Capacity of Sparse Wideband Channels with Limited Feedback

Gautham Hariharan, Vasanthan Raghavan, Akbar M. Sayeed
gharihar@qualcomm.com, vasanth@uiuc.edu, akbar@engr.wisc.edu

***Abstract*— Recent results in the wideband/low-SNR regime show that even one bit of perfect feedback is sufficient to achieve the same rate of capacity scaling as in the benchmark case of perfect CSI at the transmitter. However, the capacity achieving signals are peaky in the non-coherent scenario when CSI is estimated at the receiver. Signal peakiness is related to channel coherence and 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. With perfect receiver CSI, we show that limited feedback, *even* with an instantaneous power constraint, is sufficient to achieve the benchmark capacity in sparse multipath channels. Our analysis reveals the benefits of channel sparsity in the non-coherent scenario, where we employ a training-based signaling 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 even 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 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*. From a capacity perspective, it is shown in [1] that peaky signaling schemes are necessary to achieve the wideband limit in the non-coherent regime. 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 signal space dimension. Recent work by Zheng *et al* [2] has emphasized the crucial role of channel coherence in the low SNR regime and the importance of channel learning schemes in bridging the gap between the coherent and non-coherent extremes.

Motivated by these works as well as by recent measurement campaigns for UWB [3], [4], [5], we recently introduced the notion of *multipath sparsity* as a physical source of channel coherence and proposed a channel modeling framework in [6] that captures the effect of sparsity in delay and Doppler. A key implication of sparse multipath is that the DoF in the channel scale *sub-linearly* with the signal space dimension (time-bandwidth product). This is in contrast to rich multipath, where the DoF scale linearly, and which is the assumption in most existing works. The analysis in [6] reveals the fundamental role of multipath sparsity in reducing/eliminating peaky signaling to achieve wideband capacity.

Building on the results in [6], the focus of this paper is on the impact of *channel state feedback* on the ergodic capacity of sparse wideband channels. Although earlier works (for example [7], [8] and references therein) have explored capacity with transmitter CSI, it is only recently [9], [10] that the impact of feedback in the low-SNR, non-coherent regime has received attention. In particular, it is shown in [9] 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 approximately $\log\left(\frac{1}{\mathsf{SNR}}\right)$ and is obtained with the well-known waterfilling solution [7]. More interestingly, it is shown that this gain can be achieved with limited feedback: when there is just one bit of CSI per channel coefficient at the transmitter and on-off signaling is employed. However, for both waterfilling and on-off signaling, the capacity achieving input is 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 [9] proposes peaky training and communication.

Our main focus is on the case when there is no receiver CSI *a priori* and training-based signaling is employed. Different from [9], [10], we use orthogonal short-time Fourier signaling for jointly exploiting coherence in time and frequency. The analysis is performed under two types of transmit power constraints: (i) average or long-term and (ii) instantaneous or short-term. We restrict our attention to *causal* signaling schemes. In Sec. III, we consider the coherent case where a threshold given by $\mathsf{h_t} = \lambda \log\left(\frac{1}{\mathsf{SNR}}\right)$ for any $\lambda \in (0,1)$ directly provides a measure of capacity which behaves as $(1+\mathsf{h_t})\,\mathsf{SNR}$ 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 average number of "active" coherence subspaces, the number of independent channel coefficients (degrees of freedom) that exceed the threshold in the power allocation scheme. In particular, with an instantaneous power constraint, the benchmark capacity gain is achieved when $\mathbf{E}\left[D_{\text{eff}}\right] - \mathsf{h_t} \to \infty$ as $\mathsf{SNR} \to 0$.

In Sec. IV, we discuss the achievable rates of training-based communication schemes with limited feedback. 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 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 signaling 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 any finite constraint on the instantaneous power, channel sparsity is sufficient to achieve both $\mu > 1$ *and* $\mathbf{E}\left[D_{\text{eff}}\right] - \mathsf{h_t} \to \infty$ and thus helps attain the benchmark with both average and instantaneous power constraints. Proofs of all results can be found in [11], which are omitted here due to lack of space.

II. System Model

We now briefly summarize the model developed in [6] for sparse multipath channels. Our results are based on an orthogonal short-time Fourier (STF) signaling framework [12], [13] that naturally relates multipath sparsity in delay-Doppler to coherence in time and frequency. We consider signaling over a duration T and bandwidth W.

