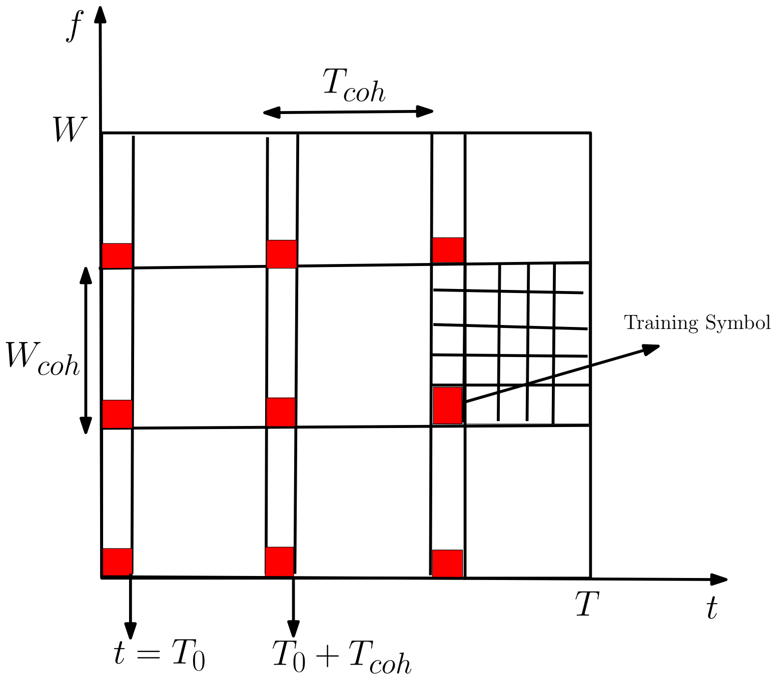


Fig. 3. *Training-based signaling scheme in the STF domain. The D estimated channel coefficients determine the D feedback bits for the communication scheme with limited feedback.*

$\widehat{C}_{\mathsf{train},1,\mathsf{LT}}(\mathsf{SNR})$. Let $\mathbf{H}$ be the actual channel, $\widehat{\mathbf{H}}$ be the estimated channel and $\mathbf{\Delta} = \mathbf{H} - \widehat{\mathbf{H}}$ denote the estimation error matrix. We begin with the following well-known lower-bound [26] to $\widehat{C}_{\mathsf{train},1,\mathsf{LT}}(\mathsf{SNR})$ corresponding to the no-feedback case

$$\widehat{C}_{\mathsf{train},1,\mathsf{LT}}(\mathsf{SNR}) \geq \sup_{\mathbf{Q}} \frac{\mathbf{E}\left[\log\det\left(\mathbf{I}_{(N_c-1)D} + \widehat{\mathbf{H}}\mathbf{Q}\widehat{\mathbf{H}}^H \left(\mathbf{I} + \Sigma_{\mathbf{\Delta x}}\right)^{-1}\right)\right]}{N_c D} \tag{57}$$

where the input is taken to be zero-mean Gaussian with covariance matrix $\mathbf{Q}$ and the supremum is over $\{\mathbf{Q} : \mathrm{Tr}(\mathbf{Q}) \leq (1-\eta)TP\}$. For the limited-feedback scheme, conditioned on the D feedback bits, $\{b_i\}$, the optimal $\mathbf{Q}$ is again diagonal and, analogous to (23), is given by

$$\mathbf{Q}(\{b_i\}) = \mathrm{diag}\big(\underbrace{q_1, \cdots, q_1}_{N_c-1}, \underbrace{q_2, \cdots, q_2}_{N_c-1}, \cdots, \underbrace{q_D, \cdots, q_D}_{N_c-1}\big) \tag{58}$$

$$q_i = \frac{(1-\eta)TP}{(N_c-1)D} \cdot \frac{\chi\left(|\widehat{h}_i|^2 \geq \mathsf{h}_{\mathsf{t}}^{\mathsf{train}}\right)}{\mathbf{E}\left[\chi\left(|\widehat{h}|^2 \geq \mathsf{h}_{\mathsf{t}}^{\mathsf{train}}\right)\right]} \tag{59}$$

where $\mathsf{h}_\mathsf{t}^{\mathsf{train}}$ is the threshold in this setting. The following theorem describes conditions under which the achievable rate with the training scheme converges to that with perfect receiver CSI. In particular, the choice of the threshold is discussed in the proof.

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

$$\lim_{\mathsf{SNR}\to 0} \frac{\widehat{C}_{\mathsf{train},1,\mathsf{LT}}(\mathsf{SNR})}{\widehat{C}_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR})} = 1. \tag{60}$$

Proof: Using the choice of $\mathbf{Q}$ from (59) in (57) and proceeding along the lines of (47), we obtain

$$\begin{aligned} \widehat{C}_{\mathsf{train},1,\mathsf{LT}}\left(\mathsf{h}_\mathsf{t}^{\mathsf{train}}, \eta, N_c, \mathsf{SNR}\right) &= \kappa_1 \cdot \left[\log\left(1 + \frac{(1-\eta)(1+\eta N_c\mathsf{SNR})\mathsf{h}_\mathsf{t}^{\mathsf{train}}\mathsf{SNR}}{(1-\eta)\mathsf{SNR} + \kappa_1\kappa_2}\right)\right. \\ &\quad \left. + \nu_{\frac{(1-\eta)(1+\eta N_c\mathsf{SNR})\mathsf{h}_\mathsf{t}^{\mathsf{train}}\mathsf{SNR}+(1-\eta)\mathsf{SNR}+\kappa_1\kappa_2}{\eta(1-\eta)N_c\mathsf{SNR}^2}}\right], \end{aligned} \tag{61}$$

$$\kappa_1 = e^{-\frac{\mathsf{h}_\mathsf{t}^{\mathsf{train}}(1+\eta N_c\mathsf{SNR})}{\eta N_c\mathsf{SNR}}}, \;\; \kappa_2 = \eta(N_c - 1)\mathsf{SNR} + \left(1 - \frac{1}{N_c}\right) \tag{62}$$

where $\nu_{\bullet}$ is as defined following (39). The tightest lower bound to (61) 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 η, the fraction of energy spent on training, and over $\mathsf{h}_\mathsf{t}^{\mathsf{train}}$, and is given as

$$C^{*}_{\mathsf{train},1,\mathsf{LT}} = \max_{\mathsf{h}_\mathsf{t}^{\mathsf{train}}} \left[\max_{\eta} \widehat{C}_{\mathsf{train},1}\left(\mathsf{h}_\mathsf{t}^{\mathsf{train}}, \eta, N_c, \mathsf{SNR}\right)\right]. \tag{63}$$

Performing the optimization in (63) seems difficult. Motivated by our study in Sec. III, we now assume a specific form for the threshold:

$$\mathsf{h}_\mathsf{t}^{\mathsf{train}} = \epsilon \log\left(\frac{1}{\mathsf{SNR}}\right), \;\; \epsilon \in (0,1). \tag{64}$$

It is shown in Appendix C that with this choice of $\mathsf{h}_\mathsf{t}^{\mathsf{train}}$, the optimal choice for η and N_c can be obtained in closed form and the desired result in (60) is established.

