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 particularly focused on the impact of channel state information (CSI), more specifically the non-coherent regime, when there is no CSI at the receiver a priori. From an ergodic capacity perspective, it is argued 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 dimensions. 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 sparsity is that the DoF in the channel scale sub-linearly with the signal space dimension (time-bandwidth product). Based on this model, we investigated the ergodic capacity of training-based signaling schemes. The analysis in [6] reveals the impact of channel sparsity on channel coherence scaling and the role played by sparsity in reducing/eliminating peaky signaling to achieve wideband capacity.

Building on the results in [6], the focus of this paper is on the impact of 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 \( \log \left( \frac{SNR}{\mu} \right) \) 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 tends to be peaky (or) bursty in time, leading to a high peak-to-average power ratio, and difficulties from an implementation standpoint. The peakiness aspect is much more relevant in the non-coherent scenario, for which [9] proposes peaky training and communication.

We analyze the capacity of sparse wideband channels with limited feedback as in [9], [10]. With a sparse channel model as in [6], the main focus is on the case when there is no receiver CSI a priori and training-based signaling is employed. Unlike [9], [10], we consider 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 \( h_t = \lambda \log \left( \frac{SNR}{\mu} \right) \) for any \( \lambda \in (0,1) \) directly provides a measure of capacity which behaves as \( (1 + h_t)SNR \) in the wideband limit. Thus with \( \lambda \rightarrow 1 \), 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 \( E[D_{eff}] \), 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 \( E[D_{eff}] \rightarrow \infty \) as \( SNR \rightarrow 0 \) and we discuss its feasibility when the channel is rich or sparse.

In Sec. IV, we discuss the achievable rates of training-based signaling scheme 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 \( N_c = \frac{1}{SNR} \) for some \( \mu > 1 \), the capacity of the training-based scheme converges to the coherent capacity, the performance benchmark, in the wideband limit. Furthermore, this condition is achievable only when the channel is sparse and we provide guidelines on choosing the 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 \( E[D_{eff}] \rightarrow \infty \) 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

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

(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 $L = [T_mW]$ and $M = [TW_d/2]$ 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 $D = D_{TW} + D_T$ with $D_T$ and $D_W$ denoting the Doppler/time and frequency/diversity degree of freedom, respectively. We have

$$D = D_{TW} + D_T$$

$$D_{TW} = D_{TW,\text{max}}$$

$$D_{T,\text{max}} = [TW_d], \quad D_{W,\text{max}} = [T_mW]$$

(2)

where $D_{T,\text{max}}$ and $D_{W,\text{max}}$ 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 $W$:

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

(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:

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

(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

$$y = H x + w$$

(5)

where $w$ represents the additive noise vector whose entries are i.i.d. $CN(0, 1)$, and the $N \times N$ channel matrix $H$ 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

$$H = \text{diag}(h_{11}, \ldots, h_{1N_c}, h_{21}, \ldots, h_{2N_c}, \ldots, h_{1D_c} \ldots, h_{D_c N_c})$$

where $N = TW = N_c D$ 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 $h_{i1}, \ldots, h_{iN_c}$ 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 $E[|h_i|^2] = \sum_n E[|\beta_n|^2] = 1$ [12].

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

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

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

(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

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

(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 $x$ satisfy an average power constraint $E[\|x\|^2] \leq PT$. Since there are $N = TW$ symbols per codeword, we define the parameter SNR as $\text{SNR} = \frac{PT}{TW_T} = \frac{P}{W_T}$. In this work, the focus is on the wideband, noncoherent regime where $\text{SNR} \rightarrow 0$ (as $W \rightarrow \infty$ 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: $N_c = SNR^{-\mu}, \mu > 0$ where the parameter $\mu$ 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 now study the coherent scenario. On one extreme, with no transmitter CSI, the coherent capacity per dimension (in
The achievable rate with this power allocation is

$$C_{\text{coh,0}}(\text{SNR}) = \sup_{Q: \text{TV}(Q) \leq TP} \frac{1}{N_c D} \log_2 \det \left( I_{N_c D} + H Q H^T \right).$$

It can be checked that uniform power allocation, $Q = \frac{TP}{N_c D} I_{N_c D} = \text{SNRI}_{N_c D}$, is optimal. The low-SNR capacity satisfies

$$C_{\text{coh,0}}(\text{SNR}) \approx \log_2 \left( e \cdot \left[ \text{SNR} - \text{SNR}^2 \right] \right).$$