A. Sparse Multipath Channel Modeling

We consider a virtual representation [14] of the channel in the delay-Doppler domain

$$y(t) = \sum_{\ell=0}^{L}\sum_{m=-M}^{M} h_{\ell m} x(t-\ell/W) e^{j2\pi mt/T} + w(t) \quad (1)$$

where $x(t), y(t)$ and $w(t)$ denote the input, output and noise, T_m and W_d denote the delay and Doppler spreads, and $L = \lceil T_m W \rceil$ and $M = \lceil TW_d/2 \rceil$ denote the number of resolvable delay and Doppler shifts. Distinct $h_{\ell m}$'s correspond to approximately *disjoint* subsets of physical paths and are hence approximately statistically independent. In this work, we assume that $\{h_{\ell m}\}$ are perfectly independent and that they are zero-mean Gaussian random variables.

Let D denote the number of dominant non-vanishing $\{h_{\ell m}\}$ which represent statistically independent degrees of freedom in the channel and also signifies the delay-Doppler diversity afforded by the channel [14]. We decompose D as $D = D_T D_W$ with D_T and D_W denoting the Doppler/time and frequency/delay diversity. We have

$$\begin{aligned} D &= D_T D_W \le D_{\max} = D_{T,\max} D_{W,\max} \\ D_{T,\max} &= \lceil TW_d \rceil \ , \ D_{W,\max} = \lceil T_m W \rceil \end{aligned} \quad (2)$$

where $D_{T,\max}$ and $D_{W,\max}$ denote the maximum Doppler and delay diversity, which increase linearly with T and W, respectively, and represent a *rich multipath* environment.

However, there is strong experimental evidence ([3], [4] and references therein) 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 (see (2)) and leads to sparsity in delay and Doppler. In this paper, we model multipath sparsity by a *sub-linear* scaling in D_T and D_W with T and W:

$$D_W \sim g_1(W) \ , \ D_T \sim g_2(T) \quad (3)$$

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:

$$D_T = (TW_d)^{\delta_1} \ , \quad D_W = (WT_m)^{\delta_2} \quad (4)$$

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 (3) and (4) 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, as opposed to linear scaling in rich multipath ($\delta_1 = \delta_2 = 1$).

B. Orthogonal Short-Time Fourier Signaling

We consider signaling using an orthonormal STF basis [12], [13] that is a natural generalization of OFDM for time-varying channels. Representing (1) with respect to the STF basis gives

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

where $\mathbf{w}$ represents the additive noise vector whose entries are i.i.d. $\mathcal{CN}(0,1)$, and the $N \times N$ channel matrix $\mathbf{H}$ is diagonal. The STF basis allows an intuitive block fading interpretation of the channel in terms of *time-frequency coherence subspaces* [12]. The channel matrix is partitioned as

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

with $N = TW = N_c D$ where D represents the number of statistically independent time-frequency coherence subspaces, reflecting the independent 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 above, 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 $\mathbf{E}[|h_i|^2] = 1$ [12].

Using the DoF scaling for sparse channels in (3), the scaling behavior for the coherence dimension can be computed as

$$\begin{aligned} W_{coh} &= \frac{W}{D_W} \sim f_1(W), \quad T_{coh} = \frac{T}{D_T} \sim f_2(T) \\ N_c &= W_{coh} T_{coh} \sim f_1(W) f_2(T) \end{aligned} \quad (6)$$

where T_{coh} is the *coherence time* and W_{coh} is the *coherence bandwidth* of the channel. As a consequence of the sub-linearity of g_1 and g_2 in (3), f_1 and f_2 are also sub-linear. In particular, corresponding to the power-law scaling in (4), we obtain

$$T_{coh} = T^{1-\delta_1}/W_d^{\delta_1}, \quad W_{coh} = W^{1-\delta_2}/T_m^{\delta_2}. \quad (7)$$

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.

We assume that the input symbols that form the transmit codeword $\mathbf{x}$ satisfy an average power constraint $\mathbf{E}\left[\|\mathbf{x}\|^2\right] \le PT$. Since there are $N = TW$ symbols per codeword, we define the parameter SNR as $\mathsf{SNR} = \frac{TP}{TW} = \frac{P}{W}$. In this work, the focus is on the wideband, noncoherent regime where $\mathsf{SNR} \to 0$ (as $W \to \infty$ for a fixed P). As we will see, the achievable rates are a function only of N_c and SNR. In order to analyze the low-SNR asymptotics, the following relation between N_c and SNR is used: $N_c = \mathsf{SNR}^{-\mu}$, $\mu > 0$ where the parameter μ reflects the level of channel coherence. We also assume throughout this paper that both the transmitter and the receiver have statistical CSI - so that the scaling in D and N_c are known.