Alternatively, we now demonstrate a sub-optimal, but simpler approach that suffices to obtain (60). This approach uses the choice of η that optimizes the average mutual information in the no feedback case [13, Lemma 2]. This choice, denoted by η^*, is given as

$$\eta^* = \frac{N_c\mathsf{SNR}+N_c-1}{(N_c-2)N_c\mathsf{SNR}} \cdot \left[\sqrt{1 + \frac{N_c\mathsf{SNR}(N_c-2)}{N_c\mathsf{SNR}+N_c-1}} - 1\right]. \tag{65}$$

Let $\mathsf{h}_\mathsf{t}^{\mathsf{train},\star} \triangleq \frac{\eta^* N_c\mathsf{SNR}}{1+\eta^* N_c\mathsf{SNR}}\mathsf{h}_\mathsf{t}$ where $\mathsf{h}_\mathsf{t} \sim \lambda\log\left(\frac{1}{\mathsf{SNR}}\right)$, $\kappa_1^\star = \kappa_1\big|_{\eta^*,\,\mathsf{h}_\mathsf{t}^{\mathsf{train},\star}}$ and $\kappa_2^\star = \kappa_2\big|_{\eta^*}$. If we define,

$$A_1 = \frac{(1-\eta^*)(1+\eta^* N_c\mathsf{SNR})\,\mathsf{h}_\mathsf{t}^{\mathsf{train},\star}\mathsf{SNR}}{(1-\eta^*)\mathsf{SNR}+\kappa_1^\star\kappa_2^\star}, \tag{66}$$

$$A_2 = \frac{(1-\eta^*)(1+\eta^* N_c\mathsf{SNR})\,\mathsf{h}_\mathsf{t}^{\mathsf{train},\star}\mathsf{SNR}+(1-\eta^*)\mathsf{SNR}+\kappa_1^\star\kappa_2^\star}{\eta^*(1-\eta^*)N_c\mathsf{SNR}^2}, \tag{67}$$

it is straightforward, but cumbersome, to show that

$$\lim_{\mathsf{SNR}\to 0} A_1 = 0 \ \text{ and } \ \lim_{\mathsf{SNR}\to 0} \frac{1}{A_2} = 0 \tag{68}$$

for any $\mu > 0$. From (61), we then have

$$\begin{aligned}
\max_{\mathsf{h}_{\mathsf{t}}^{\mathsf{train}},\eta} \widehat{C}_{\mathsf{train},1,\mathsf{LT}}(\mathsf{h}_{\mathsf{t}}^{\mathsf{train}}, \eta, N_c, \mathsf{SNR}) &\geq \widehat{C}_{\mathsf{train},1,\mathsf{LT}}(\mathsf{h}_{\mathsf{t}}^{\mathsf{train},\star}, \eta^*, N_c, \mathsf{SNR}) && (69)\\
&= \kappa_1 \cdot [\log(1 + A_1) + \nu_{A_2}] && (70)\\
&\overset{(a)}{\geq} \kappa_1 \cdot \left[\log(1 + A_1) + \frac{1}{2}\log\left(1 + \frac{2}{A_2}\right)\right] && (71)\\
&\overset{(b)}{\approx} \kappa_1 \cdot \left[A_1 + \frac{1}{A_2}\right] && (72)
\end{aligned}$$

where (a) follows from (41) and (b) is the low-SNR approximation to (71). Substituting for $\mathsf{h}_{\mathsf{t}}^{\mathsf{train},\star}$ and simplifying we can reduce the lower bound in (72) to

$$\widehat{C}_{\mathsf{train},1,\mathsf{LT}}(\mathsf{SNR}) \geq (1 - \eta^*) \left(\frac{N_c}{N_c - 1}\right) \left(\frac{\eta^* N_c \mathsf{SNR}}{1 + \eta^* N_c \mathsf{SNR}}\right) [1 + \mathsf{h}_{\mathsf{t}}]\, \mathsf{SNR}. \tag{73}$$

Now, substituting for η^* from (65) and $N_c = \frac{1}{\mathsf{SNR}^{\mu}}$, it can be checked that when $\mu > 1$ the leading term is $[1 + \mathsf{h}_{\mathsf{t}}]\, \mathsf{SNR}$ which equals the first-order term of the coherent capacity as described by Corollary 1. On the other hand when $\mu < 1$, the leading term takes the form $\mathcal{O}\left(\mathsf{SNR}^{\frac{3-\mu}{2}}\right)$ and hence, $\mu > 1$ is necessary. ■

Having established the result with an average power constraint, let us consider the instantaneous power constraint case.

B. Achievable Rates under Instantaneous Power Constraint

We impose an instantaneous power constraint similar to (45) for the $(N_c - 1)D$ communication dimensions of the training scheme. Let $\widehat{C}_{\mathsf{train},1,\mathsf{ST}}(\mathsf{SNR})$ denote the achievable rate of this scheme with the short-term constraint. With the power allocation scheme similar to that in (46) (Sec. III-B), we obtain

