On Communication over Unknown Sparse Frequency-Selective Block-Fading Channels

Arun Pachai Kannu
Dept. of Electrical Engineering
Indian Institute of Technology, Madras, India 600036
Email: arunpachai@ee.iitm.ac.in

Philip Schniter
Dept. of Electrical and Computer Engineering
The Ohio State University, Columbus OH 43210.
Email: schniter@ece.osu.edu

***Abstract*—The problem of reliable communication over unknown frequency-selective block-fading channels with sparse impulse responses is considered. In particular, discrete-time impulse responses with block-fading interval N, length $L < N$, and exactly $S \le L$ non-zero coefficients are considered, where both the locations and values of non-zero coefficients change independently across blocks and are apriori unknown. Assuming that the non-zero coefficients and noise are both Gaussian, it is first shown that the ergodic noncoherent channel capacity has pre-log factor $1 - \frac{S}{N}$ for any L. Then, a pilot-aided transmission (PAT) scheme and noncoherent decoder are proposed which are capable of communicating with arbitrarily small error probability using only S pilots per fading block. Furthermore, the achievable rate of this scheme is shown to have the optimal pre-log factor, i.e., $1 - \frac{S}{N}$. Finally, a lower complexity PAT scheme is proposed, whose ϵ-achievable rate has the pre-log factor $1 - \frac{S+1}{N}$ for any $\epsilon > 0$.**[1]

I. System Model

The problem of communicating across sparse channels has recently gained significant attention through the framework of compressed channel sensing [1]. We study communication across sparse channels from an information theoretic perspective, extending existing results on non-sparse noncoherent capacity and pilot aided transmission (PAT) to the sparse case [2], [3].

In particular, we consider the problem of communicating reliably over an unknown, single-input single-output, S-sparse, L-length N-block-fading frequency-selective channel, as described by the discrete-time complex-baseband input/output model

$$y^{(k)}[n] \;=\; \sqrt{\rho} \sum_{l=0}^{L-1} h^{(k)}[l] x^{(k)}[n-l] + v^{(k)}[n], \quad (1)$$

where $k \in \{1, \ldots, K\}$ is the fading-block index, $n \in \{0, \ldots, N-1\}$ is the channel-use index, $x^{(k)}[n]$ is the transmitted signal, $y^{(k)}[n]$ is the received signal, and $v^{(k)}[n]$ is additive white Gaussian noise (AWGN). Throughout, it will be assumed that the channel length obeys $L < N$. The channel is "sparse" in the sense that exactly S of the L channel coefficients $\{h^{(k)}[l]\}_{l=0}^{L-1}$ are non-zero during each fading block k, where the indices of these non-zero coefficients, collected in the set $\mathcal{L}^{(k)}$, can change with fading block k. In the sequel, we refer to S as the "sparsity" and to the non-zero locations $\mathcal{L}^{(k)}$ as the "support."

As there are $M \triangleq \binom{L}{S}$ distinct S-element subsets of $\{0, \ldots, L-1\}$, we write this collection of subsets as $\{\mathcal{L}_i\}_{i=1}^{M}$. We then assume that the support $\mathcal{L}^{(k)}$ is drawn so that the event $\mathcal{L}^{(k)} = \mathcal{L}_i$ occurs with prior probability λ_i, where $\mathcal{L}^{(k)}$ is drawn independently of $\mathcal{L}^{(k')}$ for $k' \neq k$. We also assume that the vector $\boldsymbol{h}_{\mathsf{nz}}^{(k)} \in \mathbb{C}^S$ containing the non-zero coefficients $\{h^{(k)}[l] : l \in \mathcal{L}^{(k)}\}$ has the circular Gaussian distribution $\boldsymbol{h}_{\mathsf{nz}}^{(k)} \sim \mathcal{CN}(\boldsymbol{0}, S^{-1}\boldsymbol{I})$ with $\boldsymbol{h}_{\mathsf{nz}}^{(k)}$ independent of $\boldsymbol{h}_{\mathsf{nz}}^{(k')}$ for $k' \neq k$. Finally, we assume that $v^{(k)}[n] \sim \mathcal{CN}(0,1)$ with $v^{(k)}[n]$ independent of $v^{(k')}[n']$ for $(k', n') \neq (k, n)$. We impose the power constraint $\frac{1}{N}\sum_{n=0}^{N-1} \mathrm{E}\{|x^{(k)}[n]|^2\} = 1 \; \forall k$, so that the signal-to-noise ratio (SNR) becomes ρ in (1).

Throughout the paper, we assume that the prefix samples $\{x^{(k)}[-l]\}_{l=1}^{L-1}$ are chosen as a *cyclic prefix* (CP), i.e., $x^{(k)}[-l] = x^{(k)}[N-l]$ for $l = 1, \ldots, L-1$. Defining the vectors $\boldsymbol{y}^{(k)} \triangleq (y^{(k)}[0], \ldots, y^{(k)}[N-1])^{\mathsf{T}}$, $\boldsymbol{v}^{(k)} \triangleq (v^{(k)}[0], \ldots, v^{(k)}[N-1])^{\mathsf{T}}$, $\boldsymbol{h}^{(k)} \triangleq (h^{(k)}[0], \ldots, h^{(k)}[L-1], 0, \ldots, 0)^{\mathsf{T}} \in \mathbb{C}^N$, and $\boldsymbol{x}^{(k)} \triangleq (x^{(k)}[0], \ldots, x^{(k)}[N-1])^{\mathsf{T}}$, we get the following input-output model in the frequency domain:

$$\boldsymbol{y}_{\mathsf{f}}^{(k)} \;=\; \sqrt{\rho}\,\mathcal{D}(\boldsymbol{x}_{\mathsf{f}}^{(k)})\boldsymbol{h}_{\mathsf{f}}^{(k)} + \boldsymbol{v}_{\mathsf{f}}^{(k)}, \quad (2)$$

where $\boldsymbol{y}_{\mathsf{f}}^{(k)} \triangleq \boldsymbol{F}\boldsymbol{y}^{(k)}$, $\boldsymbol{x}_{\mathsf{f}}^{(k)} \triangleq \boldsymbol{F}\boldsymbol{x}^{(k)}$, $\boldsymbol{v}_{\mathsf{f}}^{(k)} \triangleq \boldsymbol{F}\boldsymbol{v}^{(k)}$, $\boldsymbol{h}_{\mathsf{f}}^{(k)} \triangleq \sqrt{N}\boldsymbol{F}\boldsymbol{h}^{(k)}$, and where $\boldsymbol{F}$ denotes the N-dimensional unitary discrete Fourier transform (DFT) matrix and $\mathcal{D}(\cdot)$ converts a vector to a diagonal matrix.

The non-zero tap coefficients and their locations are "unknown" in the sense that neither the transmitter nor the receiver knows the channel realization, though they both know the channel statistics.