III. Coherent Capacity with Limited Feedback

We first study the coherent scenario. On one extreme, with no transmitter CSI, the coherent capacity per dimension (in b/s/Hz) is

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

It can be checked that uniform power allocation, $\mathbf{Q} = \frac{TP}{N_c D}\mathbf{I}_{N_c D} = \mathsf{SNR}\,\mathbf{I}_{N_c D}$, is optimal. The low-SNR capacity (in nats/s/Hz) satisfies

$$C_{\mathsf{coh},0}(\mathsf{SNR}) \approx \mathsf{SNR} - \mathsf{SNR}^2. \tag{8}$$

On the other extreme, with perfect transmitter CSI, the optimal power allocation is waterfilling [7] over the different coherence subspaces. Here, it can be shown that the capacity scales as $\log\left(\frac{1}{\mathsf{SNR}}\right)\mathsf{SNR}$ [9]. That is, the capacity gain compared with no transmitter CSI is directly proportional to $\mathsf{h_w} \sim \log\left(\frac{1}{\mathsf{SNR}}\right)$, which is also the waterfilling threshold.

More interestingly, it is shown in [9] that this maximal gain can be achieved with just one bit of feedback per channel coefficient. In the limited feedback case, both the transmitter and the receiver have *a priori* knowledge of a common threshold, $\mathsf{h_t}$. In our setting, 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 $b_i = 1$ if $|h_i|^2 \geq \mathsf{h_t}$ and 0 otherwise. 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 $b_i = 0$. Thus, conditioned on the $\{b_i\}_{i=1}^D$, the input covariance matrix is given by

$$\mathbf{Q}\left(\{b_i\}\right) = \mathrm{diag}\big(\underbrace{q_1, \cdots, q_1}_{N_c}, \underbrace{q_2, \cdots, q_2}_{N_c}, \cdots, \underbrace{q_D, \cdots, q_D}_{N_c}\big) \tag{9}$$

$$q_i = P_{\rm o} b_i = P_{\rm o} \cdot \chi(|h_i|^2 \geq \mathsf{h_t}). \tag{10}$$

where the indicator function $\chi(\mathcal{A}) = 1$ if the event $\mathcal{A}$ is true and 0 otherwise. The choice of $P_{\rm o}$ depends on the type of power constraint and also on the choice of $\mathsf{h_t}$. To explore this further, define

$$D_{\rm eff} = \sum_{i=1}^{D} \chi(|h_i|^2 \geq \mathsf{h_t}) \tag{11}$$

that denotes the number of "active" coherence subspaces, those which exceed the threshold. It can be checked that $\mathbf{E}[D_{\rm eff}] = De^{-\mathsf{h_t}}$.

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

$$P_{\mathsf{o,nc}} = \frac{TP}{N_c D_{\rm eff}}. \tag{12}$$

The achievable rate with this power allocation is

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

The power allocation in (12) satisfies the power constraint instantaneously as well as on average. However, noncausality implies that the above scheme cannot be realized in practice. This is especially true 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 (10) depends only on b_i and $P_{\rm o}$ is independent of $\{b_i\}_{i=1}^D$. From (10), we have

$$\mathbf{E}\left[\|\mathbf{x}\|^2\right] = N_c \sum_{i=1}^{D} \mathbf{E}[q_i] \stackrel{(a)}{=} N_c P_{\rm o} \mathbf{E}[D_{\rm eff}]$$

where (a) follows from the definition of $D_{\rm eff}$. To satisfy the average power constraint, we need

$$P_{\rm o,c} = \frac{TP}{N_c \mathbf{E}[D_{\rm eff}]} = \frac{TP}{N_c D e^{-\mathsf{h_t}}}. \tag{13}$$

The achievable rate here is

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

While $P_{\rm o,c}$ satisfies the average power constraint, it may have a large instantaneous power. This is because

$$P_{\mathsf{inst,c}} = \tfrac{N_c}{T}\sum_{i=1}^{D} \tfrac{TP}{N_c D e^{-\mathsf{h_t}}}\chi\left(|h_i|^2 \geq \mathsf{h_t}\right) = \left(\tfrac{D_{\rm eff}}{De^{-\mathsf{h_t}}}\right)P. \tag{14}$$

Thus $\mathbf{E}[P_{\mathsf{inst,c}}] \leq P$, but unlike the noncausal case, $P_{\mathsf{inst,c}} \in [0, \infty)$. We will address this important issue in Sec. III-B.