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

$$\log_2 \left( \frac{\text{SNR}}{\text{SNR}^2} \right) \text{SNR} [9].$$

That is, the capacity gain compared with no transmitter CSI is directly proportional to $h_b \sim \log \left( \frac{1}{\text{SNR}} \right)$, 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, $h_b$. In our setting, the receiver compares the channel strength $(|h_i|^2, i = 1, 2, \ldots, D)$ in each coherence subspace with $h_b$, and feeds back $b_i = 1$ if $|h_i|^2 \geq h_b$ 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 $b_i = 0$. Conditioned on the $\{b_i\}_{i=1}^D$, the input power allocation, which we still denote by $Q$ with a little abuse of notation, takes the form

$$Q = \text{diag} \left( \frac{q_1}{N_c}, \frac{q_2}{N_c}, \ldots, \frac{q_D}{N_c} \right).$$

(9)

The choice of $P_o$ depends on the type of power constraint and also on the nature of feedback. To explore this further, define

$$D_{\text{eff}} = \sum_{i=1}^D \chi \left( |h_i|^2 \geq h_b \right)$$

(11)

that denotes the number of “active” coherence subspaces, those which get power allocated to them. It can be checked that

$$E[D_{\text{eff}}] = D e^{-h_b}.$$ 

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,

$$P_{0,\text{nc}} = \frac{TP}{N_c D_{\text{eff}}}.$$  

(12)

The achievable rate with this power allocation is

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

The power allocation in (12) satisfies the power constraint instantaneously as well as on average. To see this, note that the instantaneous power satisfies

$$P_{\text{inst,nc}} = \frac{N_c D_{\text{eff}}}{TP} \sum_{i=1}^D q_i = \frac{N_c D_{\text{eff}}}{TP} \sum_{i=1}^D \chi \left( |h_i|^2 \geq h_b \right) = P$$

and clearly $E[P_{\text{inst,nc}}] \leq P$ as well. Noncausality means that the above scheme cannot be realized in practice. This is especially relevant in the more practical scenario when the receiver estimates the channel coefficients $\{h_i\}_{i=1}^D$ and feeds back $\{b_i\}_{i=1}^D$ based on these estimates. This motivates us to instead consider a causal power allocation scheme, one in which for all $i = 1, \ldots, D$, $q_i$ in (10) depends only on $b_i$ and $P_o$ is independent of $\{b_i\}_{i=1}^D$. From (10), we have

$$E[|x|^2] = N_c \sum_{i=1}^D E[q_i] = N_c I_{\text{eff}}[D_{\text{eff}}]$$

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

$$P_{0,c} = \frac{TP}{N_c D_{\text{eff}}} = \frac{TP}{N_c D e^{-h_b}}.$$  

(13)

The achievable rate here is

$$\hat{C}_{\text{coh,1,LT}}(\text{SNR}) = \max_{h_b} \frac{1}{D} \sum_{i=1}^D E \left[ \log_2 \left( 1 + \frac{TP}{N_c D_{\text{eff}}} |h_i|^2 \right) \chi \left( |h_i|^2 \geq h_b \right) \right].$$

While $P_{0,c}$ satisfies the average power constraint, it may have a large instantaneous power. This is because

$$P_{\text{inst,c}} = \frac{N_c D_{\text{eff}}}{TP} \sum_{i=1}^D \chi \left( |h_i|^2 \geq h_b \right) = \left( \frac{D_{\text{eff}}}{D} \right) P.$$  

(14)

Thus $E[P_{\text{inst,c}}] \leq P$, but unlike the noncausal case, $P_{\text{inst,c}} \in [0, \infty)$. 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

$$\lim_{\text{SNR} \to 0} \left[ \frac{\hat{C}_{\text{coh,1,LT}}(\text{SNR}) - C_{\text{coh,0}}(\text{SNR})}{C_{\text{coh,1,LT}}(\text{SNR})} \right] = 0$$

by using a threshold

$$\lim_{\text{SNR} \to 0} \lambda \log \left( \frac{1}{\text{SNR}} \right) = 1.$$  

(15)

The capacity gain for the $D$-bit feedback, causal power allocation scheme over the no transmitter CSI case is

$$\lim_{\text{SNR} \to 0} \frac{\hat{C}_{\text{coh,1,LT}}(\text{SNR})}{C_{\text{coh,0}}(\text{SNR})} = (1 + h_b) \left( 1 + \lambda \log \left( \frac{1}{\text{SNR}} \right) \right).$$