II. Noncoherent Capacity

For the unknown *non-sparse*, L-length, N-block fading frequency-selective channel, the ergodic capacity $\mathcal{C}_{\mathsf{non\text{-}sparse}}(\rho)$, in bits per channel use, obeys [3] $\lim_{\rho\to\infty} \frac{\mathcal{C}_{\mathsf{non\text{-}sparse}}(\rho)}{\log \rho} = 1 - \frac{L}{N}$. Here and throughout the paper, the base of the logarithm equals 2. We now

[1]This work was supported in part by NSF grant CCF-1018368 and DARPA/ONR grant N66001-10-1-4090.

characterize the ergodic noncoherent capacity of the *sparse* version of this channel, focusing on the high-SNR regime, i.e., the case $\rho \to \infty$.

Theorem 1. *The ergodic noncoherent capacity $C_{\mathsf{sparse}}(\rho)$, in bits per channel use, of the S-sparse, L-length, N-block-fading channel, obeys $\lim_{\rho\to\infty} \frac{C_{\mathsf{sparse}}(\rho)}{\log \rho} = 1 - \frac{S}{N}$ for any $L < N$.*

Proof: Using the chain rule for mutual information [4], it follows straightforwardly that

$$\mathrm{I}(\boldsymbol{y}^{(k)}; \boldsymbol{x}^{(k)}) \;=\; \mathrm{I}(\boldsymbol{y}^{(k)}; \mathcal{L}^{(k)}) + \mathrm{I}(\boldsymbol{y}^{(k)}; \boldsymbol{x}^{(k)} \,|\, \mathcal{L}^{(k)}) - \mathrm{I}(\boldsymbol{y}^{(k)}; \mathcal{L}^{(k)} \,|\, \boldsymbol{x}^{(k)}). \tag{3}$$

where $\mathrm{I}(\boldsymbol{a};\boldsymbol{b})$ denotes the mutual information between random vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ and where $\mathrm{I}(\boldsymbol{a};\boldsymbol{b}\,|\,\boldsymbol{c})$ denotes the conditional mutual information between $\boldsymbol{a}$ and $\boldsymbol{b}$ conditioned on $\boldsymbol{c}$. Then, since $|\mathcal{L}^{(k)}| = M$, we can bound the first term in (3) as follows:

$$\mathrm{I}(\boldsymbol{y}^{(k)}; \mathcal{L}^{(k)}) \;\le\; \mathrm{h}(\mathcal{L}^{(k)}) \;\le\; \log|\mathcal{L}^{(k)}| \;=\; \log M,$$

where $\mathrm{h}(\boldsymbol{a})$ denotes the entropy of $\boldsymbol{a}$. Because $\mathrm{I}(\boldsymbol{y}^{(k)}; \mathcal{L}^{(k)} \,|\, \boldsymbol{x}^{(k)}) \ge 0$, we have the upper bound $\mathrm{I}(\boldsymbol{y}^{(k)}; \boldsymbol{x}^{(k)}) \le \log M + \mathrm{I}(\boldsymbol{y}^{(k)}; \boldsymbol{x}^{(k)} \,|\, \mathcal{L}^{(k)})$. Similarly, since $\mathrm{I}(\boldsymbol{y}^{(k)}; \mathcal{L}^{(k)}) \ge 0$, equation (3) implies that $\mathrm{I}(\boldsymbol{y}^{(k)}; \boldsymbol{x}^{(k)}) \ge \mathrm{I}(\boldsymbol{y}^{(k)}; \boldsymbol{x}^{(k)} \,|\, \mathcal{L}^{(k)}) - \mathrm{I}(\boldsymbol{y}^{(k)}; \mathcal{L}^{(k)} \,|\, \boldsymbol{x}^{(k)})$ and, since $\mathrm{I}(\boldsymbol{y}^{(k)}; \mathcal{L}^{(k)} \,|\, \boldsymbol{x}^{(k)}) \le \mathrm{h}(\mathcal{L}^{(k)} \,|\, \boldsymbol{x}^{(k)}) \le \log M$, we also have that $\mathrm{I}(\boldsymbol{y}^{(k)}; \boldsymbol{x}^{(k)}) \ge \mathrm{I}(\boldsymbol{y}^{(k)}; \boldsymbol{x}^{(k)} \,|\, \mathcal{L}^{(k)}) - \log M$. In summary, we have that $\mathrm{I}(\boldsymbol{y}^{(k)}; \boldsymbol{x}^{(k)}) = \mathrm{I}(\boldsymbol{y}^{(k)}; \boldsymbol{x}^{(k)} \,|\, \mathcal{L}^{(k)}) + \Delta$ for $\Delta \in \big[-\log M, \log M\big]$.

Given knowledge of the support $\mathcal{L}^{(k)}$, the frequency-domain vector $\boldsymbol{h}_{\mathsf{f}}^{(k)}$ is zero-mean Gaussian with a rank-S covariance matrix. Thus, [3, Theorem 1] implies that $\mathcal{C}_{\mathcal{L}}(\rho)$, the pre-log factor of ergodic noncoherent capacity under knowledge of the support equals $1 - \frac{S}{N}$, i.e., $\lim_{\rho\to\infty} \frac{\mathcal{C}_{\mathcal{L}}(\rho)}{\log \rho} = 1 - \frac{S}{N}$. Since $\mathcal{C}_{\mathcal{L}}(\rho) = \frac{1}{N} \max_{p(\boldsymbol{x}_{\mathsf{f}}^{(k)}):\mathrm{E}\,\|\boldsymbol{x}_{\mathsf{f}}^{(k)}\|^2 \le N} \mathrm{I}(\boldsymbol{y}_{\mathsf{f}}^{(k)}; \boldsymbol{x}_{\mathsf{f}}^{(k)} \,|\, \mathcal{L}^{(k)})$, where $\mathrm{I}(\boldsymbol{y}_{\mathsf{f}}^{(k)}; \boldsymbol{x}_{\mathsf{f}}^{(k)} \,|\, \mathcal{L}^{(k)}) = \mathrm{I}(\boldsymbol{y}^{(k)}; \boldsymbol{x}^{(k)} \,|\, \mathcal{L}^{(k)})$ and since $\mathrm{I}(\boldsymbol{y}^{(k)}; \boldsymbol{x}^{(k)} \,|\, \mathcal{L}^{(k)})$ differs from $\mathrm{I}(\boldsymbol{y}^{(k)}; \boldsymbol{x}^{(k)})$ by a bounded ρ-invariant constant Δ, the ergodic noncoherent capacity $\mathcal{C}_{\mathsf{sparse}}(\rho) = \frac{1}{N} \max_{p(\boldsymbol{x}^{(k)}):\mathrm{E}\,\|\boldsymbol{x}^{(k)}\|^2 \le N} \mathrm{I}(\boldsymbol{y}^{(k)}; \boldsymbol{x}^{(k)})$, must also obey $\lim_{\rho\to\infty} \frac{\mathcal{C}_{\mathsf{sparse}}(\rho)}{\log \rho} = 1 - \frac{S}{N}$. ∎

It is interesting to notice that the channel multiplexing gain equals $1 - \frac{S}{N}$ whether or not the support $\boldsymbol{\mathcal{L}} \triangleq (\mathcal{L}^{(1)}, \ldots, \mathcal{L}^{(K)})$ is apriori known.