$$\begin{aligned}
\widehat{C}_{\mathsf{train},1,\mathsf{ST}}(\mathsf{SNR}) &= \left(1 - \frac{1}{N_c}\right) \frac{1}{D} \sum_{i=1}^{D} \mathbf{E}\left[\log\left(1 + \frac{|\widehat{h}_i|^2 q_i (1 + E_{tr})}{1 + q_i + E_{tr}} \right.\right. \times \\
&\qquad \left.\left. \chi\left(\sum_{j=1}^{i} \chi(|\widehat{h}_j|^2 \geq \mathsf{h}_{\mathsf{t}}^{\mathsf{train}}) \leq \frac{ADe^{-\frac{\mathsf{h}_{\mathsf{t}}^{\mathsf{train}}(1+\eta N_c \mathsf{SNR})}{\eta N_c \mathsf{SNR}}}}{(1 - \eta)}\right)\right)\right] && (74)\\
&= \widehat{C}_{\mathsf{train},1,\mathsf{LT}}(\mathsf{SNR}) \cdot \frac{\sum_{i=1}^{D} p_i^{\mathsf{train}}}{D} && (75)
\end{aligned}$$

where $E_{tr} = \eta N_c \mathsf{SNR}$ and $p_i^{\mathsf{train}} = \Pr\left(\sum_{j=1}^{i} \chi(|\widehat{h}_j|^2 \geq \mathsf{h}_{\mathsf{t}}^{\mathsf{train}}) \leq \frac{ADe^{-\frac{\mathsf{h}_{\mathsf{t}}^{\mathsf{train}}(1+\eta N_c \mathsf{SNR})}{\eta N_c \mathsf{SNR}}}}{(1-\eta)}\right)$. Understanding when $\frac{\sum_{i=1}^{D} p_i^{\mathsf{train}}}{D} \to 1$ is similar to the case studied in Sec. III-B. Taking recourse to the analysis of Prop. 1 by using a threshold of the form $\mathsf{h}_{\mathsf{t}}^{\mathsf{train},\star} = \frac{\eta^* N_c \mathsf{SNR}}{1+\eta^* N_c \mathsf{SNR}} \mathsf{h}_{\mathsf{t}}$ where η^* is as in (65) and $\mathsf{h}_{\mathsf{t}} \sim \lambda \log\left(\frac{1}{\mathsf{SNR}}\right)$, it can be shown that the $\frac{\sum_{i=1}^{D} p_i^{\mathsf{train}}}{D}$ is lower bounded by the same expression as in (48) and (49) with A replaced by $\frac{A}{1-\eta^*}$. After some simplifications, we can conclude that if $\frac{\mathbf{E}[\mathsf{D}_{\mathsf{eff}}]}{1-\eta^*} - \mathsf{h}_{\mathsf{t}} \to \infty$, then $\widehat{C}_{\mathsf{train},1,\mathsf{ST}}(\mathsf{SNR}) \to \widehat{C}_{\mathsf{train},1,\mathsf{LT}}(\mathsf{SNR})$. Note that this condition is almost equivalent to the condition in the perfect CSI since $\eta \in (0,1)$ and we know from our earlier results on the no-feedback capacity [13] that $\eta^* \to 0$ in the wideband limit.

C. Discussion

The analysis in Sec. IV-A and IV-B reveals that the following conditions are critical:

- **C1)** The channel coherence dimension, N_c, scales with SNR according to $N_c \sim \frac{1}{\mathsf{SNR}^{\mu}}$, $\mu > 1$, and
- **C2)** The independent degrees of freedom (DoF), D, in the channel scales with SNR such that $\frac{\mathbf{E}[\mathsf{D}_{\mathsf{eff}}]}{1-\eta^*} - \mathsf{h}_{\mathsf{t}} = \frac{De^{-\mathsf{h}_{\mathsf{t}}}}{1-\eta^*} - \mathsf{h}_{\mathsf{t}} \to \infty$ 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 benchmark. When there is an instantaneous power constraint, we need to satisfy *both* **C1** and **C2** so that the benchmark can be attained.

We now study the implications of these conditions. Note that **C1** predicates a certain minimum channel coherence level to ensure the fidelity of the training performance. That is, the larger the value of μ and hence, N_c, the more easier it is to meet the benchmark. On the other hand, **C2** describes the required growth rate in the DoF, D, so that $\mathbf{E}[\mathsf{D}_{\mathsf{eff}}] - \mathsf{h}_{\mathsf{t}} \to \infty$ and the instantaneous power constraint is satisfied without any rate loss. That is, the larger the value of D, the more easier it is to meet the benchmark. It is clear that the two conditions are somewhat conflicting in nature since for a richer channel, it is easier to increase D but more difficult to increase N_c, while for a sparser channel, it is the reverse. Therefore a natural question is if they can be satisfied simultaneously.

To understand this, we first study 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 sparsities 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 under the 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 (15)). 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 (16), we obtain

$$T \sim W^{\frac{\mu}{1-\delta_1}}. \tag{76}$$

Note that as δ_1 increases and the channel gets more richer, T increases monotonically in (76).

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 function $f_1(\cdot)$, we cannot satisfy $\mu > 1$. A possible solution to overcome this difficulty is to use peaky signaling where training and communication are performed only on a subset of the D coherence subspaces. Modeling peakiness as in [4], [13] and defining $\zeta = \mathsf{SNR}^\gamma,\ \gamma > 0$ as the fraction of D over which signaling is performed, it can be shown that [13, Lemma 3] the condition for asymptotic coherence gets relaxed to $N_c = \frac{1}{\mathsf{SNR}^{\mu_{\text{peaky}}}}$ from the original $N_c = \frac{1}{\mathsf{SNR}^\mu}$ where $\mu_{\text{peaky}} = \mu + \gamma$. We require $\mu_{\text{peaky}} > 1$ which is the same as $\mu > 1 - \gamma$. For the power-law scaling in (16), we have $N_c \sim f_1(W) \sim W^{1-\delta_2} \sim \frac{1}{\mathsf{SNR}^{1-\delta_2}}$. Thus, if the peakiness coefficient γ satisfies $\gamma > \delta_2$, we can satisfy the desired condition.