A. Capacity with Average Power Constraint

The main result of this section is as follows.

Theorem 1: Given any $\lambda \in (0, 1)$, the causal signaling scheme (see (10) and (13)) satisfying the average power constraint results in

$$\lim_{\mathsf{SNR}\to 0} \frac{\left|\widehat{C}_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR}) - C_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR})\right|}{C_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR})} = 0 \tag{15}$$

by using a threshold satisfying

$$\lim_{\mathsf{SNR}\to 0} \frac{\mathsf{h_t}}{\lambda \log\left(\frac{1}{\mathsf{SNR}}\right)} = 1. \tag{16}$$

In the limit of low SNR, the capacity gain for the D-bit feedback, causal power allocation scheme over the no transmitter CSI case is

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

■

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. Note from (11) and (14) 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 (14).

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

$$P_{\mathsf{inst,c}} \overset{a.s.}{\leq} AP \tag{17}$$

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. To this end, we define q_i, $i = 1, \cdots, D$ in $\mathbf{Q}$ (see (9)) as

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

The second indicator function above checks for the constraint in (17) causally, during each time-frequency coherence slot, and allocates power only if this constraint is satisfied. It can be shown that $\widehat{C}_{\mathsf{coh},1,\mathsf{ST}}(\mathsf{SNR}) = \widehat{C}_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR}) \cdot \frac{\sum_{i=1}^{D} p_i}{D}$ where $p_i \triangleq \Pr\left(\sum_{j=1}^{i} \chi(|h_j|^2 \geq \mathsf{h_t}) \leq ADe^{-\mathsf{h_t}}\right)$ [11]. 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:* If $A > 1$ and $\mathbf{E}[D_{\text{eff}}] - \mathsf{h_t} = D\mathsf{SNR}^{\lambda} + \lambda \log(\mathsf{SNR}) \to \infty$ as $\mathsf{SNR} \to 0$, then $\widehat{C}_{\mathsf{coh},1,\mathsf{ST}}(\mathsf{SNR}) \to \widehat{C}_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR})$. ■

C. Discussion: Rich vs. Sparse Multipath

We now discuss the feasibility of satisfying the conditions in Theorem 1 and Prop. 1 when the channel is rich and when it is sparse.

A1) Rich multipath: For a rich channel, we note from (2) 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 $\mathbf{E}[D_{\text{eff}}] - \mathsf{h_t} = D\,\mathsf{SNR}^{\lambda} + \lambda \log(\mathsf{SNR}) = \mathsf{SNR}^{\lambda-1} + \lambda \log(\mathsf{SNR}) \to \infty$ 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 relation in (4), we have $D \sim T^{\delta_1} W^{\delta_2}$ and $\mathbf{E}[D_{\text{eff}}] - \mathsf{h_t} \sim T^{\delta_1}\mathsf{SNR}^{\lambda-\delta_2} + \lambda \log(\mathsf{SNR})$. For a fixed T, we have

$$\mathbf{E}[D_{\text{eff}}] - \mathsf{h_t} \to \begin{cases} \infty & 0 < \lambda < \delta_2 \\ \text{constant} & \lambda = \delta_2 \\ 0 & 1 > \lambda > \delta_2. \end{cases} \tag{18}$$

Thus although we can approach the benchmark for average power constraint, (18) 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 $\mathbf{E}[D_{\text{eff}}] - \mathsf{h_t} \sim \mathsf{SNR}^{\lambda - \delta_2 - \rho\delta_1} + \lambda \log(\mathsf{SNR})$. The asymptotic behavior is then

$$\mathbf{E}[D_{\text{eff}}] - \mathsf{h_t} \to \begin{cases} \infty & 0 < \lambda < \delta_2 + \rho\delta_1 \\ \text{constant} & \lambda = \delta_2 + \rho\delta_1 \\ 0 & 1 > \lambda > \delta_2 + \rho\delta_1. \end{cases} \tag{19}$$

We have $\delta_2 + \rho\delta_1 \geq 1 \iff \rho \geq \frac{1-\delta_2}{\delta_1}$, and in such a case $\mathbf{E}[D_{\text{eff}}] - \mathsf{h_t} \to \infty$ for all $\lambda \in (0,1)$. Thus the benchmark capacity is achievable even under an instantaneous power constraint.