III. Pilot Aided Transmission

For the non-sparse frequency-selective block-fading channel, PAT is known to be *spectrally efficient* [2], [3] in that it yields an achievable rate whose pre-log factor matches that of the noncoherent channel capacity expression. We now show, constructively, that PAT is also spectrally efficient for the *sparse* frequency-selective block-fading channel.

For our PAT scheme, we partition the frequency-domain transmission vector $\boldsymbol{x}_{\mathsf{f}}^{(k)} \in \mathbb{C}^N$ into a pilot vector $\boldsymbol{x}_{\mathsf{p}} \in \mathbb{C}^P$, created from $\{x_{\mathsf{f}}^{(k)}[n] : n \in \mathcal{N}_{\mathsf{p}}\}$, and a data vector $\boldsymbol{x}_{\mathsf{d}}^{(k)} \in \mathbb{C}^{N-P}$, created from $\{x_{\mathsf{f}}^{(k)}[n] : n \in \mathcal{N}_{\mathsf{d}}\}$. Here, $\mathcal{N}_{\mathsf{p}} \subset \{0, \ldots, N-1\}$ denotes the pilot subcarrier indices and $\mathcal{N}_{\mathsf{d}}$ denotes the corresponding data subcarrier indices, where $\mathcal{N}_{\mathsf{d}} = \{0, \ldots, N-1\} \setminus \mathcal{N}_{\mathsf{p}}$. Notice that we have allocated exactly P signal-space dimensions (per fading block) to pilots, i.e., $|\mathcal{N}_{\mathsf{p}}| = P$. For simplicity, we assume that the pilot locations $\mathcal{N}_{\mathsf{p}}$ and pilot values $\boldsymbol{x}_{\mathsf{p}}$ do not change with the fading block k, and that the pilot values are constant modulus, i.e., $|x_{\mathsf{p}}[n]| = 1$.

In the parallel subchannel model (2), we partition both $\boldsymbol{y}_{\mathsf{f}}^{(k)} \in \mathbb{C}^N$ and $\boldsymbol{v}_{\mathsf{f}}^{(k)} \in \mathbb{C}^N$ in the same way as we did $\boldsymbol{x}_{\mathsf{f}}^{(k)} \in \mathbb{C}^N$, yielding

$$\boldsymbol{y}_{\mathsf{p}}^{(k)} \;=\; \sqrt{\rho}\,\mathcal{D}(\boldsymbol{x}_{\mathsf{p}})\boldsymbol{J}_{\mathsf{p}}\boldsymbol{h}_{\mathsf{f}}^{(k)} + \boldsymbol{v}_{\mathsf{p}}^{(k)} \tag{4}$$

$$\boldsymbol{y}_{\mathsf{d}}^{(k)} \;=\; \sqrt{\rho}\,\mathcal{D}(\boldsymbol{x}_{\mathsf{d}}^{(k)})\boldsymbol{J}_{\mathsf{d}}\boldsymbol{h}_{\mathsf{f}}^{(k)} + \boldsymbol{v}_{\mathsf{d}}^{(k)}, \tag{5}$$

where $\boldsymbol{J}_{\mathsf{p}}$ is a selection matrix constructed from rows $\mathcal{N}_{\mathsf{p}}$ of the $N \times N$ identity matrix, and $\boldsymbol{J}_{\mathsf{d}}$ is constructed similarly from rows $\mathcal{N}_{\mathsf{d}}$ of the identity matrix. Another formulation for $\boldsymbol{y}_{\mathsf{p}}^{(k)}$ and $\boldsymbol{y}_{\mathsf{d}}^{(k)}$, which will be useful in the sequel, is

$$\boldsymbol{y}_{\mathsf{p}}^{(k)} \;=\; \sqrt{\rho N}\,\mathcal{D}(\boldsymbol{x}_{\mathsf{p}})\boldsymbol{F}_{\mathsf{p,true}}^{(k)}\boldsymbol{h}_{\mathsf{nz}}^{(k)} + \boldsymbol{v}_{\mathsf{p}}^{(k)} \tag{6}$$

$$\boldsymbol{y}_{\mathsf{d}}^{(k)} \;=\; \sqrt{\rho N}\,\mathcal{D}(\boldsymbol{x}_{\mathsf{d}}^{(k)})\boldsymbol{F}_{\mathsf{d,true}}^{(k)}\boldsymbol{h}_{\mathsf{nz}}^{(k)} + \boldsymbol{v}_{\mathsf{d}}^{(k)}, \tag{7}$$

where $\boldsymbol{h}_{\mathsf{nz}}^{(k)} \in \mathbb{C}^S$ is formed from the non-zero elements of $\boldsymbol{h}^{(k)}$, $\boldsymbol{F}_{\mathsf{p,true}}^{(k)}$ is formed from rows $\mathcal{N}_{\mathsf{p}}$ and columns $\mathcal{L}^{(k)}$ of the DFT matrix $\boldsymbol{F}$, and $\boldsymbol{F}_{\mathsf{d,true}}^{(k)}$ is formed from rows $\mathcal{N}_{\mathsf{d}}$ and columns $\mathcal{L}^{(k)}$ of $\boldsymbol{F}$. Notice that, because $\mathcal{L}^{(k)}$ is not apriori known to the decoder, neither are $\boldsymbol{F}_{\mathsf{p,true}}^{(k)}$ or $\boldsymbol{F}_{\mathsf{d,true}}^{(k)}$.

To achieve an arbitrarily small probability of decoding error, we construct codewords that span K blocks, where K is arbitrarily large. Thus, using $\mathfrak{C} \subset \mathbb{C}^{K(N-P)}$ to denote our codebook, we partition each codeword $\boldsymbol{x}_{\mathsf{d}} \in \mathfrak{C}$ into K data vectors, i.e., $\boldsymbol{x}_{\mathsf{d}} = [\boldsymbol{x}_{\mathsf{d}}^{(1)\mathsf{T}}, \ldots, \boldsymbol{x}_{\mathsf{d}}^{(K)\mathsf{T}}]^{\mathsf{T}}$, for use in our PAT scheme. The codewords $\boldsymbol{x}_{\mathsf{d}}$ are generated independently from a Gaussian distribution such that the $\boldsymbol{x}_{\mathsf{d}}^{(k)}$ has positive definite covariance matrix $\boldsymbol{R}_{\mathsf{d}}$ for all k and $\boldsymbol{x}_{\mathsf{d}}^{(k)}$ is independent of $\boldsymbol{x}_{\mathsf{d}}^{(k')}$ for $k \neq k'$. Denoting the number of codewords in the codebook by $|\mathfrak{C}|$, the average data rate is given by $\mathcal{R} = \frac{1}{KN} \log|\mathfrak{C}|$.