B4) Delay and Doppler sparsity: Using (15), we have $W_{coh} \sim f_1(W)$ and $T_{coh} \sim f_2(T)$. Therefore, if we scale T with W according to

$$T \sim f_3(W) \quad \text{with} \quad f_3(x) = f_2^{-1}\left(\frac{x^\mu}{f_1(x)}\right), \tag{77}$$

we have $N_c = W_{coh} T_{coh} \sim f_1(W) f_2(f_3(W)) = f_1(W) f_2\left(f_2^{-1}\left(\frac{W^\mu}{f_1(W)}\right)\right) \sim \frac{1}{\mathsf{SNR}^\mu}$. Thus with $\mu > 1$ in (77), we attain the desired scaling of N_c with SNR. For the power-law scaling in (16), the desired scaling in N_c can be obtained by choosing T, W and P according to the following canonical relationship that is obtained using (16) in (77)

$$T = \frac{\left(T_m^{\delta_2} W_d^{\delta_1}\right)^{\frac{1}{1-\delta_1}} W^{\frac{\mu-1+\delta_2}{1-\delta_1}}}{P^{\frac{\mu}{1-\delta_1}}}. \tag{78}$$

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 (78). But this is necessary for achieving the benchmark capacity with an average power constraint (satisfying **C1**). We now study how this scaling law impacts the scaling of D with SNR, as in the instantaneous power case. This is critical in determining the achievability of **C2**, which we discuss next. We recall from the definition that

$$D = \frac{TW}{N_c} = TW\,\mathsf{SNR}^{\mu}. \tag{79}$$

Using (78) in (79) and simplifying, we obtain the induced scaling behavior on D with SNR as

$$D \sim \mathsf{SNR}^{\frac{\delta_1(1-\mu)-\delta_2}{1-\delta_1}}. \tag{80}$$

Therefore, we have $\mathbf{E}\left[\mathsf{D_{eff}}\right] - \mathsf{h_t}\mathsf{SNR}^{\lambda} = \mathsf{SNR}^{\lambda+\frac{\delta_1(1-\mu)-\delta_2}{1-\delta_1}} + \lambda\log(\mathsf{SNR})$ and consequently

$$\mathbf{E}\left[\mathsf{D_{eff}}\right] - \mathsf{h_t^{train}} \to \begin{cases} \infty & \text{if } 0 < \lambda < \frac{\delta_2+(\mu-1)\delta_1}{1-\delta_1} \\ -\infty & \text{if } \frac{\delta_2+(\mu-1)\delta_1}{1-\delta_1} \leq \lambda < 1. \end{cases} \tag{81}$$

It is easily seen that

$$\frac{\delta_2 + (\mu-1)\,\delta_1}{1-\delta_1} > 1 \Longleftrightarrow \mu > \frac{1-\delta_2}{\delta_1} \tag{82}$$

which yields $\mathbf{E}\left[\mathsf{D_{eff}}\right] - \mathsf{h_t} \to \infty$ for all $\lambda \in (0,1)$, and **C2** is satisfied as desired. The special cases of delay sparsity only and Doppler sparsity only (as in **B2** and **B3**) are simple extensions and follow naturally.

To summarize,

$$\mu > 1 \quad \Longrightarrow \quad \textbf{C1 is achievable} \tag{83}$$

$$\mu > \frac{1-\delta_2}{\delta_1} \quad \Longrightarrow \quad \textbf{C2 is achievable}. \tag{84}$$

Therefore,

$$\mu > \max\left(1, \frac{1-\delta_2}{\delta_1}\right) \Longrightarrow \textbf{C1 and C2 are achievable}. \tag{85}$$

We now elucidate the optimal packet configurations for different levels of channel sparsity. Analogous to the discussion in Section III-D, we focus on the power-law scaling and illustrate rules of thumb for choosing T and W for a given $N = TW$ by assuming a symmetrically sparse channel ($\delta_1 = \delta_2 = \delta$). We note the following two cases:

$$\textbf{Case 1:}\ \frac{1-\delta}{\delta} > 1 \Longleftrightarrow \delta < 0.5, \quad T \sim W^{\rho},\ \rho > \frac{1-\delta}{\delta} \tag{86}$$

$$\textbf{Case 2:}\ \frac{1-\delta}{\delta} < 1 \Longleftrightarrow \delta > 0.5, \quad T \sim W^{\rho},\ \rho > \frac{\delta}{1-\delta}. \tag{87}$$

The corresponding packet configurations are shown in Fig. 4 for $\delta \to 0$, $\delta = 0.5$ and $\delta \to 1$. It is observed that the slowest scaling in T with W is obtained for $\delta = 0.5$ when the DoF follow a *square-root* scaling law with signal space dimension. On either extreme of this square-root law, the required scaling in T with W only gets worse. This conclusion is expected and is consistent with the contradictory requirements presented by **C1** and **C2**. When $\delta < 0.5$, the channel conditions are more favorable towards scaling N_c as a function of SNR (specified by **C1**). However, the required scaling of D with SNR (specified by **C2**) is non-trivial and ultimately dominates the required scaling of T with W. On the other hand, when $\delta > 0.5$, the relatively less sparse channel conditions are favorably disposed towards the scaling of D as a function of SNR, but this is at the cost of scaling in N_c. For the case of asymmetrically sparse channels, it can be shown that this desirable condition (slowest scaling of T with W) generalizes to $\delta_1 + \delta_2 = 1$.

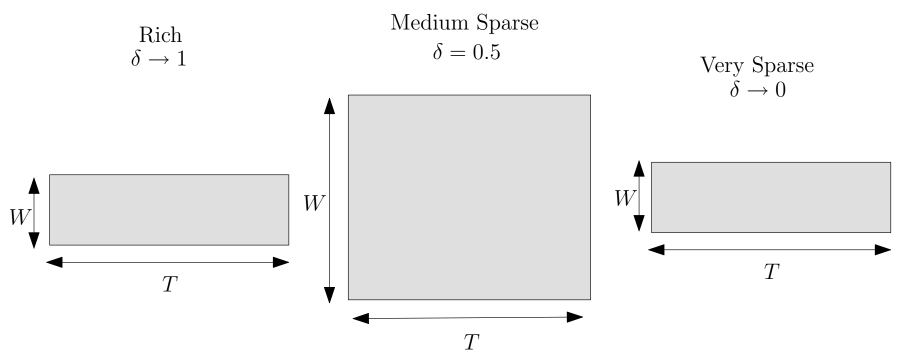


Fig. 4. *Optimal packet configurations in the training-based scenario with limited feedback. Three cases illustrated here are rich multipath* ($\delta \to 1$)*, medium sparsity* ($\delta = 0.5$) *and very high sparsity* ($\delta \to 0$).

