

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

Arun Pachai Kannu and Philip Schniter, *Senior Member, IEEE*

Abstract—This paper considers the problem of reliable communication over discrete-time channels whose impulse responses have length L and exactly $S \leq L$ non-zero coefficients, and whose support and coefficients remain fixed over blocks of $N > L$ channel uses but change independently from block to block. Here, it is assumed that the channel's support and coefficient realizations are both unknown, although their statistics are known. Assuming Gaussian non-zero-coefficients and noise, and focusing on the high-SNR regime, it is first shown that the ergodic noncoherent channel capacity has pre-log factor $1 - \frac{S}{N}$ for any L . It is then shown that, to communicate with arbitrarily small error probability at rates in accordance with the capacity pre-log factor, it suffices to use pilot-aided orthogonal frequency-division multiplexing (OFDM) with S pilots per fading block, in conjunction with an appropriate noncoherent decoder. Since the achievability result is proven using a noncoherent decoder whose complexity grows exponentially in the number of fading blocks K , a simpler decoder, based on $S + 1$ pilots, is also proposed. Its ϵ -achievable rate is shown to have pre-log factor equal to $1 - \frac{S+1}{N}$ with the previously considered channel, while its achievable rate is shown to have pre-log factor $1 - \frac{S+1}{N}$ when the support of the block-fading channel remains fixed over time.

Index Terms—Bayes model averaging, compressed sensing, fading channels, noncoherent capacity, noncoherent communication, sparse channels.

I. INTRODUCTION

WE consider the problem of communicating reliably over an *unknown sparse* single-input single-output (SISO) frequency-selective block-fading channel that is 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 $n \in \{0, \dots, N-1\}$ is the channel-use index, $k \in \{1, \dots, K\}$ in the fading-block 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 L obeys $L < N$. The channel

is “sparse” in the sense that *exactly* S of the L channel taps $\{h^{(k)}[l]\}_{l=0}^{L-1}$ are non-zero during each fading block k , where the indices of these non-zero taps, collected in the set $\mathcal{L}^{(k)}$, can change with fading block k . We will refer to this channel as “strictly sparse” when $S < L$, and as “non-sparse” when $S = L$. Furthermore, the channel is “unknown” in the sense that the transmitter and receiver *do not* know the channel realizations, although they *do* know the channel statistics, which are described as follows.

Recalling that there are $M \triangleq \binom{L}{S}$ distinct S -element subsets of $\{0, \dots, L-1\}$, we write this collection of subsets as $\{\mathcal{L}_i\}_{i=1}^M$. We then assume that the channel 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 $\mathbf{h}_{\text{nz}}^{(k)} \in \mathbb{C}^S$ containing the non-zero taps $\{h^{(k)}[l] : l \in \mathcal{L}^{(k)}\}$ has the circular Gaussian distribution¹ $\mathbf{h}_{\text{nz}}^{(k)} \sim \mathcal{CN}(\mathbf{0}, S^{-1}\mathbf{I})$, with $\mathbf{h}_{\text{nz}}^{(k)}$ independent of $\mathbf{h}_{\text{nz}}^{(k')}$ 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 $\frac{1}{N} \sum_{n=0}^{N-1} \mathbb{E}\{|x^{(k)}[n]|^2\} = 1 \forall k$, so that the signal-to-noise ratio (SNR) becomes ρ in (1).

Our channel model is motivated by the results of recent channel-sounding experiments (e.g., [1]–[3]) which suggest that, as the communication bandwidth increases, the channel taps $\{h^{(k)}[n]\}_{n=0}^{L-1}$ become sparse in that the majority of them are “below the noise floor” [4, p. 2]. The same behavior can be seen to manifest [5] in channel taps sampled from IEEE 802.15.4a [6] “ultra wideband” propagation-path-based continuous-time impulse responses after square-root-raised-cosine pulse shaping.² Clearly, the fact that we use *exactly* zero-valued taps makes our channel model an approximation, albeit a standard one (see, e.g., [4, p. 5]). In fact, our channel model ignores many additional features³ of real-world channels in order to facilitate an information-theoretic analysis. In addition, it should be emphasized that we assume a channel with *exactly* S non-zero taps, as opposed to at most S non-zero taps, and a

¹For ease of presentation, we assume that all non-zero channel taps have equal variance. All of our results except Lemma 1 and Corollary 1 remain valid for any positive definite covariance matrix of $\mathbf{h}_{\text{nz}}^{(k)}$, and both Lemma 1 and Corollary 1 can be straightforwardly extended to the general case.

²Say that $h^{(k)}(t) = \sum_{q=1}^Q a_q e^{j\phi_q} \delta(t - \tau_q)$ is a continuous-time impulse response based on Q propagation paths. When the pulse shape $b_{\text{T}}(t)$ is used at the transmitter and $b_{\text{R}}(t)$ is used at the receiver, and the baud interval is T , the channel taps become $h^{(k)}[l] = (b_{\text{T}} * h^{(k)} * b_{\text{R}})(lT)$, where $*$ denotes convolution. For a detailed derivation, see, e.g., [5].

³For example, in practice, the active taps $\{h^{(k)}[l]\}_{l \in \mathcal{L}^{(k)}}$ and additive noise might be non-Gaussian and/or correlated within a fading block; the active taps, support, and noise might be statistically dependent and/or non-stationary across fading blocks; and the linear channel assumption might break down due to power-amplifier non-linearities.

Manuscript received June 08, 2010; revised March 17, 2011; accepted June 20, 2011. Date of current version October 07, 2011. This work was supported in part by NSF Grant CCF-1018368 and DARPA/ONR Grant N66001-10-1-4090.

A. P. Kannu is with the Department of Electrical Engineering at the Indian Institute of Technology, Chennai-600 036, Tamil Nadu, India (e-mail: arunpachai@ee.iitm.ac.in).

P. Schniter is with the Department of Electrical and Computer Engineering at The Ohio State University, Columbus, OH 43210 USA (e-mail: schniter@ece.osu.edu).

Communicated by F. Hlawatsch, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2011.2165802

TABLE I
REVIEW OF COMMONLY USED VARIABLES, WHERE $(\cdot)^{(k)}$ DENOTES DEPENDENCE ON k -TH FADING BLOCK

$\mathbf{y}^{(k)}, \mathbf{y}_f^{(k)}$	observation vector in time domain, in frequency domain
$\mathbf{h}^{(k)}, \mathbf{h}_f^{(k)}$	channel vector in time domain, in frequency domain
$\mathbf{x}^{(k)}, \mathbf{x}_f^{(k)}$	data vector in time domain, in frequency domain
$\mathbf{v}^{(k)}, \mathbf{v}_f^{(k)}$	AWGN vector in time domain, in frequency domain
$\mathbf{y}_p^{(k)}, \mathbf{y}_d^{(k)}$	pilot, data portions of in frequency-domain observation vector
$\mathbf{x}_p^{(k)}, \mathbf{x}_d^{(k)}$	pilot, data portions of in frequency-domain data vector
$\mathbf{v}_p^{(k)}, \mathbf{v}_d^{(k)}$	pilot, data portions of in frequency-domain noise vector
$\mathcal{N}_p, \mathcal{N}_d$	pilot, data subcarrier index sets
$\mathbf{h}_{nz}^{(k)}$	non-zero portion of time-domain channel vector
$\mathcal{L}^{(k)}$	set of channel-support indices for k^{th} block
\mathcal{L}_i	set of channel-support indices for i^{th} hypothesis
$\mathbf{F}_{p,true}^{(k)}$	unitary DFT matrix restricted to true columns $\mathcal{L}^{(k)}$ and rows \mathcal{N}_p
\mathbf{F}_i	unitary DFT matrix restricted to columns \mathcal{L}_i
$\mathbf{F}_{p,i}$	unitary DFT matrix restricted to pilot rows \mathcal{N}_p and columns \mathcal{L}_i
$\mathbf{F}_{d,i}$	unitary DFT matrix restricted to data rows \mathcal{N}_d and columns \mathcal{L}_i
$\hat{\mathbf{h}}_{f,p,i}^{(k)}, \tilde{\mathbf{h}}_{f,p,i}^{(k)}$	\mathcal{L}_i -conditional pilot-based MMSE estimate of $\mathbf{h}_f^{(k)}$, associated error
$\hat{\mathbf{h}}_{nz,p,i}^{(k)}, \tilde{\mathbf{h}}_{nz,p,i}^{(k)}$	\mathcal{L}_i -conditional pilot-based MMSE estimate of $\mathbf{h}_{nz}^{(k)}$, associated error
$\mathbf{e}_{d,i}^{(k)}$	\mathcal{L}_i -conditional effective noise on $\mathbf{y}_d^{(k)}$ for WMD decoding
$\mathbf{z}_p^{(k)}$	normalized pilot observations used for PASE
$\mathbf{v}_p^{(k)}$	normalized AWGN on $\mathbf{z}_p^{(k)}$ used for PASE
$\mathbf{e}_{p,i}^{(k)}$	\mathcal{L}_i -conditional projection error vector used for PASE

decoder that knows the channel statistics perfectly (including S , L , $\{\lambda_i\}_{i=1}^M$, and ρ).

Notation: Above and in the sequel, we use lowercase boldface quantities to denote vectors, uppercase boldface quantities to denote matrices, and we use \mathbf{I} to denote the identity matrix. Also, we use $(\cdot)^T$ to denote transpose, $(\cdot)^*$ conjugate, $(\cdot)^H$ conjugate transpose, $(\cdot)^+$ pseudo-inverse, and $\mathcal{D}(\mathbf{b})$ the diagonal matrix created from vector \mathbf{b} . Furthermore, \odot element-wise multiplication, $\|\mathbf{x}\| \triangleq \sqrt{\mathbf{x}^H \mathbf{x}}$, and $\|\mathbf{x}\|_A \triangleq \sqrt{\mathbf{x}^H \mathbf{A} \mathbf{x}}$ for Hermitian positive semi-definite \mathbf{A} . Throughout, “log” denotes the base-2 logarithm. For random variables, we use $\mathbb{E}\{\cdot\}$ to denote expectation, $\text{cov}\{\mathbf{b}\}$ auto-covariance, $h(\mathbf{a})$ entropy, and $\mathbb{I}(\mathbf{a}, \mathbf{b})$ the mutual information between \mathbf{a} and \mathbf{b} . Finally, we write $\mathcal{CN}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \triangleq (\pi^N \det(\boldsymbol{\Sigma}))^{-1} \exp(-\|\mathbf{x} - \boldsymbol{\mu}\|_{\boldsymbol{\Sigma}^{-1}}^2)$ for the circular Gaussian pdf with mean $\boldsymbol{\mu} \in \mathbb{C}^N$ and positive definite covariance matrix $\boldsymbol{\Sigma}$, and we write $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ to indicate that random vector \mathbf{x} has this pdf. In Table I we list commonly used quantities, along with their definitions.

A. Preliminaries

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 $l = 1, \dots, L-1$.

In this case, we can write the k th block observations $\mathbf{y}^{(k)} \triangleq (y^{(k)}[0], \dots, y^{(k)}[N-1])^T$ as

$$\mathbf{y}^{(k)} = \sqrt{\rho} \mathbf{X}^{(k)} \mathbf{h}^{(k)} + \mathbf{v}^{(k)}, \quad (2)$$

where $\mathbf{v}^{(k)} \triangleq (v^{(k)}[0], \dots, v^{(k)}[N-1])^T$, $\mathbf{h}^{(k)} \triangleq (h^{(k)}[0], \dots, h^{(k)}[L-1], 0, \dots, 0)^T \in \mathbb{C}^N$, and $\mathbf{X}^{(k)} \in \mathbb{C}^{N \times N}$ is the circulant matrix with first column $\mathbf{x}^{(k)} \triangleq (x^{(k)}[0], \dots, x^{(k)}[N-1])^T$. An equivalent model results⁴ from converting all signals into the frequency domain:

$$\mathbf{y}_f^{(k)} = \sqrt{\rho} \mathcal{D}(\mathbf{x}_f^{(k)}) \mathbf{h}_f^{(k)} + \mathbf{v}_f^{(k)}, \quad (3)$$

where $\mathbf{y}_f^{(k)} \triangleq \mathbf{F} \mathbf{y}^{(k)}$, $\mathbf{x}_f^{(k)} \triangleq \mathbf{F} \mathbf{x}^{(k)}$, $\mathbf{v}_f^{(k)} \triangleq \mathbf{F} \mathbf{v}^{(k)}$, $\mathbf{h}_f^{(k)} \triangleq \sqrt{N} \mathbf{F} \mathbf{h}^{(k)}$, and where \mathbf{F} denotes the N -dimensional unitary discrete Fourier transform (DFT) matrix. Noting that $\mathbf{v}_f^{(k)} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, the model (3) establishes that, when viewed in the frequency domain, the frequency-selective channel (2) reduces to a set of N non-interfering scalar subchannels with average⁵

⁴Model (3) follows directly from (2) using the fact that $\mathbf{X}^{(k)} = \mathbf{F}^H \mathcal{D}(\sqrt{N} \mathbf{F} \mathbf{x}^{(k)}) \mathbf{F}$.