For PAT decoding, we employ a two-stage *decoupled* scheme: i) pilot-aided channel estimation and ii) coherent data-decoding based on the channel estimate. Furthermore, pilot-aided channel estimation is accomplished in a *support-hypothesized* manner. For this, we note that the support-conditional prior

$$p(\boldsymbol{h}_{\mathsf{f}}^{(k)} \,|\, \mathcal{L}^{(k)} = \mathcal{L}_i) \;=\; \mathcal{CN}\big(\boldsymbol{0}, \tfrac{N}{S}\boldsymbol{F}_i\boldsymbol{F}_i^{\mathsf{H}}\big) \tag{8}$$

is Gaussian for any sparsity S and any support hypothesis $\mathcal{L}_i$. In the sequel, we use $\boldsymbol{F}_{\mathsf{p},i} \in \mathbb{C}^{P\times S}$ to denote the matrix formed from rows $\mathcal{N}_\mathsf{p}$ of $\boldsymbol{F}_i$ (i.e., rows $\mathcal{N}_\mathsf{p}$ and columns $\mathcal{L}_i$ of $\boldsymbol{F}$).

For PAT decoding, we compute—at each fading block k—the pilot-aided MMSE estimate $\hat{\boldsymbol{h}}_{\mathsf{f},i_k}^{(k)}$ of the frequency domain channel $\boldsymbol{h}_\mathsf{f}^{(k)}$ under channel-support hypothesis $\mathcal{L}^{(k)} = \mathcal{L}_{i_k}$, for $i_k \in \{1, ..., M\}$:

$$\hat{\boldsymbol{h}}_{\mathsf{f},i_k}^{(k)} = \sqrt{\rho}\boldsymbol{F}_{i_k}\boldsymbol{F}_{\mathsf{p},i_k}^{\mathsf{H}}\big(\rho\boldsymbol{F}_{\mathsf{p},i_k}\boldsymbol{F}_{\mathsf{p},i_k}^{\mathsf{H}}+\tfrac{S}{N}\boldsymbol{I}\big)^{-1}\mathcal{D}(\boldsymbol{x}_\mathsf{p}^*)\boldsymbol{y}_\mathsf{p}^{(k)} \tag{9}$$

We then compute the weighted minimum-distance (WMD) codeword estimate

$$\hat{\boldsymbol{x}}_{\mathsf{d},\boldsymbol{i}}^{\mathsf{WMD}} = \arg\min_{\boldsymbol{x}_\mathsf{d}\in\mathfrak{C}} \sum_{k=1}^{K} \big\|\boldsymbol{Q}_{i_k}^{(k)}\big(\boldsymbol{y}_\mathsf{d}^{(k)} - \sqrt{\rho}\,\mathcal{D}(\boldsymbol{x}_\mathsf{d}^{(k)})\boldsymbol{J}_\mathsf{d}\hat{\boldsymbol{h}}_{\mathsf{f},i_k}^{(k)}\big)\big\|^2, \tag{10}$$

where $\boldsymbol{Q}_{i_k}^{(k)} \in \mathbb{C}^{N\times N}$ whitens the effective noise

$$\boldsymbol{e}_{\mathsf{d},i_k}^{(k)} \triangleq \sqrt{\rho}\,\mathcal{D}(\boldsymbol{x}_\mathsf{d}^{(k)})\boldsymbol{J}_\mathsf{d}\tilde{\boldsymbol{h}}_{\mathsf{f},i_k}^{(k)} + \boldsymbol{v}_\mathsf{d}^{(k)}, \tag{11}$$

for which $\tilde{\boldsymbol{h}}_{\mathsf{f},i_k}^{(k)}$ denotes the channel estimation error. For the achievable rate of the decoupled-decoder PAT system to grow logarithmically with ρ, the effective noise $\boldsymbol{e}_{\mathsf{d},i_k}^{(k)}$ must satisfy certain properties.

Lemma 1. *Say that N is prime. Then, for any pilot pattern $\mathcal{N}_\mathsf{p}$ such that $P \geq S$, there exists a constant C such that the channel estimation error obeys $\mathrm{E}\{\|\tilde{\boldsymbol{h}}_{\mathsf{f,p},i}^{(k)}\|^2\} \leq C\rho^{-1}$ for all $\rho > 0$ if $\mathcal{L}_i = \mathcal{L}_{\mathsf{true}}^{(k)}$, i.e., $\mathcal{L}_i$ is the true channel-support of the k^{th} block.*

Proof: We provide a proof sketch here; the full version appears in [5]. First, we make some observations about $\boldsymbol{F}_{\mathsf{p},i}$ and $\boldsymbol{F}_{\mathsf{p,true}}^{(k)}$. When N is prime, the Chebotarev theorem [6] guarantees that any square submatrix of the N-DFT matrix $\boldsymbol{F}$ will be full rank. Hence, any tall submatrix of $\boldsymbol{F}$ will also be full rank. Then, because $P \geq S$, it follows that $\boldsymbol{F}_{\mathsf{p},i} \in \mathbb{C}^{P\times S}$ will be full rank for all i, as will $\boldsymbol{F}_{\mathsf{p,true}}^{(k)}$.

We now focus on estimation of the channel impulse response $\boldsymbol{h}^{(k)}$, noting that the frequency-domain channel estimate can be obtained by taking a DFT of the impulse response estimate. Given knowledge of the correct channel support, the estimator can directly estimate the taps with non-zero variance. In particular, the pilot-aided zero-forcing estimate of $\boldsymbol{h}_{\mathsf{nz}}^{(k)}$ is given by $\frac{1}{\sqrt{\rho N}}(\boldsymbol{F}_{\mathsf{p,true}}^{(k)})^{+}\mathcal{D}(\boldsymbol{x}_\mathsf{p}^*)\boldsymbol{y}_\mathsf{p}^{(k)}$, where $(\cdot)^{+}$ denotes pseudo-inverse. The error variance of zero-forcing estimator is given by $\frac{1}{\rho N}\,\mathrm{E}\,\|(\boldsymbol{F}_{\mathsf{p,true}}^{(k)})^{+}\boldsymbol{v}_\mathsf{p}^{(k)}\|^2$. Since $\boldsymbol{F}_{\mathsf{p,true}}^{(k)}$ is full rank, $\mathrm{E}\,\|(\boldsymbol{F}_{\mathsf{p,true}}^{(k)})^{+}\boldsymbol{v}_\mathsf{p}^{(k)}\|^2 \leq C_1$ for some constant C_1. The desired result follows because the MMSE estimator has an error variance no larger than that of the zero-forcing estimator. ■

We note that N is assumed prime only to ensure certain submatrices of DFT matrix to be full rank. This assumption can be relaxed in exchange for the following restrictions on $\mathcal{N}_\mathsf{p}$ and L:

1) The set $\mathcal{N}_\mathsf{p}$ does not form a group with respect to modulo-N addition, nor a coset of a subgroup of $\{0, 1, \ldots, N-1\}$ under modulo-N addition.
2) The channel length L obeys $L < N/2$.