V. Concluding Remarks

In this paper, we studied the achievable rates of short-time Fourier signaling over sparse multipath channels with limited channel state feedback. The focus of our analysis is on the wideband/low-SNR regime. Our investigation includes constraining both the average and the instantaneous transmit powers. We first analyzed the case when the receiver has perfect CSI and when one bit (per channel coefficient) of this CSI is known perfectly at the transmitter. We established conditions under which the rates achievable with this scheme approach the capacity of the benchmark scheme with perfect receiver and transmitter CSI. For sparse channels, these conditions translate to certain optimal packet configurations for signaling. When the receiver has no CSI *a priori*, we studied the performance of a training scheme with limited feedback. It is shown that with only an average power constraint, channel sparsity is necessary to attain the benchmark performance. With an instantaneous power constraint, we established conditions on optimal packet configurations in order to approach the benchmark capacity gain asymptotically as $\mathsf{SNR} \to 0$.

TABLE I

CONDITIONS NECESSARY TO ACHIEVE THE PERFECT CSI BENCHMARK OF $\log\left(\frac{1}{\mathsf{SNR}}\right)\mathsf{SNR}$.

CSI Rx.	CSI Tx.	Power Const.	Necessary Conditions	Signaling Parameters
Perf.	Perf.	-	$h_w \sim \log\left(\frac{1}{\mathsf{SNR}}\right)$	Waterfilling; see [2], [17]
Perf.	1 bit	Avg.	$\mathsf{h_t} = \lambda \log\left(\frac{1}{\mathsf{SNR}}\right), \ \lambda \to 1$	No constraints on richness or T, W; see [2], [17], [18]
Perf.	1 bit	Inst.	$\mathsf{h_t} = \lambda \log\left(\frac{1}{\mathsf{SNR}}\right)$ for $\lambda < 1$, and $\mathbf{E}\left[\mathsf{D_{eff}}\right] - \mathsf{h_t} \to \infty$	Rich channel: no constraint on T or W, Sparse (T fixed): $\lambda < \delta_2$ limits rates, Sparse (general): $T \sim W^{\rho}, \ \ \rho \geq \frac{1-\delta_2}{\delta_1}$
Train.	1 bit	Avg.	$N_c \sim \frac{1}{\mathsf{SNR}^{\mu}}, \ \mu > 1$	Rich channel: Impossible, Sparsity (Doppler): Non-peaky scheme with $T \sim W^{\frac{\mu}{1-\delta_1}}$, Sparsity (delay): Peaky scheme with peakiness coefficient $\gamma > \delta_2$, Sparsity (both): Non-peaky scheme; see (77) and (78)
Train.	1 bit	Inst.	$N_c \sim \frac{1}{\mathsf{SNR}^{\mu}}, \ \mu > 1$ and $\frac{\mathbf{E}[\mathsf{D_{eff}}]}{1-\eta^*} - \mathsf{h_t} \to \infty$	Rich channel: Impossible, Sparse (both): $\mu > \frac{1-\delta_2}{\delta_1}$ for no rate loss, else $\lambda < \frac{\delta_2 + (\mu-1)\delta_1}{1-\delta_1}$

We contrast the results of this work with recent observations in [17], [18]. The focus in [17], [18] is on training schemes and on scenarios where T_{coh} increases as SNR decreases, although no physical justification is provided for the existence of such a scaling law. 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 achieved by appropriately choosing T and W. Furthermore, while [17], [18] considered only an average power constraint, we have established achievability under both average and instantaneous power constraints. Finally, while peaky training schemes are necessary in the framework of [17], our findings here show that channel sparsity is a resource that can be exploited to obtain near-coherent performance with non-peaky training schemes. Table I provides a short summary of our contributions and places them in the context of [2], [17], [18].

Lastly, we note that the results obtained here closely parallel our earlier work [13] where we studied the achievable rates with training and no feedback. We showed that when $N_c = \frac{1}{\mathsf{SNR}^{\mu}}$ with $\mu > 1$, the (necessarily sparse) channel is *asymptotically coherent*: consistent channel estimation is possible with at a vanishing energy cost of training in the wideband limit. Analogous to [13], we have shown here that under the assumption of an error-free D-bit feedback link, the rate achievable with a training-based scheme converges to the perfect CSI scheme. Furthermore, the cost of feedback, measured in terms of the number of feedback bits per dimension (D/N) converges asymptotically to zero in a sparse channel.

APPENDIX

A. Tightness of $\widehat{C}_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR})$ to $C_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR})$ as $\mathsf{SNR} \to 0$

Let χ_i denote the random variable $\chi(|h_i|^2 \geq \mathsf{h_t})$. Defining $\gamma \triangleq \frac{\left|C_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR}) - \widehat{C}_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR})\right|}{C_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR})}$, we have

$$\gamma = \frac{1}{D}\left|\sum_{i=1}^{D} \mathbf{E}\left[\log\left(1 + \frac{\frac{TP|h_i|^2\chi_i(De^{-\mathsf{h_t}} - \sum_i \chi_i)}{\sum_i \chi_i N_c De^{-\mathsf{h_t}}}}{1 + \frac{TP|h_i|^2\chi_i}{N_c De^{-\mathsf{h_t}}}}\right)\right]\right| \tag{88}$$

$$\leq \frac{1}{D}\sum_{i=1}^{D} \mathbf{E}\left|\log\left(1 + \frac{\frac{TP|h_i|^2\chi_i(De^{-\mathsf{h_t}} - \sum_i \chi_i)}{\sum_i \chi_i N_c De^{-\mathsf{h_t}}}}{1 + \frac{TP|h_i|^2\chi_i}{N_c De^{-\mathsf{h_t}}}}\right)\right| \tag{89}$$

$$\stackrel{(a)}{\leq} \frac{1}{D}\sum_{i=1}^{D} \mathbf{E}\left|\frac{\frac{TP|h_i|^2\chi_i(De^{-\mathsf{h_t}} - \sum_i \chi_i)}{\sum_i \chi_i N_c De^{-\mathsf{h_t}}}}{1 + \frac{TP|h_i|^2\chi_i}{N_c De^{-\mathsf{h_t}}}}\right| \tag{90}$$

$$= \frac{TP}{N_c D^2 e^{-\mathsf{h_t}}}\sum_{i=1}^{D} \mathbf{E}\left[\frac{|h_i|^2\chi_i\left|De^{-\mathsf{h_t}} - \sum_i \chi_i\right|}{\sum_i \chi_i\left(1 + \frac{TP|h_i|^2\chi_i}{N_c De^{-\mathsf{h_t}}}\right)}\right] \tag{91}$$