⁵The average subchannel SNR of ρ follows from the fact that $\frac{1}{N} \mathbb{E}\{\|\mathbf{h}_f^{(k)}\|^2\} = 1$.

subchannel SNR ρ . Although the subchannels are non-interfering, the subchannel gains within the k th block (i.e., the elements of the vector $\mathbf{h}_f^{(k)}$) are correlated in a way that depends on the channel support $\mathcal{L}^{(k)}$, as will be detailed in the sequel. For capacity analysis, we assume that the number of fading blocks K is arbitrarily large, and we ignore overhead due to the prefix, consistent with [7], [8]. Some implications of this choice are discussed below.

B. Existing Results on Noncoherent Channel Capacity

Much is known about the fundamental limits of reliable communication over the unknown *non-sparse* channel in the high-SNR regime (i.e., $\rho \rightarrow \infty$). For example, assuming that communication occurs over an arbitrarily large number of fading blocks K , the ergodic capacity $C_{\text{non-sparse}}(\rho)$, in bits per channel use, obeys [7], [8]

$$\lim_{\rho \rightarrow \infty} \frac{C_{\text{non-sparse}}(\rho)}{\log \rho} = 1 - \frac{L}{N}. \quad (4)$$

In other words, the “multiplexing gain” [9] of the non-sparse channel (i.e., the pre-log factor in its ergodic capacity expression) equals $1 - \frac{L}{N}$. Furthermore, it is possible to achieve this multiplexing gain using pilot aided transmission (PAT), which uses L signal-space dimensions of each fading block to transmit a known pilot signal and the remaining $N - L$ dimensions to transmit the data [7], [8]. In the sequel, we use the term “spectrally efficient” to describe a communication scheme whose achievable rate expression has a pre-log factor matching that of the channel’s ergodic capacity expression (i.e., the channel’s multiplexing gain).

C. Our Contributions

In this paper, we study the fundamental limits of reliable communication over the unknown *sparse* channel (1) in the high-SNR regime. First, we show that the ergodic capacity $C_{\text{sparse}}(\rho)$ obeys

$$\lim_{\rho \rightarrow \infty} \frac{C_{\text{sparse}}(\rho)}{\log \rho} = 1 - \frac{S}{N} \quad (5)$$

for any sparsity S such that $1 \leq S \leq L < N$. Comparing (5) to (4), it is interesting to notice that the channel’s multiplexing gain depends on the number of non-zero taps S and not the channel length L , even though the locations of these taps $\mathcal{L} \triangleq (\mathcal{L}^{(1)}, \dots, \mathcal{L}^{(K)})$ are unknown. Second, we show that the sparse frequency-selective block-fading channel admits spectrally efficient PAT, just as its non-sparse variant does. In other words, for an S -sparse channel, one can construct a PAT scheme that uses only S pilots per fading block to attain an achievable rate that grows with SNR ρ at the maximum possible rate, regardless of the channel length L . We establish this result constructively, by specifying a particular OFDM-based PAT scheme and a corresponding decoder, which—as we will see—can be interpreted as a joint channel-support/data decoder. Because our decoder is computationally demanding (e.g., it requires the evaluation of up to $|\mathcal{L}| = M^K = \mathcal{O}(L^{SK})$ support hypotheses), we also consider a much simpler PAT decoder and

find that its ϵ -achievable-rate has a pre-log factor of $1 - \frac{S+1}{N}$, for any error-rate $\epsilon > 0$.

In stating the above pre-log factors, we emphasize that the overhead due to the OFDM prefix has been ignored (for consistency with [7], [8]). If, instead, the overhead was included, then the pre-log factor of the non-sparse channel’s ergodic capacity (4) would read as $\frac{N-L}{N+L-1}$, and that for the sparse channel (5) would read as $\frac{N-S}{N+L-1}$. Although the increase in pre-log factor resulting from channel sparsity, i.e., $\frac{L-S}{N+L-1}$, is not as pronounced as when the prefix is ignored, i.e., $\frac{L-S}{N}$, the two values are very similar when $N \gg L - 1$, which is the typical case in practice.

D. Relation to Compressed Channel Sensing

The problem of communicating over sparse channels has recently gained a significant amount of attention through the framework of compressed channel sensing (CCS), as seen by the recent overview article [4] and the long list of papers cited therein. In CCS, it is assumed that pilots are embedded during transmission, and that channel estimation is performed using pilot-only observations (i.e., without the aid or interference from data). CCS then exploits channel sparsity to reduce the number of pilots needed for accurate channel estimation, in the hopes of increasing spectral efficiency. As an example, for the N -subcarrier OFDM scenario described by (3), CCS results [4] show that, when $P = \mathcal{O}(S_{\max} \ln^5 N)$ pilot subcarriers are chosen uniformly at random, any *deterministic* L -length channel $\mathbf{h}^{(k)}$ with sparsity *at most* S_{\max} yields a CCS estimate $\hat{\mathbf{h}}_{\text{CCS}}^{(k)}$ such that

$$\|\hat{\mathbf{h}}_{\text{CCS}}^{(k)} - \mathbf{h}^{(k)}\|^2 \leq C \frac{S_{\max} N \ln L}{\rho P} \text{ with high probability,} \quad (6)$$

where C is a constant. The success probability in (6) grows with L and N , but not with SNR ρ (see [4] for details). Furthermore, in the special case that the observations are noise-free, it is known that exactly $2S$ data-free observations are both necessary and sufficient for perfect recovery [10].

In comparing the CCS approach to the approach that we have taken, we notice that the two are fundamentally different. For example, CCS yields guarantees on the performance of channel estimation, but not on the rate of reliable communication. Also, CCS attacks the channel estimation problem using a “non-random parameter estimation” framework, whereas we approach channel estimation using a “random parameter estimation” framework, since we consider ergodic capacity and achievable rate, and are thus interested in average channel estimation performance. A potential weakness to the CCS approach is that it uses only pilot observations for channel estimation, even though the data-dependent observations contain valuable information about the unknown channel; our work (and related empirical results in [5], [11], [12]) suggests that significant gains can result from the use of *joint* channel-estimation and data decoding. Strengths of CCS include the facts that i) CCS focuses on reconstruction techniques that have polynomial complexity in L and S_{\max} ; ii) CCS focuses on reconstruction techniques that do not need to know the distributions of the signal and noise; iii) CCS guarantees like (6), which hold for

any sparsity $S \leq S_{\max}$, can be further extended to cover the case of approximately sparse (i.e., “compressible”) signals [4, p. 5].

II. NONCOHERENT CAPACITY

In this section, we characterize the ergodic noncoherent capacity of the sparse frequency-selective block-fading channel described in Section I. We focus on the high-SNR regime, i.e., $\rho \rightarrow \infty$.

Theorem 1: The ergodic noncoherent capacity of the sparse frequency-selective block-fading channel, $C_{\text{sparse}}(\rho)$, in bits per channel use, obeys $\lim_{\rho \rightarrow \infty} \frac{C_{\text{sparse}}(\rho)}{\log \rho} = 1 - \frac{S}{N}$ for sparsity S and block length N , whether or not the channel support realization \mathcal{L} is known a priori.

Proof: Using the chain rule for mutual information [13], it follows straightforwardly that

$$I(\mathbf{y}^{(k)}; \mathbf{x}^{(k)}) = I(\mathbf{y}^{(k)}; \mathcal{L}^{(k)}) + I(\mathbf{y}^{(k)}; \mathbf{x}^{(k)} | \mathcal{L}^{(k)}) - I(\mathbf{y}^{(k)}; \mathcal{L}^{(k)} | \mathbf{x}^{(k)}). \quad (7)$$

where $I(\mathbf{a}; \mathbf{b})$ denotes the mutual information between random vectors \mathbf{a} and \mathbf{b} and where $I(\mathbf{a}; \mathbf{b} | \mathbf{c})$ denotes the conditional mutual information between \mathbf{a} and \mathbf{b} conditioned on \mathbf{c} . Then, since $|\mathcal{L}^{(k)}| = M$, we can bound the first term in (7) as follows:

$$I(\mathbf{y}^{(k)}; \mathcal{L}^{(k)}) \leq h(\mathcal{L}^{(k)}) \leq \log |\mathcal{L}^{(k)}| = \log M, \quad (8)$$

where $h(\mathbf{a})$ denotes the entropy of \mathbf{a} . Because $I(\mathbf{y}^{(k)}; \mathcal{L}^{(k)} | \mathbf{x}^{(k)}) \geq 0$, (7)–(8) yield the upper bound $I(\mathbf{y}^{(k)}; \mathbf{x}^{(k)}) \leq \log M + I(\mathbf{y}^{(k)}; \mathbf{x}^{(k)} | \mathcal{L}^{(k)})$. Similarly, since $I(\mathbf{y}^{(k)}; \mathcal{L}^{(k)}) \geq 0$, (7) implies that $I(\mathbf{y}^{(k)}; \mathbf{x}^{(k)}) \geq I(\mathbf{y}^{(k)}; \mathbf{x}^{(k)} | \mathcal{L}^{(k)}) - I(\mathbf{y}^{(k)}; \mathcal{L}^{(k)} | \mathbf{x}^{(k)})$ and, since $I(\mathbf{y}^{(k)}; \mathcal{L}^{(k)} | \mathbf{x}^{(k)}) \leq h(\mathcal{L}^{(k)} | \mathbf{x}^{(k)}) \leq \log M$, we also have that $I(\mathbf{y}^{(k)}; \mathbf{x}^{(k)}) \geq I(\mathbf{y}^{(k)}; \mathbf{x}^{(k)} | \mathcal{L}^{(k)}) - \log M$. In summary, we have that

$$I(\mathbf{y}^{(k)}; \mathbf{x}^{(k)}) = I(\mathbf{y}^{(k)}; \mathbf{x}^{(k)} | \mathcal{L}^{(k)}) + \Delta \quad \text{for } \Delta \in [-\log M, \log M]. \quad (9)$$

Given knowledge of the support $\mathcal{L}^{(k)}$, the frequency-domain vector $\mathbf{h}_f^{(k)}$ is zero-mean Gaussian with a rank- S covariance matrix. Thus, [8, Theorem 1] implies that $C_{\mathcal{L}}(\rho)$, the pre-log factor of ergodic noncoherent capacity under knowledge of the support \mathcal{L} equals $1 - \frac{S}{N}$, i.e., $\lim_{\rho \rightarrow \infty} \frac{C_{\mathcal{L}}(\rho)}{\log \rho} = 1 - \frac{S}{N}$. Since

$$C_{\mathcal{L}}(\rho) = \max_{p(\mathbf{x}_f^{(k)}): \mathbb{E} \|\mathbf{x}_f^{(k)}\|^2 \leq N} \frac{1}{N} I(\mathbf{y}_f^{(k)}; \mathbf{x}_f^{(k)} | \mathcal{L}^{(k)}), \quad (10)$$

where $I(\mathbf{y}_f^{(k)}; \mathbf{x}_f^{(k)} | \mathcal{L}^{(k)}) = I(\mathbf{y}^{(k)}; \mathbf{x}^{(k)} | \mathcal{L}^{(k)})$ and where, due to (9), $I(\mathbf{y}^{(k)}; \mathbf{x}^{(k)} | \mathcal{L}^{(k)})$ differs from $I(\mathbf{y}^{(k)}; \mathbf{x}^{(k)})$ by a bounded ρ -invariant constant Δ , the ergodic noncoherent capacity

$$C_{\text{sparse}}(\rho) = \max_{p(\mathbf{x}^{(k)}): \mathbb{E} \|\mathbf{x}^{(k)}\|^2 \leq N} \frac{1}{N} I(\mathbf{y}^{(k)}; \mathbf{x}^{(k)}), \quad (11)$$

must also obey $\lim_{\rho \rightarrow \infty} \frac{C_{\text{sparse}}(\rho)}{\log \rho} = 1 - \frac{S}{N}$. ■

It is interesting to notice that the channel multiplexing gain equals $1 - \frac{S}{N}$ whether or not the support \mathcal{L} is a priori known.