Furthermore, the converse of Lemma 1 also holds [5]: for the wrong support hypotheses, there exists no C such that $\mathrm{E}\{\|\tilde{\boldsymbol{h}}_{\mathsf{f,p},i}^{(k)}\|^2\} \leq C\rho^{-1}$ for all $\rho > 0$.

When the support is known correctly for all blocks, the pre-log factor of the achievable-rate of our PAT scheme is characterized by the following lemma.

Lemma 2. *Say that N is prime, and that receiver knows the correct channel support for each fading block. Then, for any pilot pattern $\mathcal{N}_\mathsf{p}$ such that $P \geq S$, the achievable rate of the support-hypothesized estimator-decoder satisfies $\lim_{\rho\to\infty}\frac{\mathcal{R}(\rho)}{\log\rho} = 1 - \frac{P}{N}$.*

Proof: We provide a proof sketch here; the full version appears in [5]. The achievable rate of WMD decoding under imperfect channel state information and Gaussian coding was studied in [7], where it was shown that the achievable rate behaves as if the effective noise (11) was Gaussian. Given correct support knowledge, and using the channel estimation error variance bound from Lemma 1, it follows that the effective signal to noise ratio grows linearly with ρ. Then, because we transmit $N - P$ independent data symbols in each block, the achievable rate obeys $\lim_{\rho\to\infty}\frac{\mathcal{R}(\rho)}{\log\rho} = 1 - \frac{P}{N}$. ■

In summary, the PAT scheme with the decoupled decoder will suffice for spectral efficient communication over the sparse frequency-selective block-fading channel *if* we can establish a reliable means of determining the correct support.

IV. Support Decoding

In this section, we consider schemes for reliably decoding the channel support for each fading block.

A. Data-Aided Support Decoding

We now construct a so-called *data-aided support decoder* (DASD) that leverages the error-detecting capabilities of the codebook $\mathfrak{C}$. To enable DASD, we assume that the transmitter embeds cyclic redundancy check (CRC) bits in the data-bit stream. Denoting the rate of information bits as R and the rate of CRC bits as δ (per channel use), it is clear that, over $m = KN$ channel uses, we transmit a total of mR bits of information and $m\delta$ bits of CRC. Then we can write the "binning function" that maps information bits to CRC bits as $\mu(\cdot)$, where

$$\mu\ :\{1, \ldots, 2^{mR}\}\ \to\ \{1, \ldots, 2^{m\delta}\}.$$

For the information message $\boldsymbol{w}$, the corresponding CRC bits are $\boldsymbol{u} = \mu(\boldsymbol{w})$, which are sometimes referred to as the "auxiliary check message." The encoder then maps the "composite message" $(\boldsymbol{w}, \boldsymbol{u})$, which contains $m(R+\delta)$ bits, to one of the $2^{m(R+\delta)}$ codewords in the codebook $\mathfrak{C}$.

The DASD procedure is defined as follows.

For each hypothesis of support index $\boldsymbol{i} = (i_1, \ldots, i_K) \in \{1, \ldots, M\}^K$,

1) Compute conditional channel estimates $\{\hat{\boldsymbol{h}}^{(k)}_{\mathsf{f},i_k}\}_{k=1}^K$ according to (9).
2) Compute the WMD codeword estimate $\hat{\boldsymbol{x}}_{\mathsf{d},\boldsymbol{i}}$ according to (11).
3) From the codeword $\hat{\boldsymbol{x}}_{\mathsf{d},\boldsymbol{i}}$, obtain the composite message $(\hat{\boldsymbol{w}}_{\boldsymbol{i}}, \hat{\boldsymbol{u}}_{\boldsymbol{i}})$
4) Perform error detection on $(\hat{\boldsymbol{w}}_{\boldsymbol{i}}, \hat{\boldsymbol{u}}_{\boldsymbol{i}})$, i.e., check if $\mu(\hat{\boldsymbol{w}}_{\boldsymbol{i}}) = \hat{\boldsymbol{u}}_{\boldsymbol{i}}$.
5) If no error is detected or there are no more hypotheses to consider, stop and declare the decoded message as $\hat{\boldsymbol{w}}_{\boldsymbol{i}}$, else continue with the next $\boldsymbol{i}$.

The asymptotic performance of DASD is characterized by the following theorem.

Theorem 2. *For the S-sparse frequency-selective N-block-fading channel with prime N, the PAT scheme, when used with S pilots and DASD, yields an achievable rate $\mathcal{R}^{\mathsf{DASD}}(\rho)$ that obeys $\lim_{\rho\to\infty} \frac{\mathcal{R}^{\mathsf{DASD}}(\rho)}{\log \rho} = 1 - \frac{S}{N}$.*

Proof: In our proof, instead of considering a specific binning function μ, we consider the error performance averaged over all possible random binning assignments and establish that the average error approaches zero. For a given support hypothesis $\boldsymbol{\mathcal{L}}_{\boldsymbol{i}}$, the DASD computes the support-conditional channel estimate and the corresponding WMD codeword estimate, from which the composite message bits $(\hat{\boldsymbol{w}}_{\boldsymbol{i}}, \hat{\boldsymbol{u}}_{\boldsymbol{i}})$ are obtained. There are two situations under which the DASD terminates, producing the final estimate $\hat{\boldsymbol{w}}^{\mathsf{DASD}} = \hat{\boldsymbol{w}}_{\boldsymbol{i}}$: i) when $\boldsymbol{i} \neq \boldsymbol{i}_{\mathsf{last}}$ and $\mu(\hat{\boldsymbol{w}}_{\boldsymbol{i}}) = \hat{\boldsymbol{u}}_{\boldsymbol{i}}$, or ii) when $\boldsymbol{i} = \boldsymbol{i}_{\mathsf{last}}$. Here we use $\boldsymbol{i}_{\mathsf{last}}$ to denote the last of the M^K hypotheses.