$$\stackrel{(b)}{=} \frac{TP}{N_c De^{-\mathsf{h_t}}}\mathbf{E}\left[\frac{|h_1|^2\chi_1\left|De^{-\mathsf{h_t}} - \sum_i \chi_i\right|}{\sum_i \chi_i\left(1 + \frac{TP|h_1|^2\chi_1}{N_c De^{-\mathsf{h_t}}}\right)}\right] \triangleq \gamma_0 \tag{92}$$

where (a) follows from the log-inequality and (b) from the fact that $\{h_i\}$ are i.i.d. Conditioning on χ_1, we now have

$$\gamma_0 = \frac{TP}{N_c De^{-\mathsf{h_t}}}\mathbf{E}[\chi_1]\,\mathbf{E}_{h_1,\{\chi_j, j>1\}}\left[\frac{|h_1|^2\left|De^{-\mathsf{h_t}} - (1 + \sum_{j>1}\chi_j)\right|}{(1 + \sum_{j>1}\chi_j)\left(1 + \frac{TP|h_1|^2}{N_c De^{-\mathsf{h_t}}}\right)}\right] \tag{93}$$

$$= \mathsf{SNR} \cdot \mathbf{E}_{h_1,\{\chi_j, j>1\}}\left[\frac{|h_1|^2\left|De^{-\mathsf{h_t}} - (1 + \sum_{j>1}\chi_j)\right|}{(1 + \sum_{j>1}\chi_j)\left(1 + \frac{TP|h_1|^2}{N_c De^{-\mathsf{h_t}}}\right)}\right] \tag{94}$$

$$\stackrel{(a)}{=} \mathsf{SNR} \cdot \mathbf{E}_{h_1}\left[\frac{|h_1|^2}{1 + \frac{TP|h_1|^2}{N_c De^{-\mathsf{h_t}}}}\right] \cdot \mathbf{E}_{\{\chi_j, j>1\}}\left[\frac{\left|De^{-\mathsf{h_t}} - (1 + \sum_{j>1}\chi_j)\right|}{(1 + \sum_{j>1}\chi_j)}\right] \tag{95}$$

$$\leq \mathsf{SNR} \cdot \mathbf{E}[|h_1|^2] \cdot \mathbf{E}_{\{\chi_j, j>1\}}\left[\frac{\left|De^{-\mathsf{h_t}} - (1 + \sum_{j>1}\chi_j)\right|}{(1 + \sum_{j>1}\chi_j)}\right] \triangleq \gamma_1 \tag{96}$$

where (a) follows from the fact that h_1 and $\{\chi_j, j > 1\}$ are independent.

To show the closeness of $\widehat{C}_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR})$ to $C_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR})$, we now produce an upper bound for γ_1 that tends to 0 as $\mathsf{SNR} \to 0$. Our goal is to show that given any choice of D, $\frac{\gamma_1}{\mathsf{SNR}}$ is bounded. Consider

$$\mathbf{E}_{\{\chi_j,j>1\}}\left[\frac{\left|De^{-\mathsf{h}_\mathsf{t}}-(1+\sum_{j>1}\chi_j)\right|}{(1+\sum_{j>1}\chi_j)}\right] = \mathbf{E}_{\{\chi_j,j>1\}}\left[\left|\frac{De^{-\mathsf{h}_\mathsf{t}}}{(1+\sum_{j>1}\chi_j)}-1\right|\right]$$
$$\overset{(a)}{\leq} \underbrace{\sqrt{\mathbf{E}_{\chi_j}\left[\left(\frac{De^{-\mathsf{h}_\mathsf{t}}}{(1+\sum_{j>1}\chi_j)}\right)^2+1-2\frac{De^{-\mathsf{h}_\mathsf{t}}}{(1+\sum_{j>1}\chi_j)}\right]}}_{\triangleq\gamma_2}$$

where (a) is a consequence of Cauchy-Schwarz inequality. Let E denote $e^{-\mathsf{h}_\mathsf{t}}$. We then have

$$\gamma_2 \overset{(b)}{\leq} \sqrt{1+D^2\mathsf{E}^2\cdot\mathbf{E}_{\chi_j}\left[\frac{1}{(1+\sum_{j>1}\chi_j)^2}\right]-\frac{2D\mathsf{E}}{1+(D-1)\mathsf{E}}} \tag{97}$$

where in (b) we have used the fact that $\mathbf{E}\left[\frac{1}{\mathbf{X}}\right]\geq\frac{1}{\mathbf{E}[\mathbf{X}]}$ for a positive random variable $\mathbf{X}$. We now estimate $\alpha\triangleq\mathbf{E}_{\chi_j}\left[\frac{1}{(1+\sum_{j>1}\chi_j)^2}\right]$. It is easy to check that

$$\alpha=\sum_{i=0}^{D-1}\binom{D-1}{i}\frac{\mathsf{E}^i(1-\mathsf{E})^{D-1-i}}{(i+1)^2}. \tag{98}$$

Noting that

$$(1+y)^{D-1}=\sum_{i=0}^{D-1}\binom{D-1}{i}y^i \tag{99}$$

and integrating twice both sides of (99) with respect to y, we have

$$\frac{(1+y)^{D+1}}{D(D+1)}=\sum_{i=0}^{D-1}\binom{D-1}{i}\frac{y^{i+2}}{(i+1)(i+2)}. \tag{100}$$

Using $y=\frac{\mathsf{E}}{1-\mathsf{E}}$ in (100), we have

$$\frac{1}{D(D+1)\mathsf{E}^2}=\sum_{i=0}^{D-1}\binom{D-1}{i}\frac{\mathsf{E}^i(1-\mathsf{E})^{D-1-i}}{(i+1)(i+2)}. \tag{101}$$

Observe that $\frac{1}{(i+1)^2}\leq\frac{2}{(i+1)(i+2)}$ for all $i\geq 0$ and an upper bound for γ_2 is

$$\gamma_2\leq\sqrt{1+\frac{2D^2\mathsf{E}^2}{D(D+1)\mathsf{E}^2}-\frac{2D\mathsf{E}}{1+(D-1)\mathsf{E}}}=\sqrt{\frac{D^2\mathsf{E}-4D\mathsf{E}+3D-\mathsf{E}+1}{(D+1)(D\mathsf{E}-\mathsf{E}+1)}} \tag{102}$$

which is bounded for any choice of D. (In fact, the upper bound converges to 1 as $D\to\infty$). Note that the bound in (102) is loose and one might expect that $\frac{\gamma_1}{\mathsf{SNR}}\to 0$ as $D\to\infty$ as a consequence of the law of large numbers. However, for our purpose, the proposed loose upper bound in (102) is sufficient.