III. PILOT AIDED TRANSMISSION AND DECODING

For the *non-sparse* frequency-selective block-fading channel, it has been shown [7] that pilot aided transmission (PAT) is *spectrally efficient* as defined in Section I, i.e., that it is possible to design a PAT scheme for which the pre-log factor in its achievable rate expression coincides with the pre-log factor in the noncoherent ergodic capacity expression (i.e., the channel multiplexing gain). The question remains as to whether PAT is spectrally efficient for *sparse* channels as well.

Interestingly, Theorem 1 showed that the multiplexing gain of the sparse channel does not change with knowledge of the channel support \mathcal{L} . Realizing⁶ that an S -sparse channel with known support has the same capacity characteristics as a non-sparse length- S channel, and recalling that PAT is spectrally efficient for non-sparse channels, one might suspect that PAT is spectrally efficient for sparse channels. As we shall see, this is indeed the case. To prove this, we construct an appropriate PAT scheme and a corresponding decoder, as detailed in Sections III-A–III-C.

A. PAT Definition

For the transmission scheme outlined in Section I-A, we consider a PAT scheme in which the elements in the frequency-domain transmission vector $\mathbf{x}_f^{(k)} \in \mathbb{C}^N$ can be partitioned into a pilot vector $\mathbf{x}_p \in \mathbb{C}^P$, created from $\{x_f^{(k)}[n] : n \in \mathcal{N}_p\}$, and a data vector $\mathbf{x}_d \in \mathbb{C}^{N-P}$, created from $\{x_f^{(k)}[n] : n \in \mathcal{N}_d\}$. Here, we use $\mathcal{N}_p \subset \{0, \dots, N-1\}$ to denote the pilot subcarrier indices and \mathcal{N}_d to denote the corresponding data subcarrier indices, where $\mathcal{N}_d = \{0, \dots, N-1\} \setminus \mathcal{N}_p$. Notice that exactly P signal-space dimensions (per fading block) have been allocated to pilots, i.e., $|\mathcal{N}_p| = P$. For simplicity, we assume that the pilot locations \mathcal{N}_p and pilot values \mathbf{x}_p do not change with the fading block k , and that the pilot values are constant modulus, i.e., $|x_p[n]| = 1$. By definition, the pilot quantities \mathbf{x}_p and \mathcal{N}_p are known a priori to the decoder.

In the parallel subchannel model (3), we partition both $\mathbf{y}_f^{(k)} \in \mathbb{C}^N$ and $\mathbf{v}_f^{(k)} \in \mathbb{C}^N$ in the same way as we did $\mathbf{x}_f^{(k)} \in \mathbb{C}^N$, yielding

$$\mathbf{y}_p^{(k)} = \sqrt{\rho} \mathbf{D}(\mathbf{x}_p) \mathbf{J}_p \mathbf{h}_f^{(k)} + \mathbf{v}_p^{(k)} \quad (12)$$

$$\mathbf{y}_d^{(k)} = \sqrt{\rho} \mathbf{D}(\mathbf{x}_d) \mathbf{J}_d \mathbf{h}_f^{(k)} + \mathbf{v}_d^{(k)}, \quad (13)$$

where \mathbf{J}_p is a selection matrix constructed from rows \mathcal{N}_p of the $N \times N$ identity matrix, and \mathbf{J}_d is constructed similarly from rows \mathcal{N}_d of the identity matrix. Another way to write $\mathbf{y}_p^{(k)}$ and $\mathbf{y}_d^{(k)}$, which will be useful in the sequel, is

$$\mathbf{y}_p^{(k)} = \sqrt{\rho N} \mathbf{D}(\mathbf{x}_p) \mathbf{F}_{p, \text{true}}^{(k)} \mathbf{h}_{\text{nz}}^{(k)} + \mathbf{v}_p^{(k)} \quad (14)$$

$$\mathbf{y}_d^{(k)} = \sqrt{\rho N} \mathbf{D}(\mathbf{x}_d) \mathbf{F}_{d, \text{true}}^{(k)} \mathbf{h}_{\text{nz}}^{(k)} + \mathbf{v}_d^{(k)}, \quad (15)$$

where $\mathbf{h}_{\text{nz}}^{(k)} \in \mathbb{C}^S$ is formed from the non-zero elements of $\mathbf{h}^{(k)}$, $\mathbf{F}_{p, \text{true}}^{(k)}$ is formed from rows \mathcal{N}_p and columns $\mathcal{L}^{(k)}$ of the DFT

⁶The equivalence in pre-log factor between S -sparse channel with known support and a non-sparse length- S channel follows directly from [8, Theorem 1] and the fact that, in both cases, $\mathbf{h}_f^{(k)}$ is zero-mean Gaussian with rank- S covariance matrix.

matrix \mathbf{F} , and $\mathbf{F}_{\text{d,true}}^{(k)}$ is formed from rows \mathcal{N}_{d} and columns $\mathcal{L}^{(k)}$ of \mathbf{F} . Notice that, because $\mathcal{L}^{(k)}$ is not a priori known to the decoder, neither are $\mathbf{F}_{\text{p,true}}^{(k)}$ or $\mathbf{F}_{\text{d,true}}^{(k)}$.

To achieve an arbitrarily small probability of decoding error, we construct codewords that span K blocks, where K is arbitrarily large. Thus, using $\mathcal{C} \subset \mathbb{C}^{K(N-P)}$ to denote our codebook, we partition each codeword $\mathbf{x}_{\text{d}} \in \mathcal{C}$ into K data vectors, i.e., $\mathbf{x}_{\text{d}} = [\mathbf{x}_{\text{d}}^{(1)\top}, \dots, \mathbf{x}_{\text{d}}^{(K)\top}]^\top$, for use in our PAT scheme. The codewords \mathbf{x}_{d} are generated independently from a Gaussian distribution such that the $\mathbf{x}_{\text{d}}^{(k)}$ has positive definite covariance matrix \mathbf{R}_{d} for all k , and such that $\mathbf{x}_{\text{d}}^{(k)}$ is independent of $\mathbf{x}_{\text{d}}^{(k')}$ for $k \neq k'$. Denoting the number of codewords in the codebook by $|\mathcal{C}|$, the average data rate is given by $\mathcal{R} = \frac{1}{KN} \log |\mathcal{C}|$.

B. Optimal Decoding for PAT

The reader may naturally wonder: what is the *optimal* decoder for the above PAT scheme in the case of the sparse channel described in Section I, and how does it compare to optimal decoding in the non-sparse case? To answer these questions, we detail the optimal decoder for the sparse and non-sparse cases below. In the sequel, we use $\mathbf{F}_i \in \mathbb{C}^{N \times S}$ to denote the matrix formed from columns \mathcal{L}_i of the DFT matrix \mathbf{F} , we use $\mathbf{F}_{\text{p},i} \in \mathbb{C}^{P \times S}$ to denote the matrix formed from rows \mathcal{N}_{p} of \mathbf{F}_i , and we use $\mathbf{F}_{\text{d},i} \in \mathbb{C}^{(N-P) \times S}$ to denote the matrix formed from rows \mathcal{N}_{d} of \mathbf{F}_i .

Lemma 1: The maximum likelihood decoder for PAT over the S -sparse L -length frequency-selective N -block-fading channel takes the form of (16) at the bottom of the page, where $\hat{\lambda}_{\text{p},i}^{(k)} \triangleq \Pr\{\mathcal{L}^{(k)} = \mathcal{L}_i | \mathbf{y}_{\text{p}}^{(k)}, \mathbf{x}_{\text{p}}\}$ is the pilot-aided channel-support posterior, where $\hat{\mathbf{h}}_{\text{nz,p},i}^{(k)}$ is the \mathcal{L}_i -conditional pilot-aided MMSE estimate of $\mathbf{h}_{\text{nz}}^{(k)}$ and $\Sigma_{\text{nz,p},i}$ is its error covariance, which take the form

$$\hat{\mathbf{h}}_{\text{nz,p},i}^{(k)} = \sqrt{\frac{\rho}{N}} \mathbf{F}_{\text{p},i}^{\text{H}} (\rho \mathbf{F}_{\text{p},i} \mathbf{F}_{\text{p},i}^{\text{H}} + \frac{S}{N} \mathbf{I})^{-1} \mathcal{D}(\mathbf{x}_{\text{p}}^*) \mathbf{y}_{\text{p}}^{(k)}, \quad (17)$$

$$\Sigma_{\text{nz,p},i} = \frac{1}{S} (\mathbf{I} - \mathbf{F}_{\text{p},i}^{\text{H}} (\mathbf{F}_{\text{p},i} \mathbf{F}_{\text{p},i}^{\text{H}} + \frac{S}{N} \mathbf{I})^{-1} \mathbf{F}_{\text{p},i}), \quad (18)$$

and where $\hat{\mathbf{h}}_{\text{nz},i}^{(k)}(\mathbf{x}_{\text{d}}^{(k)})$ denotes the MMSE estimate of $\mathbf{h}_{\text{nz}}^{(k)}$ conditioned on the data hypothesis $\mathbf{x}_{\text{d}}^{(k)}$ and based on the pilot-aided channel statistics (17)–(18), i.e., see (19) at the bottom of the page.

Proof: See Appendix A. ■

Paraphrasing Lemma 1, the optimal decoder (16) for sparse-channel PAT first uses pilots to compute support posteriors $\{\hat{\lambda}_{\text{p},i}^{(k)}\}_{i=1}^M$ and support-conditional channel posteriors $\{\hat{\mathbf{h}}_{\text{nz,p},i}^{(k)}\}_{i=1}^M$ for each fading block k . Then, it averages over the M support hypotheses to obtain a joint data-channel decoding metric for each fading block k . Finally, it searches for the codeword that maximizes the product of the decoding metrics (over all fading blocks k). We note that optimal decoding is an example of Bayes model averaging [14] and differs markedly from the decoding approach implied in the compressed channel sensing (CCS) framework [4], which aims to compute a *single* sparse channel estimate $\{\hat{\mathbf{h}}_{\text{nz,p},i}^{(k)}, \mathcal{L}^{(k)} = \mathcal{L}_i\}$ for later use in data decoding. We also note that ML decoding complexity is $\mathcal{O}(|\mathcal{C}|MKN^3)$.

For illustrative purposes, we compare the optimal decoder for a sparse channel (as specified in Lemma 1 above) to the optimal decoder for a *non-sparse* channel, as detailed below in Corollary 1.

Corollary 1: The maximum likelihood decoder for PAT over the non-sparse L -length frequency-selective N -block-fading channel takes the form

$$\begin{aligned} \hat{\mathbf{x}}_{\text{d}}^{\text{ML}} = \arg \min_{\mathbf{x}_{\text{d}} \in \mathcal{C}} \sum_{k=1}^K & \left(\ln \det(\rho N \mathbf{F}_{\text{d}}^{\text{H}} \text{diag}(\mathbf{x}_{\text{d}}^{(k)} \odot \mathbf{x}_{\text{d}}^{(k)*}) \mathbf{F}_{\text{d}} + \Sigma_{\text{nz,p}}^{-1}) \right. \\ & + \left\| \mathbf{y}_{\text{d}}^{(k)} - \sqrt{\rho N} \mathcal{D}(\mathbf{x}_{\text{d}}^{(k)}) \mathbf{F}_{\text{d}} \hat{\mathbf{h}}_{\text{nz}}^{(k)}(\mathbf{x}_{\text{d}}^{(k)}) \right\|^2 \\ & \left. + \left\| \hat{\mathbf{h}}_{\text{nz}}^{(k)}(\mathbf{x}_{\text{d}}^{(k)}) - \hat{\mathbf{h}}_{\text{nz,p}}^{(k)} \right\|_{\Sigma_{\text{nz,p}}^{-1}}^2 \right), \end{aligned} \quad (20)$$

where $\hat{\mathbf{h}}_{\text{nz,p}}^{(k)}$ is the pilot-aided MMSE estimate of $\mathbf{h}_{\text{nz}}^{(k)}$ and $\Sigma_{\text{nz,p}}$ is its error covariance, which take the form

$$\hat{\mathbf{h}}_{\text{nz,p}}^{(k)} \triangleq \sqrt{\frac{\rho}{N}} \mathbf{F}_{\text{p}}^{\text{H}} (\rho \mathbf{F}_{\text{p}} \mathbf{F}_{\text{p}}^{\text{H}} + \frac{L}{N} \mathbf{I})^{-1} \mathcal{D}(\mathbf{x}_{\text{p}}^*) \mathbf{y}_{\text{p}}^{(k)}, \quad (21)$$

$$\Sigma_{\text{nz,p}} \triangleq \frac{1}{L} (\mathbf{I} - \mathbf{F}_{\text{p}}^{\text{H}} (\mathbf{F}_{\text{p}} \mathbf{F}_{\text{p}}^{\text{H}} + \frac{L}{N} \mathbf{I})^{-1} \mathbf{F}_{\text{p}}), \quad (22)$$

⁷Note that $\{\Sigma_{\text{nz,p},i}\}_{i=1}^M$ can be precomputed since they do not depend on the observations.