IV. Feedback Capacity with Channel Learning

We now consider the more realistic case where CSI is learned at the receiver via a training scheme. 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. 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. We consider MMSE estimation at the receiver. See [6, Sec. II(c)] for details.

A. Capacity with Average Power Constraint

The following theorem describes the conditions under which the achievable rate with the training scheme converges to the coherent capacity.

***Theorem** 2:* Let $\widehat{C}_{\mathsf{train},1,\mathsf{LT}}(\mathsf{SNR})$ denote the average mutual information achievable (per-dimension) with the causal training-based scheme satisfying the average power constraint (appropriately modified versions of (10) and (13)). If $N_c = \frac{1}{\mathsf{SNR}^{\mu}}$ for some $\mu > 1$, we have

$$\lim_{\mathsf{SNR}\to 0} \frac{\widehat{C}_{\mathsf{train},1,\mathsf{LT}}(\mathsf{SNR})}{\widehat{C}_{\mathsf{coh},1,\mathsf{LT}}(\mathsf{SNR})} = 1. \tag{20}$$

■

B. Capacity with Instantaneous Power Constraint

With a constraint on the instantaneous transmit power as in (17) and the same power allocation scheme as in Sec. III-B, we have

$$\widehat{C}_{\mathsf{train},1,\mathsf{ST}}(\mathsf{SNR}) = \widehat{C}_{\mathsf{train},1,\mathsf{LT}}(\mathsf{SNR}) \cdot \frac{\sum_{i=1}^{D} p_i^{\mathsf{train}}}{D} \tag{21}$$

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 checking whether $\frac{\sum_{i=1}^{D} p_i^{\mathsf{train}}}{D} \to 1$. It can be shown that if $\mathbf{E}[D_{\text{eff}}] - \mathsf{h_t} \to \infty$ and $\mu > 1$, then $\widehat{C}_{\mathsf{train},1,\mathsf{ST}}(\mathsf{SNR}) \to \widehat{C}_{\mathsf{train},1,\mathsf{LT}}(\mathsf{SNR})$.

C. Discussion of Results

We therefore present the following two conditions for achieving the benchmark capacity in the noncoherent case.

- **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, D, in the channel scales with SNR such that $\mathbf{E}[D_{\text{eff}}] - \mathsf{h_t} = D\mathsf{SNR}^{\lambda} + \lambda \log(\mathsf{SNR}) \to \infty$ 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 benchmark - the capacity with perfect CSI. When there is an instantaneous power constraint, we need to satisfy *both* C1 and C2 so that the benchmark can be attained.

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 the instantaneous power constraint is satisfied. It is clear that the two conditions are conflicting in nature since for a richer channel, it is easier to increase D but difficult to increase N_c, while for a sparser channel, it is vice versa. Can they be satisfied simultaneously?

We first analyze the achievability of C1. We study the conditions on the channel parameters (T_m, W_d, δ_1 and δ_2) and their interaction with the signal space (T, W and P) so that $\mu > 1$ is feasible.

B1) Rich multipath: In this case $N_c = \frac{1}{T_m W_d}$ 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 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 (6)). Therefore, by scaling T with W according to $T \sim {f_2}^{-1}(W^\mu)$ and choosing $\mu > 1$, we have $N_c \sim T_{coh} \sim f_2\left({f_2}^{-1}(W^\mu)\right) \sim \mathsf{SNR}^{-\mu}$. For the power-law scaling in (7), we obtain $T \sim W^{\frac{\mu}{1-\delta_1}}$.

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 [2], [6] 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 [6, Lemma 3] that the condition for asymptotic coherence gets relaxed to $N_c = \mathsf{SNR}^{-\mu_{\mathsf{peaky}}}$ from the original $N_c = \mathsf{SNR}^{-\mu}$ where $\mu_{\mathsf{peaky}} = \mu + \gamma$. Thus now we require $\mu_{\mathsf{peaky}} > 1$, that is $\mu > 1 - \gamma$. For the power-law scaling in (7), we have $N_c \sim f_1(W) = W^{1-\delta_2} \sim \mathsf{SNR}^{-1+\delta_2}$ (that is, $\mu = 1 - \delta_2$). Thus with $\gamma > \delta_2$, we can satisfy the desired condition.