B. 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 [27, 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

$$\mathbf{E}[\mathbf{X}_i^m] \leq \frac{1}{2} m! \sigma_i^2 H^{m-2} \tag{103}$$

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

$$\Pr\left(\sum_{j=1}^{i} \chi(|h_j|^2 \geq \mathsf{h_t}) - i\mathsf{E} > (AD-i)\mathsf{E}\right) \leq \begin{cases} \exp\left(-\frac{(AD-i)\mathsf{E}}{4(1-\mathsf{E})}\right) & \text{if } i \leq \lfloor \frac{AD}{2} \rfloor, \\ \exp\left(-\frac{(AD-i)^2\mathsf{E}}{4i(1-\mathsf{E})}\right) & \text{if } i \geq \lfloor \frac{AD}{2} \rfloor + 1. \end{cases} \tag{104}$$

If $1 < A < 2$, with $\kappa = \frac{\mathsf{E}}{4(1-\mathsf{E})}$ using (104), the following lower bound, L, holds for $\frac{\sum_{i=1}^{D} p_i}{D}$:

$$\begin{aligned} L &= 1 - \left[e^{-AD\kappa} \sum_{i \leq \lfloor \frac{AD}{2} \rfloor} e^{i\kappa} + \sum_{i \geq \lfloor \frac{AD}{2} \rfloor + 1} e^{-\frac{(AD-i)^2\kappa}{i}}\right] & (105) \\ &\overset{(a)}{=} 1 - \left[\frac{e^{-\kappa(AD-1)} \cdot (e^{\kappa \lfloor \frac{AD}{2} \rfloor} - 1)}{e^{\kappa} - 1} + \left(D - \left\lfloor \frac{AD}{2} \right\rfloor\right) e^{-(A-1)^2 D\kappa}\right] & (106) \\ &\geq 1 - \left[\frac{1}{e^{\kappa} - 1} \cdot e^{-\kappa\left(\frac{AD}{2} - 1\right)} + (1 + D(1 - A/2))\, e^{-(A-1)^2 D\kappa}\right] & (107) \end{aligned}$$

where (a) follows by 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 $\frac{\sum_{i=1}^{D} p_i}{D}$:

$$L = 1 - \exp(-AD\kappa) \sum_{1 \leq i \leq D} e^{i\kappa} \approx 1 - e^{-\kappa(D(A-1)-1)} \cdot \frac{1}{e^{\kappa} - 1}. \tag{108}$$

With $\mathsf{h_t} = \lambda \log\left(\frac{1}{\mathsf{SNR}}\right)$ as in (33), 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 (107) and (108) and simplifying, we obtain the desired bounds in (48) and (49). It is also straightforward to check that when D satisfies $D\,\mathsf{SNR}^{\lambda} + \lambda\log(\mathsf{SNR}) \to \infty$ as $\mathsf{SNR} \to 0$, $L \to 1$ in both the cases. ■

C. Completing the Proof of Theorem 2

The choice of $\mathsf{h_t}$ we study is $\mathsf{h_t} = \epsilon\log\left(\frac{1}{\mathsf{SNR}}\right)$ for some $\epsilon > 0$. First, with this fixed choice of $\mathsf{h_t}$, note that maximizing $\widehat{C}_{\mathsf{train},1,\mathsf{LT}}\left(\eta, N_c, \mathsf{SNR}\right)$ is equivalent to setting its derivative (with respect to η) to zero. Then, it is straightforward to check that the derivative is

$$\underbrace{\frac{\nu_\beta \mathsf{h_t}}{\eta}}_{\mathsf{I}} + \underbrace{\frac{\mathsf{h_t}}{\eta}\log_e\left(1 + \frac{(1-\eta)(1+\eta N_c\mathsf{SNR})\mathsf{h_t}\mathsf{SNR}}{(1-\eta)\mathsf{SNR} + \kappa_1\kappa_2}\right)}_{\mathsf{II}}$$
$$+ \underbrace{\frac{\left(\nu_\beta - \frac{1}{\beta}\right)}{\mathsf{SNR}\eta}\left[\kappa_1\left(1 - \frac{1}{N_c}\right)\left(\frac{N_c\eta^2\mathsf{SNR} + 2\eta - 1}{(1-\eta)^2} + \frac{\mathsf{h_t}(1+\eta N_c\mathsf{SNR})}{\eta N_c\mathsf{SNR}(1-\eta)}\right) - \mathsf{SNR}(\mathsf{h_t}+1)\right]}_{\mathsf{III}}$$
$$+ \underbrace{\frac{\mathsf{h_t}\mathsf{SNR}^2 N_c\eta}{(1-\eta)\mathsf{SNR} + \kappa_1\kappa_2} \cdot \frac{N_c\mathsf{SNR}^2(1-\eta)^2 - \kappa_1\kappa_2(1+\eta\mathsf{SNR}N_c)\left(1 + \frac{\mathsf{h_t}(1-\eta)}{N_c\eta^2\mathsf{SNR}}\right)}{(1-\eta)\mathsf{SNR} + \kappa_1\kappa_2 + (1-\eta)(1+\eta N_c\mathsf{SNR})\mathsf{h_t}\mathsf{SNR}}}_{\mathsf{IV}}. \quad (109)$$

For simplicity, we will denote the four terms in (109) by I, II, III and IV. We will further assume that $\eta = \mathsf{SNR}^x, x \geq 0$ and $N_c = \frac{1}{\mathsf{SNR}^y}, y > 0$. For a given choice of ϵ, our goal is to determine the relationship between x and y such that the derivative in (109) can be zero. We consider three cases: i) $y > 1 + x$, ii) $y < 1 + x$ and iii) $y = 1 + x$.