⁸The term after the sum in (16) must be computed for every triple $(i, k, \mathbf{x}_{\text{d}}^{(k)})$, where the complexity of each computation is $\mathcal{O}(N^3)$ due to the matrix inversion in (19).

$$\begin{aligned} \hat{\mathbf{x}}_{\text{d}}^{\text{ML}} = \arg \max_{\mathbf{x}_{\text{d}} \in \mathcal{C}} & \prod_{k=1}^K \sum_{i=1}^M \hat{\lambda}_{\text{p},i}^{(k)} \det \left(\rho N \mathbf{F}_{\text{d},i}^{\text{H}} \mathcal{D}(\mathbf{x}_{\text{d}}^{(k)} \odot \mathbf{x}_{\text{d}}^{(k)*}) \mathbf{F}_{\text{d},i} + \Sigma_{\text{nz,p},i}^{-1} \right)^{-1} \\ & \exp \left(- \left\| \mathbf{y}_{\text{d}}^{(k)} - \sqrt{\rho N} \mathcal{D}(\mathbf{x}_{\text{d}}^{(k)}) \mathbf{F}_{\text{d},i} \hat{\mathbf{h}}_{\text{nz},i}^{(k)}(\mathbf{x}_{\text{d}}^{(k)}) \right\|^2 - \left\| \hat{\mathbf{h}}_{\text{nz},i}^{(k)}(\mathbf{x}_{\text{d}}^{(k)}) - \hat{\mathbf{h}}_{\text{nz,p},i}^{(k)} \right\|_{\Sigma_{\text{nz,p},i}^{-1}}^2 \right) \end{aligned} \quad (16)$$

$$\begin{aligned} \hat{\mathbf{h}}_{\text{nz},i}^{(k)}(\mathbf{x}_{\text{d}}^{(k)}) = & \hat{\mathbf{h}}_{\text{nz,p},i}^{(k)} + \sqrt{\rho N} \Sigma_{\text{nz,p},i} \mathbf{F}_{\text{d},i}^{\text{H}} \mathcal{D}(\mathbf{x}_{\text{d}}^{(k)*}) (\rho N \mathcal{D}(\mathbf{x}_{\text{d}}^{(k)}) \mathbf{F}_{\text{d},i} \Sigma_{\text{nz,p},i} \mathbf{F}_{\text{d},i}^{\text{H}} \mathcal{D}(\mathbf{x}_{\text{d}}^{(k)*}) + \mathbf{I})^{-1} \\ & \times \left(\mathbf{y}_{\text{d}}^{(k)} - \sqrt{\rho N} \mathcal{D}(\mathbf{x}_{\text{d}}^{(k)}) \mathbf{F}_{\text{d},i} \hat{\mathbf{h}}_{\text{nz,p},i}^{(k)} \right). \end{aligned} \quad (19)$$

and where $\hat{\mathbf{h}}_{\text{nz}}^{(k)}(\mathbf{x}_d^{(k)})$ denotes the MMSE estimate of $\mathbf{h}_{\text{nz}}^{(k)}$ conditioned on the data hypothesis $\mathbf{x}_d^{(k)}$ and based on the pilot-aided channel statistics (21)–(22), i.e., see equation (23) at the bottom of the page.

To paraphrase Corollary 1, the optimal decoder (16) for *non-sparse-channel* PAT computes a single pilot-aided MMSE channel estimate $\hat{\mathbf{h}}_{\text{nz},p}^{(k)}$, which is then used to construct a joint data-channel decoding metric, for each fading block k . Finally, it searches for the codeword that minimizes the sum of the decoding metrics (over k). It can be seen that optimal decoding in the sparse case differs from that in the non-sparse cases by the need to compute, at each fading block k , the support posteriors $\{\hat{\lambda}_{p,i}^{(k)}\}_{i=1}^M$ and the corresponding support-conditional tap estimates $\{\hat{\mathbf{h}}_{\text{nz},p,i}^{(k)}\}_{i=1}^M$ and then average the decoding metrics over the M support hypotheses.

C. Decoupled Decoding of PAT

For both sparse and non-sparse channels, the optimal decoder of PAT, as detailed in Section III-B, takes the form of a joint-channel/data decoder. In practice, for reasons of simplicity, decoding is often *decoupled* into two stages: i) pilot-aided channel estimation and ii) coherent data-decoding based on the channel estimate. We now detail a decoupled decoder for the sparse channel of Section I and the PAT scheme of Section III that, while suboptimal, performs well enough to yield spectrally efficient communication *when provided with the correct value of the channel support \mathcal{L}* . In the sequel, we will refer to the case of known \mathcal{L} as the *support-genie* case. Later, in Sections IV-A and IV-B, we will propose schemes to reliably infer the support \mathcal{L} .

For our decoupled decoder, pilot-aided channel estimation is accomplished in a *support-hypothesized* manner. More precisely, we compute—at each fading block k —the pilot-aided MMSE estimate $\hat{\mathbf{h}}_{f,p,i_k}^{(k)}$ of the non-zero taps $\mathbf{h}_f^{(k)}$ under channel-support hypothesis $\mathcal{L}^{(k)} = \mathcal{L}_{i_k}$. To do this, we set $\hat{\mathbf{h}}_{f,p,i_k}^{(k)} = \sqrt{N}\mathbf{F}_{i_k}\hat{\mathbf{h}}_{\text{nz},p,i_k}^{(k)}$ for the $\hat{\mathbf{h}}_{\text{nz},p,i_k}^{(k)}$ specified by (17). Note that $\hat{\mathbf{h}}_{f,p,i_k}^{(k)}$ is a linear estimate due to the fact that $\mathbf{h}_f^{(k)}$ becomes Gaussian when conditioned on a particular support. In contrast, the (support-unconditional) pilot-aided MMSE estimate of $\mathbf{h}_f^{(k)}$ is in general non-linear. The support-hypothesized channel estimates $\{\hat{\mathbf{h}}_{f,p,i_k}^{(k)}\}_{k=1}^K$ and their covariances $\{\Sigma_{f,p,i_k}^{(k)}\}_{k=1}^K$ are then used in coherent data decoding. (Note that $\Sigma_{f,p,i_k}^{(k)} = N\mathbf{F}_{i_k}\Sigma_{\text{nz},p,i_k}\mathbf{F}_{i_k}^H$, where Σ_{nz,p,i_k} is given by (18)). For coherent data decoding, we employ the weighted minimum-distance (WMD) decoder, defined [15] as

$$\hat{\mathbf{x}}_{d,i}^{\text{WMD}} = \arg \min_{\mathbf{x}_d \in \mathcal{C}} \sum_{k=1}^K \left\| \mathbf{Q}_{i_k}^{(k)}(\mathbf{y}_d^{(k)} - \sqrt{\rho}D(\mathbf{x}_d^{(k)})\mathbf{J}_d\hat{\mathbf{h}}_{f,p,i_k}^{(k)}) \right\|^2, \quad (24)$$

where $\mathbf{Q}_{i_k}^{(k)}$ is a weighting matrix and $\mathbf{i} = (i_1, \dots, i_K)$. Writing the observation as

$$\mathbf{y}_d^{(k)} = \sqrt{\rho}D(\mathbf{x}_d^{(k)})\mathbf{J}_d\hat{\mathbf{h}}_{f,p,i_k}^{(k)} + \underbrace{\sqrt{\rho}D(\mathbf{x}_d^{(k)})\mathbf{J}_d\tilde{\mathbf{h}}_{f,p,i_k}^{(k)}}_{\triangleq \mathbf{e}_{d,i_k}^{(k)}} + \mathbf{v}_d^{(k)}, \quad (25)$$

the standard [15] choice for $\mathbf{Q}_{i_k}^{(k)}$ is a whitening matrix for the “effective noise” $\mathbf{e}_{d,i_k}^{(k)}$. We note that the covariance $\mathbf{C}_{d,i_k} \triangleq \text{cov}\{\mathbf{e}_{d,i_k}^{(k)}\}$ (and thus $\mathbf{Q}_{i_k}^{(k)}$) depends on Σ_{f,p,i_k} , \mathbf{R}_d , and ρ .

For the achievable rate of the decoupled-decoder PAT system to grow logarithmically with ρ , the effective noise $\mathbf{e}_{d,i_k}^{(k)}$ must satisfy certain properties. Towards this aim, we establish that, with $P \geq S$ pilot tones, the support hypothesized channel estimation error variance decays at the rate of $\frac{1}{\rho}$ as $\rho \rightarrow \infty$, if and only if the support hypothesis is correct.

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

Proof: We begin by recalling that, under support hypothesis $\mathcal{L}^{(k)} = \mathcal{L}_i$, the frequency-domain channel coefficients $\mathbf{h}_f^{(k)}$ are related to the non-zero channel taps $\mathbf{h}_{\text{nz}}^{(k)}$ via $\mathbf{h}_f^{(k)} = \sqrt{N}\mathbf{F}_i\mathbf{h}_{\text{nz}}^{(k)}$, where \mathbf{F}_i contains columns \mathcal{L}_i of the unitary DFT matrix \mathbf{F} . Thus, $\hat{\mathbf{h}}_{f,p,i}^{(k)}$, the \mathcal{L}_i -conditional pilot-aided MMSE estimate of $\mathbf{h}_f^{(k)}$ is related to $\hat{\mathbf{h}}_{\text{nz},p,i}^{(k)}$, the \mathcal{L}_i -conditional MMSE pilot-aided estimate of $\mathbf{h}_{\text{nz}}^{(k)}$, via $\hat{\mathbf{h}}_{f,p,i}^{(k)} = \sqrt{N}\mathbf{F}_i\hat{\mathbf{h}}_{\text{nz},p,i}^{(k)}$. Because the columns of \mathbf{F}_i are orthonormal, the estimation error obeys

$$\begin{aligned} \|\tilde{\mathbf{h}}_{f,p,i}^{(k)}\|^2 &= \|\mathbf{h}_f^{(k)} - \hat{\mathbf{h}}_{f,p,i}^{(k)}\|^2 = N\|\mathbf{h}_{\text{nz}}^{(k)} - \hat{\mathbf{h}}_{\text{nz},p,i}^{(k)}\|^2 \\ &= N\|\tilde{\mathbf{h}}_{\text{nz},p,i}^{(k)}\|^2. \end{aligned} \quad (26)$$

Plugging (14) into (17), the estimation error $\tilde{\mathbf{h}}_{\text{nz},p,i}^{(k)} \triangleq \mathbf{h}_{\text{nz}}^{(k)} - \hat{\mathbf{h}}_{\text{nz},p,i}^{(k)}$ becomes

$$\begin{aligned} \tilde{\mathbf{h}}_{\text{nz},p,i}^{(k)} &= \left(\mathbf{I} - \mathbf{F}_{p,i}^H(\mathbf{F}_{p,i}\mathbf{F}_{p,i}^H + \frac{S}{\rho N}\mathbf{I})^{-1}\mathbf{F}_{p,\text{true}}^{(k)} \right) \mathbf{h}_{\text{nz}}^{(k)} \\ &\quad - \frac{1}{\sqrt{\rho N}}\mathbf{F}_{p,i}^H(\mathbf{F}_{p,i}\mathbf{F}_{p,i}^H + \frac{S}{\rho N}\mathbf{I})^{-1}D(\mathbf{x}_p^*)\mathbf{v}_p^{(k)}. \end{aligned} \quad (27)$$

$$\begin{aligned} \hat{\mathbf{h}}_{\text{nz}}^{(k)}(\mathbf{x}_d^{(k)}) &= \hat{\mathbf{h}}_{\text{nz},p}^{(k)} + \sqrt{\rho N}\Sigma_{\text{nz},p}\mathbf{F}_d^H D(\mathbf{x}_d^{(k)*})(\rho N D(\mathbf{x}_d^{(k)})\mathbf{F}_d\Sigma_{\text{nz},p}\mathbf{F}_d^H D(\mathbf{x}_d^{(k)*}) + \mathbf{I})^{-1} \\ &\quad \times (\mathbf{y}_d^{(k)} - \sqrt{\rho N}D(\mathbf{x}_d^{(k)})\mathbf{F}_d\hat{\mathbf{h}}_{\text{nz},p}^{(k)}). \end{aligned} \quad (23)$$