B4) Delay and Doppler sparsity: Using (6), we have $W_{coh} \sim f_1(W)$ and $T_{coh} \sim f_2(T)$. Therefore, if we scale T with W according to

$$T \sim f_3(W) \quad \text{with} \quad f_3(x) = {f_2}^{-1}\left(\frac{x^\mu}{f_1(x)}\right) \tag{22}$$

implying $N_c = W_{coh} T_{coh} \sim f_1(W) f_2(f_3(W)) = f_1(W) f_2\left({f_2}^{-1}\left(\frac{W^\mu}{f_1(W)}\right)\right) \sim \mathsf{SNR}^{-\mu}$. Thus with $\mu > 1$ in (22), we attain the desired scaling of N_c with SNR. For the power-law scaling in (7), the desired scaling in N_c can be obtained by choosing T, W and P satisfying the following canonical relationship that is obtained using (7) in (22)

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

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 (23). How does this relationship impact the scaling of D with SNR? This is critical in determining the achievability of C2, which we discuss next. Recall that $D = TW\,\mathsf{SNR}^\mu$. Using (23), we obtain $D \sim \mathsf{SNR}^{\frac{\delta_1(1-\mu)\delta_2}{1-\delta_1}}$. Therefore, we have $\mathbf{E}\left[D_{\text{eff}}\right] - \mathsf{h_t} = D\mathsf{SNR}^\lambda + \lambda \log(\mathsf{SNR}) = \mathsf{SNR}^{\lambda + \frac{\delta_1(1-\mu)-\delta_2}{1-\delta_1}} + \lambda \log(\mathsf{SNR})$ and consequently

$$\mathbf{E}\left[D_{\text{eff}}\right] - \mathsf{h_t} \rightarrow \begin{cases} \infty & 0 < \lambda < \frac{\delta_2 + (\mu-1)\delta_1}{1-\delta_1} \\ \text{constant} & \lambda = \frac{\delta_2 + (\mu-1)\delta_1}{1-\delta_1} \\ 0 & 1 > \lambda > \frac{\delta_2 + (\mu-1)\delta_1}{1-\delta_1}. \end{cases} \tag{24}$$

It is easily seen that if $\frac{\delta_2 + (\mu-1)\delta_1}{1-\delta_1} > 1$, that is, $\mu > \frac{1-\delta_2}{\delta_1}$, $\mathbf{E}\left[D_{\text{eff}}\right] - \mathsf{h_t} \rightarrow \infty$ 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 rate achievable with the training-based scheme converges to the coherent capacity and achieves the benchmark provided (i) the channel is sparse and (ii) the canonical scaling law in (23) is satisfied. With only an average power constraint, we require $\mu > 1$, whereas with an instantaneous power constraint, we require $\mu > \max\left(1, \frac{1-\delta_2}{\delta_1}\right)$.

V. Concluding Remarks

We contrast the results of this work with the conclusions in [9], [10]. The focus in [9] is on *peaky* training-based signaling 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 channel coherence scales naturally with T and W with multipath sparsity and the benchmark gain, $\log\left(\frac{1}{\mathsf{SNR}}\right)$ can always be attained by appropriately choosing T and W. Furthermore, while [9], [10] considered only an average power constraint, we have established achievability under both average and instantaneous power constraints. Our findings 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.

References

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

[2] L. Zheng, D. N. C. Tse, and M. Medard, "Channel Coherence in the Low-SNR Regime," *IEEE Trans. Inform. Theory*, Mar. 2007.

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

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

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

[6] 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.*, Oct. 2007.

[7] A. Goldsmith and P. Varaiya, "Capacity of Fading Channels with Channel Side Information," *IEEE Trans. Inform. Theory*, Nov. 1997.

[8] G. Caire, G. Tarrico, and E. Biglieri, "Optimum Power Control Over Fading Channels," *IEEE Trans. Inform. Theory*, July 1999.

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

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

[11] G. Hariharan, V. Raghavan, and A. Sayeed, "Capacity of sparse wideband channels with partial channel feedback," *Submitted to European Trans. on Telecommun.*, June 2007, Available: [Online]. `http://www.ifp.uiuc.edu/~vasanth`.

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

[13] W. Kozek, *Adaptation of Weyl–Heisenberg Frames to Underspread Environments*, *Gabor Analysis and Algorithm: Theory and Applications*, Eds. Boston, MA, Birkhäuser, 1997.

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