Case i: First, note that $\eta N_c\mathsf{SNR} = \mathsf{SNR}^{-z}$ for some $z > 0$. The dominant terms of β can be seen to be $\frac{1}{\mathsf{SNR}^{1-\epsilon}} + \epsilon\log\left(\frac{1}{\mathsf{SNR}}\right)$ and thus, up to first order $\beta = \frac{1}{\mathsf{SNR}^{1-\epsilon}}$. Similarly, $(1-\eta)\mathsf{SNR} + \kappa_1\kappa_2$ up to first order equals $\mathsf{SNR}^{\epsilon - z}$. Note from [25, 5.1.20, p. 229] that $\nu_\beta = \mathcal{O}\left(\frac{1}{\beta}\right)$ if $\beta \to \infty$ and hence I is $\epsilon\log\left(\frac{1}{\mathsf{SNR}}\right)\frac{1}{\mathsf{SNR}^{\epsilon+x-1}}$. It can also be checked that II is $\left(\epsilon\log\left(\frac{1}{\mathsf{SNR}}\right)\right)^2\frac{1}{\mathsf{SNR}^{\epsilon+x-1}}$, $\nu_\beta - \frac{1}{\beta} = \mathcal{O}\left(\frac{1}{\beta^2}\right)$ and hence III is $\epsilon\log\left(\frac{1}{\mathsf{SNR}}\right)\frac{1}{\mathsf{SNR}^{\epsilon+x-1}}$ as long as $y < 1 + 2x$. Under the same assumption, $y < 1 + 2x$, IV is $-\left(\epsilon\log\left(\frac{1}{\mathsf{SNR}}\right)\right)^2\frac{1}{\mathsf{SNR}^{\epsilon+x-1}}$. Thus, by playing with constants the derivative can be set to zero in this case. If $y \geq 1 + 2x$, I and II remain unchanged, but III is $\mathsf{SNR}^{2+x-y-\epsilon}$ and IV is $-\epsilon\log\left(\frac{1}{\mathsf{SNR}}\right)\mathsf{SNR}^{2+x-y-\epsilon}$. By comparing the coefficients, we see that the only way the derivative can be zero is if $y = 1 + 2x$.

Case ii: In this case, the first order terms show the following behavior. With $w = 1 + x - y > 0$, I is SNR^{w-x}, II is $\epsilon\log\left(\frac{1}{\mathsf{SNR}}\right)\log\log\left(\frac{1}{\mathsf{SNR}}\right)\frac{1}{\mathsf{SNR}^x}$, III is $-\mathsf{SNR}^{2w-x}\frac{1}{\epsilon\log\left(\frac{1}{\mathsf{SNR}}\right)}$, and IV is SNR^{2-2y+x}. It can be seen that the derivative can never be zero and hence this case is ruled out.

Case iii: In this case, based on a similar analysis, we see that the derivative can again be set to zero.

Therefore, if $\epsilon \in (0,1)$, $x \geq 0$ and $1+x < y \leq 1+2x$, we have

$$\widehat{C}_{\mathsf{train},1,\mathsf{LT}}(\mathsf{SNR}) \geq \mathsf{SNR}^{\epsilon} \log\left(1 + \frac{\epsilon \log\left(\frac{1}{\mathsf{SNR}}\right)\mathsf{SNR}^{1-\epsilon}(1-\mathsf{SNR}^{x})}{1-\mathsf{SNR}^{y}}\right) + \mathsf{SNR}. \tag{110}$$

Thus, $\widehat{C}_{\mathsf{train},1,\mathsf{LT}}(\mathsf{SNR})$ is up to first order the same as $\widehat{C}_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR})$ and $C_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR})$. If $y = 1+x$ and $\eta N_c \mathsf{SNR} = a$ for some choice of a (positive, finite and independent of SNR), we need $a > \frac{\epsilon}{1-\epsilon}$ and we have

$$\widehat{C}_{\mathsf{train},1,\mathsf{LT}}(\mathsf{SNR}) \geq \mathsf{SNR}^{\frac{\epsilon(1+a)}{a}} \log\left(1 + \epsilon \mathsf{SNR}^{1-\frac{\epsilon(1+a)}{a}} \log\left(\frac{1}{\mathsf{SNR}}\right)\right) + \frac{a}{1+a}\cdot \mathsf{SNR}. \tag{111}$$

If $y < 1+x$, the training scheme is strictly sub-optimal (in the limit of SNR) from an ergodic capacity point-of-view. Putting things together, we obtain the desired condition, $\mu > 1$. ■

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, Mar. 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," *Proc. IEEE 2005 Spring Veh. Tech. Conf.*, vol. 5, pp. 3147–3151, May 2005.

[7] C. C. Chong, Y. Kim, and S. S. Lee, "A Modified S-V Clustering Channel Model for the UWB Indoor Residential Environment," *Proc. IEEE 2005 Spring Veh. Tech. Conf.*, vol. 1, pp. 58–62, May 2005.

[8] J. Tsao, D. Porrat, and D. N. C. Tse, "Prediction and Modeling for the Time-Evolving Ultra-Wideband Channel," *IEEE Journ. Selected Topics in Sig. Proc.*, vol. 1, no. 3, pp. 340–356, Oct. 2007.

[9] J. Karedal, S. Wyne, P. Almers, F. Tufvesson, and A. F. Molisch, "Statistical Analysis of the UWB Channel in an Industrial Environment," *Proc. IEEE 2004 Fall Veh. Tech. Conf.*, vol. 1, pp. 81–85, Sept. 2004.

[10] A. F. Molisch *et al.*, "A Comprehensive Standardized Model for Ultrawideband Propagation Channels," *IEEE Trans. Antennas Propagat.*, vol. 54, no. 11, pp. 3151–3166, Nov. 2006.

[11] 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.

[12] J. Foerster *et al.*, "Channel Modeling Sub-committee Report Final," *IEEE Document IEEE P802.15-02/490r1-SG3a*, 2003.

[13] V. Raghavan, G. Hariharan, and A. M. Sayeed, "Capacity of Sparse Multipath Channels in the Ultra-Wideband Regime," *IEEE Journ. Selected Topics in Sig. Proc.*, vol. 1, no. 3, pp. 357–371, Oct. 2007.

[14] 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.

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

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

[17] S. Borade and L. Zheng, "Wideband Fading Channels with Feedback," *Proc. Allerton Conf. Commun. Cont. and Comp.*, Sept. 2004, Available: [Online]. `http://web.mit.edu/lizhong/www`.

[18] M. Agarwal and M. Honig, "Wideband Channel Capacity with Training and Partial Feedback," *Proc. Allerton Conf. Commun. Cont. and Comp.*, Sept. 2005, Available: [Online]. `http://www.ece.northwestern.edu/~mh`.

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

[20] 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.

[21] 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.

[22] 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.

[23] 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.

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

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

[26] 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.

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