Then, since $\mathbf{h}_{\text{nz}}^{(k)}$ is independent of $\mathbf{v}_p^{(k)}$,

$$\begin{aligned} \mathbb{E}\{\|\tilde{\mathbf{h}}_{\text{nz},p,i}^{(k)}\|^2\} &= \frac{1}{S} \text{tr}\left\{\left(\mathbf{I} - \mathbf{F}_{p,i}^H \left(\mathbf{F}_{p,i} \mathbf{F}_{p,i}^H + \frac{S}{\rho N} \mathbf{I}\right)^{-1} \mathbf{F}_{p,\text{true}}^{(k)}\right)\right. \\ &\quad \times \left.\left(\mathbf{I} - \mathbf{F}_{p,i}^H \left(\mathbf{F}_{p,i} \mathbf{F}_{p,i}^H + \frac{S}{\rho N} \mathbf{I}\right)^{-1} \mathbf{F}_{p,\text{true}}^{(k)}\right)^H\right\} \\ &\quad + \frac{1}{\rho N} \text{tr}\left\{\mathbf{F}_{p,i}^H \left(\mathbf{F}_{p,i} \mathbf{F}_{p,i}^H + \frac{S}{\rho N} \mathbf{I}\right)^{-2} \mathbf{F}_{p,i}\right\}. \end{aligned} \quad (28)$$

We now make a few observations about $\mathbf{F}_{p,i}$ and $\mathbf{F}_{p,\text{true}}^{(k)}$. When N is prime, the Chebotarev theorem [16], [17] guarantees that any square submatrix of the N -DFT matrix \mathbf{F} will be full rank. Hence, any tall submatrix of \mathbf{F} will also be full rank. Then, because $P \geq S$, it follows that $\mathbf{F}_{p,i} \in \mathbb{C}^{P \times S}$ will be full rank for all i , as will $\mathbf{F}_{p,\text{true}}^{(k)}$. Furthermore, when $\mathcal{L}_i \neq \mathcal{L}_{\text{true}}^{(k)}$, it follows that $\mathbf{F}_{p,i} \neq \mathbf{F}_{p,\text{true}}^{(k)}$.

To proceed, we use the singular value decomposition $\mathbf{F}_{p,i} = \mathbf{U}_i \boldsymbol{\Sigma}_i \mathbf{V}_i^H$, where $\boldsymbol{\Sigma}_i \in \mathbb{C}^{P \times S}$ is a full-rank diagonal matrix and where \mathbf{U}_i and \mathbf{V}_i are both unitary. Then

$$\mathbf{F}_{p,i}^H \left(\mathbf{F}_{p,i} \mathbf{F}_{p,i}^H + \frac{S}{\rho N} \mathbf{I}\right)^{-1} = \mathbf{V}_i \underbrace{\boldsymbol{\Sigma}_i^H (\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_i^H + \frac{S}{\rho N} \mathbf{I})^{-1}}_{\triangleq \mathbf{D}_i^H} \mathbf{U}_i^H, \quad (29)$$

where $\mathbf{D}_i \in \mathbb{C}^{P \times S}$ is full-rank diagonal with non-zero elements $\left\{\frac{\sigma_{i,l}}{\sigma_{i,l}^2 + S/(\rho N)}\right\}_{l=1}^S$, using $\sigma_{i,l}$ to denote the l^{th} singular value in $\boldsymbol{\Sigma}_i$.

In the case that $\mathcal{L}_i = \mathcal{L}_{\text{true}}^{(k)}$, we have $\mathbf{F}_{p,\text{true}}^{(k)} = \mathbf{F}_{p,i}$, and so

$$\begin{aligned} \mathbb{E}\{\|\tilde{\mathbf{h}}_{\text{nz},p,i}^{(k)}\|^2\} &= \frac{1}{S} \text{tr}\left\{(\mathbf{I} - \mathbf{V}_i \mathbf{D}_i^H \boldsymbol{\Sigma}_i \mathbf{V}_i^H)(\mathbf{I} - \mathbf{V}_i \boldsymbol{\Sigma}_i^H \mathbf{D}_i \mathbf{V}_i^H)\right\} \\ &\quad + \frac{1}{\rho N} \text{tr}\left\{\mathbf{V}_i \mathbf{D}_i^H \mathbf{D}_i \mathbf{V}_i^H\right\} \end{aligned} \quad (30)$$

$$\begin{aligned} &= \frac{1}{S} \text{tr}\left\{(\mathbf{I} - \mathbf{D}_i^H \boldsymbol{\Sigma}_i)(\mathbf{I} - \boldsymbol{\Sigma}_i^H \mathbf{D}_i)\right\} \\ &\quad + \frac{1}{\rho N} \text{tr}\left\{\mathbf{D}_i^H \mathbf{D}_i\right\} \end{aligned} \quad (31)$$

$$\begin{aligned} &= \frac{1}{S} \sum_{l=1}^S \left(1 - \frac{\sigma_{i,l}^2}{\sigma_{i,l}^2 + S/(\rho N)}\right)^2 \\ &\quad + \frac{1}{\rho N} \sum_{l=1}^S \frac{\sigma_{i,l}^2}{(\sigma_{i,l}^2 + S/(\rho N))^2} \end{aligned} \quad (32)$$

$$= \sum_{l=1}^S \frac{1}{N \sigma_{i,l}^2 \rho + S} \quad (33)$$

$$\leq \rho^{-1} \sum_{l=1}^S \frac{1}{N \sigma_{i,l}^2}. \quad (34)$$

Thus, we have the upper bound $\mathbb{E}\{\|\tilde{\mathbf{h}}_{\text{f},p,i}^{(k)}\|^2\} = N \mathbb{E}\{\|\tilde{\mathbf{h}}_{\text{nz},p,i}^{(k)}\|^2\} \leq C \rho^{-1}$ with $C = \sum_{l=1}^S \sigma_{i,l}^{-2}$.

For the case $\mathcal{L}_i \neq \mathcal{L}_{\text{true}}^{(k)}$, we have $\mathbf{F}_{p,\text{true}}^{(k)} \neq \mathbf{F}_{p,i}$, and so we can use the previously defined SVD quantities to write $\mathbf{F}_{p,\text{true}}^{(k)} =$

$\mathbf{U}_i (\boldsymbol{\Sigma}_i + \boldsymbol{\Delta}_i) \mathbf{V}_i^H$, where $\boldsymbol{\Delta}_i \in \mathbb{C}^{P \times S}$ is some non-zero matrix. It then follows that

$$\begin{aligned} \mathbb{E}\{\|\tilde{\mathbf{h}}_{\text{nz},p,i}^{(k)}\|^2\} &= \frac{1}{S} \text{tr}\left\{(\mathbf{I} - \mathbf{V}_i \mathbf{D}_i^H (\boldsymbol{\Sigma}_i + \boldsymbol{\Delta}_i) \mathbf{V}_i^H)(\mathbf{I} - \mathbf{V}_i (\boldsymbol{\Sigma}_i + \boldsymbol{\Delta}_i)^H \mathbf{D}_i \mathbf{V}_i^H)\right\} \\ &\quad + \frac{1}{\rho N} \text{tr}\left\{\mathbf{V}_i \mathbf{D}_i^H \mathbf{D}_i \mathbf{V}_i^H\right\} \end{aligned} \quad (35)$$

$$\begin{aligned} &= \frac{1}{S} \text{tr}\left\{(\mathbf{I} - \mathbf{D}_i^H \boldsymbol{\Sigma}_i - \mathbf{D}_i^H \boldsymbol{\Delta}_i)(\mathbf{I} - \boldsymbol{\Sigma}_i^H \mathbf{D}_i - \boldsymbol{\Delta}_i^H \mathbf{D}_i)\right\} \\ &\quad + \frac{1}{\rho N} \text{tr}\left\{\mathbf{D}_i^H \mathbf{D}_i\right\} \end{aligned} \quad (36)$$

$$\begin{aligned} &= \mathbb{E}\{\|\tilde{\mathbf{h}}_{\text{nz},p,\text{true}}^{(k)}\|^2\} \\ &\quad - \frac{1}{S} \text{tr}\left\{(\mathbf{I} - \mathbf{D}_i^H \boldsymbol{\Sigma}_i) \boldsymbol{\Delta}_i^H \mathbf{D}_i + \mathbf{D}_i^H \boldsymbol{\Delta}_i (\mathbf{I} - \boldsymbol{\Sigma}_i^H \mathbf{D}_i)\right\} \\ &\quad + \frac{1}{S} \text{tr}\left\{\mathbf{D}_i^H \boldsymbol{\Delta}_i \boldsymbol{\Delta}_i^H \mathbf{D}_i\right\} \end{aligned} \quad (37)$$

As established above, $\mathbb{E}\{\|\tilde{\mathbf{h}}_{\text{nz},p,\text{true}}^{(k)}\|^2\} \rightarrow 0$ as $\rho \rightarrow \infty$. Since $\mathbf{I} - \mathbf{D}_i^H \boldsymbol{\Sigma}_i$ is diagonal with elements $\left\{\frac{1}{1 + \rho N \sigma_{i,l}^2}\right\}_{l=1}^S$, the second term in (37) also vanishes as $\rho \rightarrow \infty$. The third term in (37), however, converges to the quantity $\frac{1}{S} \text{tr}\{\boldsymbol{\Sigma}_i^+ \boldsymbol{\Delta}_i \boldsymbol{\Delta}_i^H \boldsymbol{\Sigma}_i^{+H}\}$ as $\rho \rightarrow \infty$, where $(\cdot)^+$ denotes pseudo-inverse. Now, since $\mathbf{F}_{p,\text{true}}^{(k)}$ and $\mathbf{F}_{p,i}$ are distinct full rank matrices with $\text{tr}\{\mathbf{F}_{p,\text{true}}^{(k)H} \mathbf{F}_{p,i}^{(k)}\} = \text{tr}\{\mathbf{F}_{p,i}^H \mathbf{F}_{p,i}\}$, it follows that $\boldsymbol{\Sigma}_i^+ \boldsymbol{\Delta}_i \neq \mathbf{0}$ and hence $\text{tr}\{\boldsymbol{\Sigma}_i^+ \boldsymbol{\Delta}_i \boldsymbol{\Delta}_i^H \boldsymbol{\Sigma}_i^{+H}\} > 0$. So there does not exist C such that $\mathbb{E}\{\|\tilde{\mathbf{h}}_{\text{nz},p,i}^{(k)}\|^2\} \leq C \rho^{-1}$ for all $\rho > 0$. ■

Corollary 2: Lemma 2, and several other results in the paper, are stated under prime N , arbitrary \mathcal{N}_p , and $L < N$. The requirement that N is prime can be relaxed in exchange for the following restrictions on \mathcal{N}_p and L .

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

Proof: Throughout the paper, the prime- N property is used only to guarantee that certain square submatrices of the N -DFT matrix \mathbf{F} remain full rank. When forming these submatrices, we use S row indices from \mathcal{N}_p (where $\mathcal{N}_p \subset \{0, \dots, N-1\}$ and $|\mathcal{N}_p| = P \geq S$) and S column indices from \mathcal{L}_i (where $\mathcal{L}_i \subset \{0, \dots, L-1\}$ and $|\mathcal{L}_i| = S$). In the case that N is prime, the Chebotarev theorem [16], [17] guarantees that our square submatrix will be full rank, as discussed in the proof of Lemma 2. However, even when N is not prime, our square submatrix will be full rank whenever both \mathcal{N}_p and \mathcal{L}_i do not form groups with respect to modulo- N addition, nor cosets of subgroups of $\{0, 1, \dots, N-1\}$ w.r.t modulo- N addition [10, p.491]. These conditions on \mathcal{N}_p and \mathcal{L}_i are ensured by the two conditions stated in the corollary. ■

For a given communication scheme, we say that a rate \mathcal{R} (in bits per channel use) is *achievable* if the probability of decoding error can be made arbitrarily small at that rate. Now, using the bound on the estimation error variance from Lemma 2, we establish that when the true channel support is apriori known at

receiver (i.e., the support-genie case), the achievable rates satisfy $\lim_{\rho \rightarrow \infty} \frac{\mathcal{R}(\rho)}{\log \rho} = 1 - \frac{P}{N}$, where $P \geq S$ denotes the number of pilot tones.

Lemma 3: Say that N is prime, and that the true channel support is known apriori at the receiver for each fading block. Then, for any pilot pattern \mathcal{N}_p such that $P \geq S$, the achievable rate of the support-hypothesized estimator-decoder satisfies $\lim_{\rho \rightarrow \infty} \frac{\mathcal{R}(\rho)}{\log \rho} = 1 - \frac{P}{N}$.