We now upper bound the probability that the DASD infers the wrong information bits, i.e., that $\hat{\boldsymbol{w}}^{\mathsf{DASD}} \neq \boldsymbol{w}$. Say that $\boldsymbol{i}_{\mathsf{stop}}$ denotes the value of $\boldsymbol{i}$ used to produce $\hat{\boldsymbol{w}}^{\mathsf{DASD}}$, i.e., $\hat{\boldsymbol{w}}^{\mathsf{DASD}} = \hat{\boldsymbol{w}}_{\boldsymbol{i}_{\mathsf{stop}}}$ and $\boldsymbol{i}_{\mathsf{true}}$ denote the index corresponding to the true support. We can partition the error event $\hat{\boldsymbol{w}}_{\boldsymbol{i}_{\mathsf{stop}}} \neq \boldsymbol{w}$ into three mutually exclusive events:

E1) $\boldsymbol{i}_{\mathsf{stop}} = \boldsymbol{i}_{\mathsf{true}}$ and $\hat{\boldsymbol{w}}_{\boldsymbol{i}_{\mathsf{stop}}} \neq \boldsymbol{w}$,

E2) $\boldsymbol{i}_{\mathsf{stop}} = \boldsymbol{i}_{\mathsf{last}} \neq \boldsymbol{i}_{\mathsf{true}}$ and both $\mu(\hat{\boldsymbol{w}}_{\boldsymbol{i}_{\mathsf{true}}}) \neq \hat{\boldsymbol{u}}_{\boldsymbol{i}_{\mathsf{true}}}$ and $\hat{\boldsymbol{w}}_{\boldsymbol{i}_{\mathsf{stop}}} \neq \boldsymbol{w}$.

E3) $\exists \boldsymbol{i}_{\mathsf{stop}} \notin \{\boldsymbol{i}_{\mathsf{true}}, \boldsymbol{i}_{\mathsf{last}}\}$ s.t. both $\mu(\hat{\boldsymbol{w}}_{\boldsymbol{i}_{\mathsf{stop}}}) = \hat{\boldsymbol{u}}_{\boldsymbol{i}_{\mathsf{stop}}}$ and $\hat{\boldsymbol{w}}_{\boldsymbol{i}_{\mathsf{stop}}} \neq \boldsymbol{w}$.

Notice that E1 is the event of a data-decoding error under the correct support hypothesis (i.e., $\hat{\boldsymbol{w}}_{\boldsymbol{i}_{\mathsf{true}}} \neq \boldsymbol{w}$). As long as the total rate $R + \delta$ is less than $\mathcal{R}$, the achievable rate under known-support, the probability of E1 can be made arbitrarily small. E2 characterizes the event in which the true support is falsely discarded and a data-decoding error results later (under an incorrect support hypothesis). The probability of E2 can be upper bounded by the probability of a decoding error under the correct support-hypothesis, which (like $\Pr\{\text{E1}\}$) can be made arbitrarily small for any achievable rate. E3 describes the event that both the detection of a support-error is missed and a data-decoding error results. We have, $\Pr\{\text{E3}\} = \Pr\big\{\exists\, \boldsymbol{i}_{\mathsf{stop}} \notin \{\boldsymbol{i}_{\mathsf{true}}, \boldsymbol{i}_{\mathsf{last}}\} \text{ s.t. } \mu(\hat{\boldsymbol{w}}_{\boldsymbol{i}_{\mathsf{stop}}}) = \hat{\boldsymbol{u}}_{\boldsymbol{i}_{\mathsf{stop}}} \,\big|\, \hat{\boldsymbol{w}}_{\boldsymbol{i}_{\mathsf{stop}}} \neq \boldsymbol{w}\big\} \Pr\{\hat{\boldsymbol{w}}_{\boldsymbol{i}_{\mathsf{stop}}} \neq \boldsymbol{w}\}$ which can be upper bounded as $\Pr\{\text{E3}\} \leq \sum_{\boldsymbol{i}\neq\boldsymbol{i}_{\mathsf{true}}} \Pr\{\mu(\hat{\boldsymbol{w}}_{\boldsymbol{i}}) = \hat{\boldsymbol{u}}_{\boldsymbol{i}} \,|\, \hat{\boldsymbol{w}}_{\boldsymbol{i}} \neq \boldsymbol{w}\}$. Now, to find the probability of missing a support-error, we assume that, when $\hat{\boldsymbol{w}}_{\boldsymbol{i}} \neq \boldsymbol{w}$, the auxiliary check estimate $\mu(\hat{\boldsymbol{w}}_{\boldsymbol{i}})$ is uniformly distributed over all possibilities of $\boldsymbol{u}$. This can be justified by letting the function μ be constructed by a random binning assignment of the codewords onto $2^{m\delta}$ bins, and averaging over the ensemble of random binning assignments [8]. In this case, for any $\boldsymbol{i} \neq \boldsymbol{i}_{\mathsf{true}}$, the probability of missing the detection of a support-error becomes $\Pr\{\mu(\hat{\boldsymbol{w}}_{\boldsymbol{i}}) = \hat{\boldsymbol{u}}_{\boldsymbol{i}} \,|\, \boldsymbol{i} \neq \boldsymbol{i}_{\mathsf{true}}, \hat{\boldsymbol{w}}_{\boldsymbol{i}} \neq \boldsymbol{w}\} = \frac{1}{2^{m\delta}}$, so that $\Pr\{\text{E3}\} \leq \frac{\binom{L}{S}^K}{2^{m\delta}} = \frac{\binom{L}{S}^K}{2^{KN\delta}} = \left(\frac{\binom{L}{S}}{2^{N\delta}}\right)^K$. So, when $\delta > \frac{\log\binom{L}{S}}{N}$, by choosing K large enough, we can make $\Pr\{\text{E3}\}$ averaged over all the random binning CRC assignments arbitrarily small. Notice that the rate δ sacrificed to make $\Pr\{\text{E3}\}$ arbitrarily small does not grow with SNR ρ. As long as we choose the SNR-dependent information rate $R(\rho) \leq \mathcal{R}(\rho) - \delta$, where $\mathcal{R}(\rho)$ is an achievable rate for the sparse channel with known support described in Lemma 2, the error probability that $\Pr\{\hat{\boldsymbol{w}}^{\mathsf{DASD}} \neq \boldsymbol{w}\} = \Pr\{\text{E1}\} + \Pr\{\text{E2}\} + \Pr\{\text{E3}\}$ can be made arbitrarily small. Since δ is fixed with respect to SNR, the information rate of DASD satisfies $\lim_{\rho\to\infty} \frac{R(\rho)}{\log\rho} = 1 - \frac{S}{N}$. ∎

As we have seen, the DASD achieves the optimal pre-log factor, though its complexity grows exponentially with number of fading blocks K.