The capacity gain due to feedback is directly proportional to $h_b$ and the highest gain is obtained by choosing $\lambda \to 1$, and equals the perfect CSI benchmark. Note from (11) and (14) that as $D \to \infty$, $P_{\text{inst,c}} \to P$ as a consequence of the law of large numbers. However, for any large but finite $D$, $P_{\text{inst,c}}$ 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

$$P_{\text{inst,c}} \leq AP$$

where $A > 1$ and finite. With this short-term constraint, we calculate the capacity, $\hat{C}_{\text{coh,1,ST}}(\text{SNR})$, of the causal signaling
scheme. To this end, we define \( q_i, i = 1, \ldots, D \) in \( Q \) (see (9)) as
\[
q_i = P_{o,c} \chi(|h_i|^2 \geq h_i) \chi \left( \sum_{j=1}^i \chi(|h_j|^2 \geq h_j) \leq ADc^{-h_i} \right).
\]
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 \( \tilde{C}_{coh,1,LT}(SNR) = \tilde{C}_{coh,1,LT}(SNR) \cdot \sum_{i=1}^D \frac{p_i^{\text{train}}}{p_i} \) where \( p_i = \Pr \left( \sum_{j=1}^{i-1} \chi(|h_j|^2 \geq h_j) \leq ADc^{-h_i} \right) \) [11]. Thus, characterizing \( \tilde{C}_{coh,1,LT}(SNR) \) is equivalent to characterizing \( p_i \). In particular, under what condition does \( \sum_{i=1}^D \frac{p_i^{\text{train}}}{p_i} \rightarrow 1? \) This is discussed in the following proposition.

**Proposition 1:** If \( A > 1 \) and \( E[D_{\text{eff}}] = DSNR^\lambda \rightarrow \infty \) as \( \text{SNR} \rightarrow 0 \), then \( \tilde{C}_{coh,1,LT}(SNR) \rightarrow \tilde{C}_{coh,1,LT}(SNR) \).

### C. Discussion: Rich vs. Sparse Multipath

Theorem 1 implies that as \( \lambda \rightarrow 1 \), the D-bit feedback scheme approaches the performance of the perfect CSI benchmark. Furthermore, this benchmark can be attained in the wideband limit, even when there is an instantaneous power constraint. As described in Prop. 1, \( E[D_{\text{eff}}] \rightarrow \infty \) provides a sufficient condition. When we now discuss the feasibility of satisfying these conditions 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 fixed \( T \), \( D \sim \text{SNR}^{-1} \) (since \( \text{SNR} = \frac{P}{F} \)). Thus \( E[D_{\text{eff}}] = DSNR^\lambda \sim \text{SNR}^{-\lambda - 2} \). For a fixed \( T \), we have
\[
E[D_{\text{eff}}] \rightarrow \begin{cases} 
\infty & 0 < \lambda \leq \delta_2 \\
0 & 1 > \lambda > \delta_2
\end{cases}
(18)
\]
Thus although we can approach the benchmark for average power constraint, (18) suggests a cap \( \lambda \rightarrow \delta_2 \) on the highest achievable gain with an instantaneous power constraint.

**A2) Sparse multipath:** From the relation in (4), we have \( D \sim T^{\delta_2}W^{\delta_3} \) and \( E[D_{\text{eff}}] = DSNR^\lambda \sim T^{\delta_2}SNR^{\lambda - \delta_2} \). For a fixed \( T \), we have
\[
E[D_{\text{eff}}] \rightarrow \begin{cases} 
\infty & 0 < \lambda \leq \delta_2 \\
0 & 1 > \lambda > \delta_2
\end{cases}
(19)
\]
We have \( \delta_2 + \rho_{\delta_1} \geq 1 \iff \rho \geq \frac{1 - \delta_2}{\delta_1} \), and in such a case \( E[D_{\text{eff}}] \rightarrow \infty \) 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 \( \eta \) 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 \( \tilde{C}_{\text{train},1,LT}(SNR) \) 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 \( N_c = \frac{1}{\text{SNR}^\lambda} \) for some \( \mu > 1 \), we have
\[
\lim_{\text{SNR} \rightarrow 0} \frac{\tilde{C}_{\text{train},1,LT}(SNR)}{\tilde{C}_{coh,1,LT}(SNR)} = 1.
(20)
\]

**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
\[
\tilde{C}_{\text{train},1,LT}(SNR) = \tilde{C}_{\text{train},1,LT}(SNR) \cdot \sum_{i=1}^D \frac{p_i^{\text{train}}}{D}
\]
where \( p_i^{\text{train}} = \Pr \left( \sum_{j=1}^{i-1} \chi(|\hat{h}_j|^2 \geq h_i^{\text{train}}) \leq ADc^{-h_i^{\text{train}}} \right) \). Once again the problem reduces to checking whether \( \sum_{i=1}^D \frac{p_i^{\text{train}}}{D} \rightarrow 1 \). It can be shown that if \( E[D_{\text{eff}}] \rightarrow \infty \) and \( \mu > 1 \), then \( \tilde{C}_{\text{train},1,LT}(SNR) \rightarrow \tilde{C}_{\text{train},1,LT}(SNR) \).