Proof: The achievable rate of WMD decoding under imperfect channel state information (CSI) and Gaussian coding was studied in [15], where the rate expressions were obtained under certain restrictions on the statistical properties of the imperfect CSI. In the support-genie case, our support-hypothesized channel estimator satisfies all of the standard requirements in [15] except for time-invariance, since the support varies over the fading blocks. However, our model does satisfy the alternative ergodic condition in [15]. To see this, we need to verify that, for any function $f(\cdot)$, we have $\lim_{K \rightarrow \infty} \frac{1}{K} \sum_{k=1}^K f(\mathbf{y}_d^{(k)}, \hat{\mathbf{h}}_{f,p,i_k,\text{true}}^{(k)}) = \mathbb{E}\{f(\mathbf{y}_d^{(k)}, \hat{\mathbf{h}}_{f,p,i_k,\text{true}}^{(k)})\}$, using $i_{k,\text{true}}$ to denote the index of the true support during the k th fading block, and $\hat{\mathbf{h}}_{f,p,i_k,\text{true}}^{(k)} \triangleq \sqrt{N} \mathbf{F}_{i_k,\text{true}} \hat{\mathbf{h}}_{\text{nz},p,\text{true}}^{(k)}$. Let us define $\mathcal{K}_i = \{k : \mathcal{L}_{\text{true}}^{(k)} = \mathcal{L}_i\}$ for $i = 1, \dots, M$. Then it follows that,

$$\begin{aligned} & \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{k=1}^K f(\mathbf{y}_d^{(k)}, \hat{\mathbf{h}}_{f,p,i_k,\text{true}}^{(k)}) \\ &= \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{i=1}^M \sum_{k \in \mathcal{K}_i} f(\mathbf{y}_d^{(k)}, \hat{\mathbf{h}}_{f,p,i}^{(k)}), \end{aligned} \quad (38)$$

$$= \sum_{i=1}^M \lim_{K \rightarrow \infty} \frac{|\mathcal{K}_i|}{K} \frac{1}{|\mathcal{K}_i|} \sum_{k \in \mathcal{K}_i} f(\mathbf{y}_d^{(k)}, \hat{\mathbf{h}}_{f,p,i}^{(k)}), \quad (39)$$

$$= \sum_{i=1}^M \lambda_i \mathbb{E}\{f(\mathbf{y}_d^{(k)}, \hat{\mathbf{h}}_{f,p,i}^{(k)}) \mid \mathcal{L}_{\text{true}}^{(k)} = \mathcal{L}_i\}, \quad (40)$$

$$= \mathbb{E}\{f(\mathbf{y}_d^{(k)}, \hat{\mathbf{h}}_{f,p,i_k,\text{true}}^{(k)})\}. \quad (41)$$

Hence [15, Theorem 2] can be applied to find the achievable rates for our decoupled decoding scheme under the support genie. In particular, by rewriting the data observations from (25) as

$$\mathbf{y}_d^{(k)} = \sqrt{\rho} \mathbf{D}(\mathbf{x}_d^{(k)}) \mathbf{J}_d \hat{\mathbf{h}}_{f,p,i_k,\text{true}}^{(k)} + \mathbf{e}_{d,i_k,\text{true}}^{(k)}, \quad (42)$$

for effective noise $\mathbf{e}_{d,i_k,\text{true}}^{(k)} \triangleq \sqrt{\rho} \mathbf{D}(\mathbf{x}_d^{(k)}) \mathbf{J}_d \tilde{\mathbf{h}}_{f,p,i_k,\text{true}}^{(k)} + \mathbf{v}_d^{(k)}$, it follows [15] that the achievable rate (in bits per channel use) is shown in (43) at the bottom of the page, where $\mathbf{C}_{d,i}(\rho) \triangleq \text{cov}\{\mathbf{e}_{d,i}^{(k)}\}$ for $\mathbf{e}_{d,i}^{(k)}$ defined in (25). Similar to (40)–(41), we can rewrite (43) as (44) at the bottom of the page. When $\mathcal{L}_{\text{true}}^{(k)} = \mathcal{L}_i$, Lemma 2 specifies that there exists some constant C such that $\mathbb{E}\{\|\tilde{\mathbf{h}}_{\text{nz},p,i}^{(k)}\|^2\} \leq C\rho^{-1}$ for all ρ . In this case, the eigenvalues of $\mathbf{C}_{d,i}(\rho)$ will be positive and bounded from above for all ρ , and thus eigenvalues of $\mathbf{C}_{d,i}^{-1}(\rho)$ will be positive and bounded from below for all ρ . Thus, using a standard high-SNR analysis (see, e.g., [18] for details), $\lim_{\rho \rightarrow \infty} \frac{\mathcal{R}_i(\rho)}{\log \rho} = 1 - \frac{P}{N}$ for any i , from which the stated result of this lemma follows. ■

In [7], it has been shown that, for L -length non-sparse channels, PAT can be designed to achieve data rates that satisfy $\lim_{\rho \rightarrow \infty} \frac{\mathcal{R}(\rho)}{\log \rho} = 1 - \frac{P}{N}$, for $P \geq L$. Our Lemma 3 can be interpreted as an extension of the result from [7] to L -length S -sparse channels with known support.

IV. CHANNEL-SUPPORT DECODING

In summary, the PAT scheme of Section III-A and the decoupled decoder of Section III-C 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 (i.e., i such that $\mathcal{L}_i = \mathcal{L}_{\text{true}}$). In this section, we consider schemes for reliably decoding the channel support of each block.

A. Data-Aided Support Decoding

In this section, we show that, with prime N , the pilot aided transmission (PAT) scheme defined in Section III-A is spectrally efficient for the sparse frequency-selective block-fading channel. In other words, when the L -length channel is S -sparse, it is sufficient to sacrifice only S signal-space dimensions to maintain an achievable rate that grows at the same rate as channel capacity in the high-SNR regime. To show this, we construct a so-called data-aided support decoder (DASD) that leverages certain error-detecting capabilities in the codebook \mathcal{C} . We first describe the error detection mechanism and later propose a procedure for channel support decoding.

In our DASD scheme, we attach error detection parity bits, which we refer to as cyclic redundancy check (CRC) bits, to the information bits *prior* to the channel-coding operation. Attaching parity bits to the information bits is a commonly used

$$\mathcal{R}(\rho) = \frac{1}{N} \mathbb{E}\left\{ \log \det \left[\mathbf{I} + \rho \mathbf{C}_{d,i_k,\text{true}}^{-1}(\rho) \mathbf{D}(\mathbf{J}_d \hat{\mathbf{h}}_{f,p,i_k,\text{true}}^{(k)}) \mathbf{R}_d \mathbf{D}(\mathbf{J}_d \hat{\mathbf{h}}_{f,p,i_k,\text{true}}^{(k)})^H \right] \right\}, \quad (43)$$

$$\mathcal{R}(\rho) = \frac{1}{N} \sum_{i=1}^M \lambda_i \underbrace{\mathbb{E}\left\{ \log \det \left[\mathbf{I} + \rho \mathbf{C}_{d,i}^{-1}(\rho) \mathbf{D}(\mathbf{J}_d \hat{\mathbf{h}}_{f,p,i}^{(k)}) \mathbf{R}_d \mathbf{D}(\mathbf{J}_d \hat{\mathbf{h}}_{f,p,i}^{(k)})^H \right] \mid \mathcal{L}_{\text{true}}^{(k)} = \mathcal{L}_i \right\}}_{\triangleq \mathcal{R}_i(\rho)}. \quad (44)$$

mechanism to identify the decoding errors at the receiver [19]. Let us denote the information bit rate as R , and the CRC bit rate as δ , both in units of bits-per-channel-use. Then, over $m = KN$ channel uses, we use a total⁹ of mR bits for information and a total of $m\delta$ bits for CRC. Let $\mu(\cdot)$ denote the function which specifies the $m\delta$ parity bits for every set of mR information bits. Specifically, $\mu : \{1, \dots, 2^{mR}\} \rightarrow \{1, \dots, 2^{m\delta}\}$ is a ‘‘binning function’’ mapping information bits to corresponding CRC bits, so that, for the information message \mathbf{w} , the corresponding CRC bits are $\mathbf{u} = \mu(\mathbf{w})$. Such \mathbf{u} is sometimes referred to as the ‘‘auxiliary check message.’’ The channel-encoder then maps the ‘‘composite message’’ (\mathbf{w}, \mathbf{u}) , containing $m(R + \delta)$ bits, to one of the $2^{m(R+\delta)}$ codewords in the codebook \mathcal{C} . (See Section III-A for details on the codebook.) For clarity, we use ‘‘message’’ when referring to channel-coder inputs, and ‘‘codeword’’ when referring to channel-coder outputs.

The DASD support decoding procedure is defined as follows.

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

- 1) Compute conditional channel estimates $\{\hat{\mathbf{h}}_{f,p,i_k}^{(k)}\}_{k=1}^K$ and $\{\Sigma_{f,p,i_k}\}_{k=1}^K$ using (17)–(18) with $\hat{\mathbf{h}}_{f,p,i_k}^{(k)} = \sqrt{N} \mathbf{F}_{i_k} \hat{\mathbf{h}}_{nz,p,i_k}^{(k)}$ and $\Sigma_{f,p,i_k} = N \mathbf{F}_{i_k} \Sigma_{nz,p,i_k} \mathbf{F}_{i_k}^H$.
- 2) Compute the WMD codeword estimate $\hat{\mathbf{x}}_{d,\mathbf{i}}$ according to (24).
- 3) From the codeword $\hat{\mathbf{x}}_{d,\mathbf{i}}$, recover the corresponding composite message $(\hat{\mathbf{w}}_{\mathbf{i}}, \hat{\mathbf{u}}_{\mathbf{i}})$.
- 4) Perform error detection on $(\hat{\mathbf{w}}_{\mathbf{i}}, \hat{\mathbf{u}}_{\mathbf{i}})$, i.e., check if $\mu(\hat{\mathbf{w}}_{\mathbf{i}}) \neq \hat{\mathbf{u}}_{\mathbf{i}}$.
- 5) If no error is detected or there are no more hypotheses to consider, stop and declare the decoded message as $\hat{\mathbf{w}}_{\mathbf{i}}$, else continue with the next hypothesis \mathbf{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 previously defined PAT scheme, when used with S pilots and DASD, yields an achievable rate $\mathcal{R}^{\text{DASD}}(\rho)$ that obeys $\lim_{\rho \rightarrow \infty} \frac{\mathcal{R}^{\text{DASD}}(\rho)}{\log \rho} = 1 - \frac{S}{N}$. Hence, PAT is spectrally efficient for this channel.

Proof: In our proof, instead of considering a specific binning function $\mu(\cdot)$, we consider the error performance averaged over all possible random binning assignments and establish that

⁹For ease of presentation, we have ignored the flooring $\lfloor mR \rfloor$ and $\lfloor m\delta \rfloor$ and the flooring error can be made negligible by choosing a large m .

the average error approaches zero. For a given support hypothesis $\mathcal{L}_{\mathbf{i}}$, the DASD computes the support-conditional channel estimate and the corresponding WMD codeword estimate from which the composite message bits are obtained, which we write as $(\hat{\mathbf{w}}_{\mathbf{i}}, \hat{\mathbf{u}}_{\mathbf{i}})$. There are two situations under which the DASD terminates, producing the final estimate $\hat{\mathbf{w}}^{\text{DASD}} = \hat{\mathbf{w}}_{\mathbf{i}}$: i) when $\mathbf{i} \neq \mathbf{i}_{\text{last}}$ and $\mu(\hat{\mathbf{w}}_{\mathbf{i}}) = \hat{\mathbf{u}}_{\mathbf{i}}$, or ii) when $\mathbf{i} = \mathbf{i}_{\text{last}}$. Here we use \mathbf{i}_{last} to denote the last of the M^K hypotheses. Note that, in all other cases, an error is detected, and the DASD continues under a different hypothesis $\mathcal{L}_{\mathbf{i}'}$.