B. Pilot-Aided Support Decoding

In this section, we propose a *pilot-aided support decoder* (PASD) that offers significantly reduced decoding complexity relative to DASD. Our PASD, however, requires one additional pilot dimension relative to DASD (i.e., $P = S + 1$) and is only *asymptotically reliable* (i.e., the probability of support-detection error vanishes as $\rho \to \infty$). Consider the following normalized pilot observations: $\boldsymbol{z}^{(k)}_{\mathsf{p}} \triangleq \frac{1}{\sqrt{\rho N}} \mathcal{D}(\boldsymbol{x}^*_{\mathsf{p}}) \boldsymbol{y}^{(k)}_{\mathsf{p}} = \boldsymbol{F}^{(k)}_{\mathsf{p,true}} \boldsymbol{h}^{(k)}_{\mathsf{nz}} + \frac{1}{\sqrt{\rho N}} \boldsymbol{\nu}^{(k)}_{\mathsf{p}}$, where $\boldsymbol{\nu}^{(k)}_{\mathsf{p}} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I})$. We henceforth use $\boldsymbol{\Pi}_{\mathsf{p},i} \triangleq \boldsymbol{F}_{\mathsf{p},i} (\boldsymbol{F}^{\mathsf{H}}_{\mathsf{p},i} \boldsymbol{F}_{\mathsf{p},i})^{-1} \boldsymbol{F}^{\mathsf{H}}_{\mathsf{p},i}$ to denote the matrix that projects onto the column space of $\boldsymbol{F}_{\mathsf{p},i}$, and $\boldsymbol{\Pi}^{\perp}_{\mathsf{p},i} \triangleq \boldsymbol{I} - \boldsymbol{\Pi}_{\mathsf{p},i}$ to denote its orthogonal complement.

The pilot-aided support estimator (PASE) infers the support index as that which minimizes the energy of the projection error $\boldsymbol{e}_{\mathsf{p},i}^{(k)}$:

$$\hat{i}_{\mathsf{p}}^{(k)} \triangleq \arg\min_{i\in\{1,\ldots,M\}} \|\boldsymbol{e}_{\mathsf{p},i}^{(k)}\|^2 \text{ for } \boldsymbol{e}_{\mathsf{p},i}^{(k)} \triangleq \boldsymbol{\Pi}_{\mathsf{p},i}^{\perp}\boldsymbol{z}_{\mathsf{p}}^{(k)}$$

Theorem 3. *For the S-sparse frequency-selective N-block-fading channel with prime N, and the PAT scheme with $P \geq S+1$ arbitrarily placed pilots, the probability of PASE support-detection error vanishes as $\rho \to \infty$.*

Proof: We first note that, due to the Chebotarev theorem [6], each $\boldsymbol{F}_{\mathsf{p},i} \in \mathbb{C}^{P\times S}$ is full rank when N is prime and $P \geq S+1$. Also, each column $\boldsymbol{f}$ of $\boldsymbol{F}_{\mathsf{p},i}$ is linearly independent of all columns in $\boldsymbol{F}_{\mathsf{p},j}\big|_{j\neq i}$ that are not equal to $\boldsymbol{f}$.

A PASE support-detection error results in case of the event

E4) $\exists i \neq i_{\mathsf{true}}^{(k)}$ s.t. $\|\boldsymbol{e}_{\mathsf{p},i}^{(k)}\|^2 < \|\boldsymbol{e}_{\mathsf{p},\mathsf{true}}^{(k)}\|^2$.

Now, $\Pr\{\text{E4}\} \leq \sum_{i\neq i_{\mathsf{true}}} \Pr\{\|\boldsymbol{e}_{\mathsf{p},i}^{(k)}\| < \|\boldsymbol{e}_{\mathsf{p},\mathsf{true}}^{(k)}\|\}$. Since $\|\boldsymbol{e}_{\mathsf{p},i}^{(k)}\| = \|\boldsymbol{\Pi}_{\mathsf{p},i}^{\perp}\boldsymbol{F}_{\mathsf{p},\mathsf{true}}^{(k)}\boldsymbol{h}_{\mathsf{nz}}^{(k)} + \frac{1}{\sqrt{\rho N}}\boldsymbol{\Pi}_{\mathsf{p},i}^{\perp}\boldsymbol{\nu}_{\mathsf{p}}^{(k)}\| \geq \|\boldsymbol{\Pi}_{\mathsf{p},i}^{\perp}\boldsymbol{F}_{\mathsf{p},\mathsf{true}}^{(k)}\boldsymbol{h}_{\mathsf{nz}}^{(k)}\| - \|\frac{1}{\sqrt{\rho N}}\boldsymbol{\Pi}_{\mathsf{p},i}^{\perp}\boldsymbol{\nu}_{\mathsf{p}}^{(k)}\|$, we have $\Pr\{\text{E4}\} \leq \sum_{i\neq i_{\mathsf{true}}^{(k)}} \Pr\left\{\|\boldsymbol{\Pi}_{\mathsf{p},i}^{\perp}\boldsymbol{F}_{\mathsf{p},\mathsf{true}}^{(k)}\boldsymbol{h}_{\mathsf{nz}}^{(k)}\| < \frac{1}{\sqrt{\rho N}}\|\boldsymbol{\Pi}_{\mathsf{p},i}^{\perp}\boldsymbol{\nu}_{\mathsf{p}}^{(k)}\| + \frac{1}{\sqrt{\rho N}}\|\boldsymbol{\Pi}_{\mathsf{p},\mathsf{true}}^{(k)\perp}\boldsymbol{\nu}_{\mathsf{p}}^{(k)}\|\right\}$. Since $\boldsymbol{\Pi}_{\mathsf{p},i}$ and $\boldsymbol{\Pi}_{\mathsf{p},\mathsf{true}}^{(k)}$ are projection matrices, we have $\|\boldsymbol{\Pi}_{\mathsf{p},i}^{\perp}\boldsymbol{\nu}_{\mathsf{p}}^{(k)}\| \leq \|\boldsymbol{\nu}_{\mathsf{p}}^{(k)}\|$ and $\|\boldsymbol{\Pi}_{\mathsf{p},\mathsf{true}}^{(k)\perp}\boldsymbol{\nu}_{\mathsf{p}}^{(k)}\| \leq \|\boldsymbol{\nu}_{\mathsf{p}}^{(k)}\|$ and we can further upper bound error probability as

$$\begin{aligned}&\Pr\{\text{E4}\}\\ &\leq \sum_{i\neq i_{\mathsf{true}}^{(k)}} \Pr\left\{\|\boldsymbol{\Pi}_{\mathsf{p},i}^{\perp}\boldsymbol{F}_{\mathsf{p},\mathsf{true}}^{(k)}\boldsymbol{h}_{\mathsf{nz}}^{(k)}\| < \tfrac{2}{\sqrt{\rho N}}\|\boldsymbol{\nu}_{\mathsf{p}}^{(k)}\|\right\}.\end{aligned} \quad (12)$$