**C. Discussion of Results**

The results in Sec. IV-A and IV-B reveal that the following two conditions are critical for achieving the benchmark capacity in the noncoherent case.

**C1)** The channel coherence dimension, \( N_c \), scales with SNR according to \( N_c \sim \frac{1}{\text{SNR}^\lambda} \) with \( \mu > 1 \), and

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

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

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 \( E[D_{\text{eff}}] = DSNR^\lambda \rightarrow \infty \) and the instantaneous power constraint is satisfied, without any loss in capacity. It is clear that the two conditions are somewhat conflicting in nature since for a richer channel, it is easier to increase \( D \) but
difficult to increase $N_c$, while for a sparser channel, it is vice versa. Can they be satisfied simultaneously?

To answer this, we first analyze the achievability of C1. What are the conditions on the channel parameters ($T_n$, $W_s$, $\delta_1$ and $\delta_2$) and how do they interact with the signal space parameters ($T$, $W$ and $P$) so that $\mu > 1$ is feasible? As we discuss next, by leveraging delay and Doppler sparsity and using peaky signaling (when necessary), $\mu > 1$ is achievable.

**B1) Rich multipath:** When the channel is rich in both delay and Doppler, $N_c = \frac{1}{\delta MW_n}$ is fixed and does not scale with $\text{SNR}$. 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) Delay Doppler sparsity only:** In this case $W_{coh} = \frac{1}{\delta MW_n}$ is fixed and the scaling in $N_c$ is only through $T_{coh} \sim f_3(T)$ (see (6)). Therefore, by scaling $T$ with $W$ according to $T \sim f_2^{-1}(W^{\mu})$ and choosing $\mu > 1$, we have $N_c \sim T_{coh} \sim f_2^{-1}(W^{\mu}) \sim \text{SNR}^{-\mu}$. For the power-law scaling in (7), we obtain $T \sim W^{-\mu / \mu}$. This solution is only valid for $\mu > 1$. We cannot satisfy the desired condition.

**B3) Delay sparsity only:** In this case, $T_{coh} = \frac{1}{\delta MW_n}$ and $N_c = W_{coh}T_{coh}$ scales with $\text{SNR}$ only through $W_{coh} \sim f_1(\text{SNR}^{\delta_2})$. 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 $\zeta = \text{SNR}^{\delta_2}$, $\gamma > 0$ 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 $\text{SNR}^{\delta_2} \gamma = \text{SNR}^{\delta_2}$ from the original $\text{SNR}^{\delta_2}$ where $\mu_{\text{peak}} = \mu + \gamma$. Thus now we require $\mu_{\text{peak}} > 1$, that is $\mu > 1 - \gamma$. For the power-law scaling in (7), we have $N_c \sim f_1(W) = W^{1-\delta_2} \sim \text{SNR}^{-\mu (1-\delta_2)}$ (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 $T_{coh} \sim f_3(T)$. Therefore, if we scale $T$ with $W$ according to

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

implying $N_c = W_{coh}T_{coh} \sim f_1(W)f_2(f_3(W)) = f_1(W)f_2\left(\frac{W^\mu}{f_1(W)}\right) \sim \text{SNR}^{-\mu}$. Thus with $\mu > 1$ in (22), we attain the desired scaling of $N_c$ with $\text{SNR}$. 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)

\[
T = \left(\frac{T_{coh}^{\delta_2}}{W_{coh}^{\delta_2}}\right)^{-\frac{1}{\mu + \delta_2}} W^{\mu - 1 + \frac{\delta_2}{\delta_1}} / P^{\frac{1}{\mu + \delta_2}}.
\]

From the above discussion, it is clear that channel sparseness is necessary and in addition we also require a specific scaling relationship between $T$ and $W$ as defined in (23). How does this scaling law impact the scaling of $D$ with $\text{SNR}$? This is critical in determining the achievability of C2, which we discuss next. Recall that $D = TW\text{SNR}^\delta$. Using (23), we obtain $D \sim \text{SNR}^\delta W^{\frac{\delta_1 + (\mu - 1)\delta_2}{\delta_1}}$. Therefore, we have $E[D_{\text{eff}}] = D \text{SNR}^\lambda = \text{SNR}^{\lambda + \frac{\delta_1 + (\mu - 1)\delta_2}{\delta_1 - 1}}$ and consequently

\[
E[D_{\text{eff}}] \rightarrow \begin{cases} \infty & 0 < \lambda < \frac{\delta_1 + (\mu - 1)\delta_2}{\delta_1 - 1} \\ \text{constant} & \lambda = \frac{\delta_1 + (\mu - 1)\delta_2}{\delta_1 - 1} \\ 0 & 1 > \lambda > \frac{\delta_1 + (\mu - 1)\delta_2}{\delta_1 - 1}. \end{cases}
\]