We now upper bound the probability that the DASD infers the wrong information bits, i.e., that $\hat{\mathbf{w}}^{\text{DASD}} \neq \mathbf{w}$. Say that \mathbf{i}_{stop} denotes the value of \mathbf{i} used to produce $\hat{\mathbf{w}}^{\text{DASD}}$, i.e., $\hat{\mathbf{w}}^{\text{DASD}} = \hat{\mathbf{w}}_{\mathbf{i}_{\text{stop}}}$. Notice that either 1) $\mathbf{i}_{\text{stop}} = \mathbf{i}_{\text{true}}$ or 2) $\mathbf{i}_{\text{stop}} \neq \mathbf{i}_{\text{true}}$. In the latter case, the support detector fails to detect the true support when either 2a) $\mathbf{i}_{\text{stop}} \neq \mathbf{i}_{\text{last}}$ and $\mu(\hat{\mathbf{w}}_{\mathbf{i}_{\text{stop}}}) = \hat{\mathbf{u}}_{\mathbf{i}_{\text{stop}}}$, where the error was missed, or 2b) $\mathbf{i}_{\text{stop}} = \mathbf{i}_{\text{last}}$. Finally, notice that, if event 2b occurs, the DASD must have (falsely) detected an error under the true support hypothesis, i.e., $\mu(\hat{\mathbf{w}}_{\mathbf{i}_{\text{true}}}) \neq \hat{\mathbf{u}}_{\mathbf{i}_{\text{true}}}$. Thus we can partition the error event $\hat{\mathbf{w}}_{\mathbf{i}_{\text{stop}}} \neq \mathbf{w}$ into three mutually exclusive events:

- E1) $\mathbf{i}_{\text{stop}} = \mathbf{i}_{\text{true}}$ and $\hat{\mathbf{w}}_{\mathbf{i}_{\text{stop}}} \neq \mathbf{w}$,
- E2) $\mathbf{i}_{\text{stop}} = \mathbf{i}_{\text{last}} \neq \mathbf{i}_{\text{true}}$ and both $\mu(\hat{\mathbf{w}}_{\mathbf{i}_{\text{true}}}) \neq \hat{\mathbf{u}}_{\mathbf{i}_{\text{true}}}$ and $\hat{\mathbf{w}}_{\mathbf{i}_{\text{stop}}} \neq \mathbf{w}$.
- E3) $\exists \mathbf{i}_{\text{stop}} \notin \{\mathbf{i}_{\text{true}}, \mathbf{i}_{\text{last}}\}$ s.t. both $\mu(\hat{\mathbf{w}}_{\mathbf{i}_{\text{stop}}}) = \hat{\mathbf{u}}_{\mathbf{i}_{\text{stop}}}$ and $\hat{\mathbf{w}}_{\mathbf{i}_{\text{stop}}} \neq \mathbf{w}$.

We now analyze each of these three events.

Notice that E1 is the event of a data-decoding error under the correct support hypothesis (i.e., $\hat{\mathbf{w}}_{\mathbf{i}_{\text{true}}} \neq \mathbf{w}$). We recall that the correct-support-hypothesis case was analyzed in Section III-C, under which PAT with decoupled decoding was found to be spectrally efficient, having an achievable rate \mathcal{R} that obeys $\lim_{\rho \rightarrow \infty} \frac{\mathcal{R}(\rho)}{\log \rho} = 1 - \frac{S}{N}$. Thus, the probability of E1 can be made arbitrarily small for any rates R and δ such that $R + \delta \leq \mathcal{R}$.

E2 characterizes the event in which the true support is falsely discarded and data-decoding error results later (under an incorrect support hypothesis). Recall that, when the support hypothesis is incorrect, we cannot guarantee a low probability of data-decoding error when communicating at rates that scale as $(1 - \frac{S}{N}) \log \rho$. The key, then, is to make the support-error probability small. Towards this aim, we bound E2 as shown in (45)–(49) at the bottom of the page. Thus, the probability of E2 can be upper bounded by the probability of decoding error

$$\Pr\{\text{E2}\} = \Pr\{\mu(\hat{\mathbf{w}}_{\mathbf{i}_{\text{true}}}) \neq \hat{\mathbf{u}}_{\mathbf{i}_{\text{true}}} \text{ and } \hat{\mathbf{w}}_{\mathbf{i}_{\text{stop}}} \neq \mathbf{w}\} \quad (45)$$

$$\leq \Pr\{\mu(\hat{\mathbf{w}}_{\mathbf{i}_{\text{true}}}) \neq \hat{\mathbf{u}}_{\mathbf{i}_{\text{true}}}\} \quad (46)$$

$$= \Pr\{\mu(\hat{\mathbf{w}}_{\mathbf{i}_{\text{true}}}) \neq \hat{\mathbf{u}}_{\mathbf{i}_{\text{true}}} \mid \hat{\mathbf{w}}_{\mathbf{i}_{\text{true}}} = \mathbf{w}\} \Pr\{\hat{\mathbf{w}}_{\mathbf{i}_{\text{true}}} = \mathbf{w}\} \\ + \Pr\{\mu(\hat{\mathbf{w}}_{\mathbf{i}_{\text{true}}}) \neq \hat{\mathbf{u}}_{\mathbf{i}_{\text{true}}} \mid \hat{\mathbf{w}}_{\mathbf{i}_{\text{true}}} \neq \mathbf{w}\} \Pr\{\hat{\mathbf{w}}_{\mathbf{i}_{\text{true}}} \neq \mathbf{w}\} \quad (47)$$

$$\leq \Pr\{\mu(\hat{\mathbf{w}}_{\mathbf{i}_{\text{true}}}) \neq \hat{\mathbf{u}}_{\mathbf{i}_{\text{true}}} \mid \hat{\mathbf{w}}_{\mathbf{i}_{\text{true}}} = \mathbf{w}\} + \Pr\{\hat{\mathbf{w}}_{\mathbf{i}_{\text{true}}} \neq \mathbf{w}\} \quad (48)$$

$$= \Pr\{\mathbf{u} \neq \hat{\mathbf{u}}_{\mathbf{i}_{\text{true}}}\} + \Pr\{\hat{\mathbf{w}}_{\mathbf{i}_{\text{true}}} \neq \mathbf{w}\}. \quad (49)$$

under the correct support-hypothesis, which (like $\Pr\{E1\}$) 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. Like with E2, the probability of data-decoding cannot be made arbitrarily small under an incorrect support hypothesis, and so we hope that the false alarm error is small. Towards this aim, we begin by upper bounding the probability of the event E3 as shown in (50)–(53) at the bottom of the page, where we used the union bound in (53). Now, to find the probability of missing a support-error, we assume that, when $\hat{\mathbf{w}}_{\mathbf{i}} \neq \mathbf{w}$, the auxiliary check estimate $\mu(\hat{\mathbf{w}}_{\mathbf{i}})$ is uniformly distributed over all possibilities of \mathbf{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 [20]. In this case, for any $\mathbf{i} \neq \mathbf{i}_{\text{true}}$, the probability of missing the detection of a support-error becomes

$$\Pr\{\mu(\hat{\mathbf{w}}_{\mathbf{i}}) = \hat{\mathbf{u}}_{\mathbf{i}} \mid \mathbf{i} \neq \mathbf{i}_{\text{true}}, \hat{\mathbf{w}}_{\mathbf{i}} \neq \mathbf{w}\} = \frac{1}{2^{m\delta}}, \quad (54)$$

so that

$$\Pr\{E3\} \leq \frac{M^K}{2^{m\delta}} = \frac{M^K}{2^{KN\delta}} = \left(\frac{M}{2^{N\delta}}\right)^K. \quad (55)$$

So, when $\delta > \frac{\log M}{N}$, by choosing K large enough, we can make $\Pr\{E3\}$ averaged over all the random binning CRC assignments arbitrarily small. This implies that there exists a binning function $\tilde{\mu}$ for which $\Pr\{E3\}$ can be made arbitrarily small.

Notice that the rate δ sacrificed to make $\Pr\{E3\}$ 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 3, we can construct a codebook that guarantees arbitrarily small values for $\Pr\{E1\} + \Pr\{E2\}$. This codebook, when used in conjunction with the binning function $\tilde{\mu}$, ensures that $\Pr\{\hat{\mathbf{w}}^{\text{DASD}} \neq \mathbf{w}\} = \Pr\{E1\} + \Pr\{E2\} + \Pr\{E3\}$ can be made arbitrarily small. Since δ is fixed with respect to SNR ρ , the information rate of DASD satisfies $\lim_{\rho \rightarrow \infty} \frac{R(\rho)}{\log \rho} = 1 - \frac{S}{N}$. ■

As we have seen, the DASD achieves the optimal pre-log factor, albeit at complexity¹⁰ $\mathcal{O}(|\mathcal{C}|M^K + |\mathcal{C}|MKN^2)$, which may be larger than that of the optimal decoder specified in Lemma 1. In fact, we do not propose DASD for practical use, but rather as a constructive means of proving the achievability of the optimal pre-log factor, since the optimal decoder is difficult to analyze directly. In Section V, we present a simpler suboptimal decoding scheme that also has performance guarantees.

B. Pilot-Aided Support Decoding

In this section, we propose a *pilot-aided support decoder* (PASD) with complexity¹¹ $\mathcal{O}(|\mathcal{C}|KN^2 + KMP^2)$, which is significantly less complex than both DASD and the optimal decoder in Lemma 1. Since only pilots are used to infer the channel support, the complexity of support estimation grows linearly in K . 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 $\rho \rightarrow \infty$ but is not guaranteed to be arbitrarily small at any finite ρ) unless the channel support $\mathcal{L}^{(k)}$ is fixed over fading blocks $k \in \{1, \dots, K\}$.

1) *Pilot-Aided Support Estimation*: We now present an asymptotically reliable method to infer the channel support \mathcal{L} that requires only $P = S + 1$ pilots per fading block. For this, we use the following normalized pilot observations:

$$\mathbf{z}_p^{(k)} \triangleq \frac{1}{\sqrt{\rho N}} \mathcal{D}(\mathbf{x}_p^*) \mathbf{y}_p^{(k)} = \mathbf{F}_{p,\text{true}}^{(k)} \mathbf{h}_{\text{nz}}^{(k)} + \frac{1}{\sqrt{\rho N}} \boldsymbol{\nu}_p^{(k)}, \quad (56)$$

where $\boldsymbol{\nu}_p^{(k)} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ due to the constant-modulus assumption on the pilots. Recalling that $\mathbf{F}_{p,\text{true}}^{(k)}$ is constructed from rows \mathcal{N}_p and columns $\mathcal{L}_{\text{true}}^{(k)}$ of \mathbf{F} , and that $\mathbf{F}_{p,i}$ is constructed from rows \mathcal{N}_p and columns \mathcal{L}_i of \mathbf{F} , we henceforth use $\mathbf{\Pi}_{p,i} \triangleq \mathbf{F}_{p,i} (\mathbf{F}_{p,i}^H \mathbf{F}_{p,i})^{-1} \mathbf{F}_{p,i}^H$ to denote the matrix that projects onto the

¹⁰Note that the term to the right of the sum in the WMD decoder metric (24) must be computed for every triple $(i, k, \mathbf{x}_d^{(k)})$, where the complexity of each computation is $\mathcal{O}(N^2)$. Subsequently, these terms must be summed for each of M^K support-vector hypotheses.

¹¹As described below, for support estimation, K instances of $\hat{\mathbf{i}}_p^{(k)}$ must be computed, each with complexity $\mathcal{O}(MP^2)$. Then, for (support-conditional) WMD decoding, $|\mathcal{C}|K$ instances of the term after the sum in (24) must be computed, each with complexity $\mathcal{O}(N^2)$.

$$\Pr\{E3\} = \Pr\{\exists \mathbf{i}_{\text{stop}} \notin \{\mathbf{i}_{\text{true}}, \mathbf{i}_{\text{last}}\} \text{ s.t. } \mu(\hat{\mathbf{w}}_{\mathbf{i}_{\text{stop}}}) = \hat{\mathbf{u}}_{\mathbf{i}_{\text{stop}}} \mid \hat{\mathbf{w}}_{\mathbf{i}_{\text{stop}}} \neq \mathbf{w}\} \Pr\{\hat{\mathbf{w}}_{\mathbf{i}_{\text{stop}}} \neq \mathbf{w}\} \quad (50)$$

$$\leq \Pr\{\exists \mathbf{i} \notin \{\mathbf{i}_{\text{true}}, \mathbf{i}_{\text{last}}\} \text{ s.t. } \mu(\hat{\mathbf{w}}_{\mathbf{i}}) = \hat{\mathbf{u}}_{\mathbf{i}} \mid \hat{\mathbf{w}}_{\mathbf{i}} \neq \mathbf{w}\} \quad (51)$$

$$\leq \Pr\{\exists \mathbf{i} \neq \mathbf{i}_{\text{true}} \text{ s.t. } \mu(\hat{\mathbf{w}}_{\mathbf{i}}) = \hat{\mathbf{u}}_{\mathbf{i}} \mid \hat{\mathbf{w}}_{\mathbf{i}} \neq \mathbf{w}\} \quad (52)$$

$$\leq \sum_{\mathbf{i} \neq \mathbf{i}_{\text{true}}} \Pr\{\mu(\hat{\mathbf{w}}_{\mathbf{i}}) = \hat{\mathbf{u}}_{\mathbf{i}} \mid \hat{\mathbf{w}}_{\mathbf{i}} \neq \mathbf{w}\} \quad (53)$$

column space of $\mathbf{F}_{p,i}$, and $\mathbf{\Pi}_{p,i}^\perp \triangleq \mathbf{I} - \mathbf{\Pi}_{p,i}$ 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 $\mathbf{e}_{p,i}^{(k)}$:

$$\hat{i}_p^{(k)} \triangleq \arg \min_{i \in \{1, \dots, M\}} \|\mathbf{e}_{p,i}^{(k)}\|^2 \text{ for } \mathbf{e}_{p,i}^{(k)} \triangleq \mathbf{\Pi}_{p,i}^\perp \mathbf{z}_p^{(k)} \quad (57)$$

Clearly, the complexity of PASE is proportional to $M = \binom{L}{S} = \mathcal{O}((L/S)^S)$. Thus, while the complexity of PASE is much less than the DASD proposed in Section IV-A, we note that its complexity may be significantly larger than classical compressive sensing algorithms like basis pursuit, whose complexity is polynomial in L [21].