Taking the SVD $\boldsymbol{\Pi}_{\mathsf{p},i}^{\perp}\boldsymbol{F}_{\mathsf{p},\mathsf{true}}^{(k)} = \boldsymbol{U}_i^{(k)}\boldsymbol{\Sigma}_i^{(k)}\boldsymbol{V}_i^{(k)\mathsf{H}}$ and defining $\boldsymbol{g}_i^{(k)} \triangleq \sqrt{S}\boldsymbol{V}_i^{(k)\mathsf{H}}\boldsymbol{h}_{\mathsf{nz}}^{(k)} \sim \mathcal{CN}(\boldsymbol{0},\boldsymbol{I})$, we can rewrite (12) as follows and upper bound further:
$\Pr\{\text{E4}\} \leq \sum_{i\neq i_{\mathsf{true}}^{(k)}} \Pr\left\{\|\boldsymbol{\Sigma}_i^{(k)}\boldsymbol{g}_i^{(k)}\|^2 < \frac{4S}{\rho N}\|\boldsymbol{\nu}_{\mathsf{p}}^{(k)}\|^2\right\}$
$\leq \sum_{i\neq i_{\mathsf{true}}^{(k)}} \Pr\left\{(\sigma_{i,0}^{(k)})^2|g_{i,0}^{(k)}|^2 < \frac{4S}{\rho N}\|\boldsymbol{\nu}_{\mathsf{p}}^{(k)}\|^2\right\}$
$\leq \sum_{i\neq i_{\mathsf{true}}^{(k)}} \Pr\left\{(\sigma_{i,0}^{(\min)})^2|g_{i,0}^{(k)}|^2 < \frac{4S}{\rho N}\|\boldsymbol{\nu}_{\mathsf{p}}^{(k)}\|^2\right\}$
$= \sum_{i\neq i_{\mathsf{true}}^{(k)}} \Pr\left\{\frac{|g_{i,0}^{(k)}|^2}{\|\boldsymbol{\nu}_{\mathsf{p}}^{(k)}\|^2} < \frac{4S}{(\sigma_{i,0}^{(\min)})^2\rho N}\right\}$. Here, $\sigma_{i,0}^{(k)}$ denotes the largest singular value in $\boldsymbol{\Sigma}_i^{(k)}$ and $\sigma_{i,0}^{(\min)} \triangleq \min_k \sigma_{i,0}^{(k)}$. Notice that at least one of the columns of $\boldsymbol{F}_{\mathsf{p},\mathsf{true}}^{(k)}$ lies outside the column space of $\boldsymbol{F}_{\mathsf{p},i}$. The projection of those columns onto the subspace orthogonal to the column space of $\boldsymbol{F}_{\mathsf{p},i}$ will be non-zero implying that $\boldsymbol{\Pi}_{\mathsf{p},i}^{\perp}\boldsymbol{F}_{\mathsf{p},\mathsf{true}}^{(k)}$ is not identical to $\boldsymbol{0}$ and hence the largest singular value $\sigma_{i,0}^{(k)} > 0, \forall k$. Since $g_{i,0}^{(k)} \sim \mathcal{CN}(0,1)$ is independent of $\boldsymbol{\nu}_{\mathsf{p}}^{(k)} \sim \mathcal{CN}(0,\boldsymbol{I})$, the random variable $F_i^{(k)} \triangleq |g_{i,0}^{(k)}|^2/\|\boldsymbol{\nu}_{\mathsf{p}}^{(k)}\|^2$ is F-distributed with parameters $(2,2P)$. Since the cumulative distribution function of an F-distributed random variable vanishes as its argument (in this case, $\frac{4S}{(\sigma_{i,0}^{(\min)})^2\rho N}$) approaches zero, the probability of a PASE error vanishes as $\rho \to \infty$. ■

For pilot-aided support decoding (PASD), we assume that the transmitter uses the PAT scheme with $P = S+1$ pilots and prime N. At the receiver, the PASE scheme is used to estimate the sparse channel support and, based on this estimate, support-conditional channel estimation and decoupled data decoding are performed. For some $\epsilon > 0$ and SNR ρ, let $\mathcal{R}_\epsilon(\rho)$ denote the information rate for which the probability of decoding error can be made less than ϵ. Lemma 3 characterizes $\mathcal{R}_\epsilon(\rho)$ for PAT with PASD.

Lemma 3. *For the S-sparse frequency-selective N-block-fading channel with prime N, the previously defined PAT scheme, when used with $S+1$ pilots and PASD, yields an ϵ-achievable rate $\mathcal{R}_\epsilon^{\mathsf{PASD}}$ that, for any $\epsilon > 0$, obeys $\lim_{\rho\to\infty}\frac{\mathcal{R}_\epsilon^{\mathsf{PASD}}(\rho)}{\log\rho} = 1 - \frac{S+1}{N}$.*

Proof: From Theorem 3 we know that, under the conditions stated in the lemma, there exists, for any $\epsilon > 0$, an SNR ρ_ϵ above which the error of PASE is less than $\frac{\epsilon}{2}$. In the case that the support hypothesis is correct, the channel estimation and decoupled decoding allow for the design of a codebook $\mathfrak{C}_{\rho,\epsilon}$ that guarantees data decoding with error probability less than $\frac{\epsilon}{2}$ at SNR ρ. Furthermore, from Lemma 2, this codebook can be designed with a rate $R_\epsilon(\rho)$ such that $\lim_{\rho\to\infty}\frac{R_\epsilon(\rho)}{\log\rho} = 1 - \frac{S+1}{N}$. Putting these together, we obtain the result of the lemma. ■

REFERENCES

[1] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, pp. 1058–1076, June 2010.

[2] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath, "On the capacity of frequency-selective channels in training-based transmission schemes," *IEEE Trans. Signal Process.*, pp. 2572–2583, Sept. 2004.

[3] Y. Liang and V. Veeravalli, "Capacity of noncoherent time-selective Rayleigh-fading channels," *IEEE Trans. Inform. Theory*, vol. 50, pp. 3095–3110, Dec. 2004.

[4] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[5] A. P. Kannu and P. Schniter, "On communication over unknown sparse frequency-selective block-fading channels," *arXiv:1006.1548*, June 2010.

[6] T. Tao, "An uncertainty principle for cyclic groups of prime order," *Math. Res. Letter*, vol. 12, pp. 121–127, 2005.

[7] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "Gaussian codes and weighted nearest neighbor decoding in fading multiple-antenna channels," *IEEE Trans. Inform. Theory*, vol. 50, pp. 1665–1686, Aug. 2004.

[8] H. El Gamal, G. Caire, and M. O. Damen, "The MIMO ARQ channel: Diversity-multiplexing-delay tradeoff," *IEEE Trans. Inform. Theory*, vol. 52, pp. 3601–3621, Aug. 2006.