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

Proof: We first note that, due to the Chebotarev theorem [16], [17], each $\mathbf{F}_{p,i} \in \mathbb{C}^{P \times S}$ is full rank when N is prime and $P \geq S + 1$. Also, each column \mathbf{f} of $\mathbf{F}_{p,i}$ is linearly independent of all columns in $\mathbf{F}_{p,j} \Big|_{j \neq i}$ that are not equal to \mathbf{f} . Thus, each $\mathbf{F}_{p,i}$ defines a *unique* column space. We note that this property does not hold when $P = S$.

A PASE support-detection error results when $\exists i \neq i_{\text{true}}^{(k)}$ s.t. $\|\mathbf{e}_{p,i}^{(k)}\|^2 < \|\mathbf{e}_{p,\text{true}}^{(k)}\|^2$. The probability of this event can be upper bounded as shown in (58)–(62) at the bottom of the page, where the probability of error in (60) was upper-bounded by making the left side of the inequality smaller via $\|\mathbf{x}\| - \|\mathbf{y}\| \leq \|\mathbf{x} + \mathbf{y}\|$. The upper bound (62) follows from $\|\mathbf{\Pi}_{p,i}^\perp \mathbf{\nu}_p^{(k)}\| \leq \|\mathbf{\nu}_p^{(k)}\|$ and $\|\mathbf{\Pi}_{p,\text{true}}^{(k)\perp} \mathbf{\nu}_p^{(k)}\| \leq \|\mathbf{\nu}_p^{(k)}\|$, which hold because $\mathbf{\Pi}_{p,i}^\perp$ and $\mathbf{\Pi}_{p,\text{true}}^{(k)\perp}$ are projection matrices. Taking the SVD $\mathbf{\Pi}_{p,i}^\perp \mathbf{F}_{p,\text{true}}^{(k)} = \mathbf{U}_i^{(k)} \mathbf{\Sigma}_i^{(k)} \mathbf{V}_i^{(k)\text{H}}$ and defining

$\mathbf{g}_i^{(k)} \triangleq \sqrt{S} \mathbf{V}_i^{(k)\text{H}} \mathbf{h}_{\text{nz}}^{(k)} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, we can rewrite (62) as follows and upper bound further:

$$\Pr \left\{ \exists i \neq i_{\text{true}}^{(k)} \text{ s.t. } \|\mathbf{e}_{p,i}^{(k)}\|^2 < \|\mathbf{e}_{p,\text{true}}^{(k)}\|^2 \right\} \leq \sum_{i \neq i_{\text{true}}^{(k)}} \Pr \left\{ \|\mathbf{\Sigma}_i^{(k)} \mathbf{g}_i^{(k)}\|^2 < \frac{4S}{\rho N} \|\mathbf{\nu}_p^{(k)}\|^2 \right\} \quad (63)$$

$$\leq \sum_{i \neq i_{\text{true}}^{(k)}} \Pr \left\{ (\sigma_{i,0}^{(k)})^2 |g_{i,0}^{(k)}|^2 < \frac{4S}{\rho N} \|\mathbf{\nu}_p^{(k)}\|^2 \right\} \quad (64)$$

$$\leq \sum_{i \neq i_{\text{true}}^{(k)}} \Pr \left\{ (\sigma_{i,0}^{(\min)})^2 |g_{i,0}^{(k)}|^2 < \frac{4S}{\rho N} \|\mathbf{\nu}_p^{(k)}\|^2 \right\} \quad (65)$$

$$= \sum_{i \neq i_{\text{true}}^{(k)}} \Pr \left\{ \frac{|g_{i,0}^{(k)}|^2}{\|\mathbf{\nu}_p^{(k)}\|^2} < \frac{4S}{(\sigma_{i,0}^{(\min)})^2 \rho N} \right\}. \quad (66)$$

Above, $\sigma_{i,0}^{(k)}$ denotes the largest singular value in $\mathbf{\Sigma}_i^{(k)}$ and $\sigma_{i,0}^{(\min)} \triangleq \min_k \sigma_{i,0}^{(k)}$. Notice that at least one of the columns of $\mathbf{F}_{p,\text{true}}^{(k)}$ lies outside the column space of $\mathbf{F}_{p,i}$. The projection of those columns onto the subspace orthogonal to the column space of $\mathbf{F}_{p,i}$ will be non-zero implying that $\mathbf{\Pi}_{p,i}^\perp \mathbf{F}_{p,\text{true}}^{(k)}$ is not identical to $\mathbf{0}$ 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 $\mathbf{\nu}_p^{(k)} \sim \mathcal{CN}(0, \mathbf{I})$, the random variable $F_i^{(k)} \triangleq |g_{i,0}^{(k)}|^2 / \|\mathbf{\nu}_p^{(k)}\|^2$ is F-distributed with parameters $(2, 2P)$. Since the cumulative distribution function (cdf) 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 $\rho \rightarrow \infty$. ■

We now make a few comments about Theorem 3. To perfectly recover any *arbitrary deterministic* S -sparse impulse response from noise-free frequency-domain samples, [10] established that $2S$ pilot tones are both necessary and sufficient. In contrast, to perfectly recover an S -sparse *probabilistic* Rayleigh-fading impulse response, Theorem 3 establishes that $S + 1$ noise-free pilot observations suffice with probability one. In particular, the

$$\Pr \left\{ \exists i \neq i_{\text{true}}^{(k)} \text{ s.t. } \|\mathbf{e}_{p,i}^{(k)}\|^2 < \|\mathbf{e}_{p,\text{true}}^{(k)}\|^2 \right\} \leq \sum_{i \neq i_{\text{true}}^{(k)}} \Pr \left\{ \|\mathbf{e}_{p,i}^{(k)}\|^2 < \|\mathbf{e}_{p,\text{true}}^{(k)}\|^2 \right\} \quad (58)$$

$$= \sum_{i \neq i_{\text{true}}^{(k)}} \Pr \left\{ \|\mathbf{\Pi}_{p,i}^\perp \mathbf{F}_{p,\text{true}}^{(k)} \mathbf{h}_{\text{nz}}^{(k)} + \frac{1}{\sqrt{\rho N}} \mathbf{\Pi}_{p,i}^\perp \mathbf{\nu}_p^{(k)}\| < \frac{1}{\sqrt{\rho N}} \|\mathbf{\Pi}_{p,\text{true}}^{(k)\perp} \mathbf{\nu}_p^{(k)}\| \right\} \quad (59)$$

$$\leq \sum_{i \neq i_{\text{true}}^{(k)}} \Pr \left\{ \|\mathbf{\Pi}_{p,i}^\perp \mathbf{F}_{p,\text{true}}^{(k)} \mathbf{h}_{\text{nz}}^{(k)}\| - \frac{1}{\sqrt{\rho N}} \|\mathbf{\Pi}_{p,i}^\perp \mathbf{\nu}_p^{(k)}\| < \frac{1}{\sqrt{\rho N}} \|\mathbf{\Pi}_{p,\text{true}}^{(k)\perp} \mathbf{\nu}_p^{(k)}\| \right\} \quad (60)$$

$$= \sum_{i \neq i_{\text{true}}^{(k)}} \Pr \left\{ \|\mathbf{\Pi}_{p,i}^\perp \mathbf{F}_{p,\text{true}}^{(k)} \mathbf{h}_{\text{nz}}^{(k)}\| < \frac{1}{\sqrt{\rho N}} \|\mathbf{\Pi}_{p,i}^\perp \mathbf{\nu}_p^{(k)}\| + \frac{1}{\sqrt{\rho N}} \|\mathbf{\Pi}_{p,\text{true}}^{(k)\perp} \mathbf{\nu}_p^{(k)}\| \right\} \quad (61)$$

$$\leq \sum_{i \neq i_{\text{true}}^{(k)}} \Pr \left\{ \|\mathbf{\Pi}_{p,i}^\perp \mathbf{F}_{p,\text{true}}^{(k)} \mathbf{h}_{\text{nz}}^{(k)}\| < \frac{2}{\sqrt{\rho N}} \|\mathbf{\nu}_p^{(k)}\| \right\}, \quad (62)$$

condition $P \geq S + 1$ ensures that the set of $\mathbf{h}^{(k)}$ that cannot be recovered by the PASE support detector has probability 0 with respect to the Gaussian distribution on the nonzero entries of $\mathbf{h}^{(k)}$. To see this, notice that $\text{rank}(\mathbf{F}_{p,i}) = \text{rank}(\mathbf{F}_{p,j}) = S$, but also that $\text{range}(\mathbf{F}_{p,i}) = \text{range}(\mathbf{F}_{p,j})$ only if $i = j$. In particular, if $i \neq j$, then $\dim\{\text{range}(\mathbf{F}_{p,i}) \cap \text{range}(\mathbf{F}_{p,j})\} = S - 1$. This implies that the set of vectors $\mathbf{h}_{\text{nz}} \in \mathbb{C}^S$ for which $\mathbf{F}_{p,i}\mathbf{h}_{\text{nz}}$ is in the range space of $\mathbf{F}_{p,j}$ has measure zero with respect to any continuous distribution on \mathbf{h}_{nz} . Similar results on the recovery of probabilistic sparse signals have also appeared in [22].

2) *Pilot-Aided Support Decoding*: For pilot-aided support decoding, we assume that the transmitter uses the PAT scheme defined in Section III-A with $P = S + 1$ pilots and prime N . At the receiver, the PASE scheme described in the previous section is used to estimate the sparse channel support and, based on this estimate, support-conditional channel estimation and decoupled data decoding are performed as described in Section III-C.

We now study the ϵ -achievable rate of PAT with PASD. 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 4 characterizes $\mathcal{R}_\epsilon(\rho)$ for PAT with PASD.

Lemma 4: 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 $\mathcal{R}_\epsilon^{\text{PASD}}$ that, for any $\epsilon > 0$, obeys $\lim_{\rho \rightarrow \infty} \frac{\mathcal{R}_\epsilon^{\text{PASD}}(\rho)}{\log_2 \rho} = 1 - \frac{S+1}{N}$.

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 $\epsilon/2$. In the case that the support hypothesis is correct, the channel estimation and decoupled decoding of Section III-C allow for the design of a codebook $\mathcal{C}_{\rho,\epsilon}$ that guarantees data decoding with error probability less than $\epsilon/2$ at SNR ρ . Furthermore, from Lemma 3, this codebook can be designed with a rate $R_\epsilon(\rho)$ such that $\lim_{\rho \rightarrow \infty} \frac{R_\epsilon(\rho)}{\log_2 \rho} = 1 - \frac{S+1}{N}$. Putting these together, we obtain the result of the lemma. ■

We note that, for any given finite SNR ρ , it is not possible to make ϵ , the PASD error probability, arbitrarily small. Thus, the achievable rate $\mathcal{R}(\rho)$ of PAT with PASD equals zero for any finite ρ . This behavior contrasts that of PAT with DASD, which had positive achievable rate for all $\rho > 0$.

Recall that, with the sparse block-fading channel model assumed throughout the paper, the channel support $\mathcal{L}^{(k)}$ changes independently over fading blocks k . We now consider a variation of this channel for which the support does not change¹² over k . For this fixed-support channel, it is possible to modify PASE so that it recovers the support \mathcal{L} with an arbitrarily small probability of error at any SNR $\rho > 0$, leading to the following corollary of Lemma 4.

Corollary 3: For the S -sparse frequency-selective N -block-fading channel with prime N and a support $\{\mathcal{L}^{(k)}\}_{k=1}^K$ that is constant over the fading block index k , the previously defined

¹²Although the support $\mathcal{L}^{(k)}$ remains fixed over k , the nonzero channel taps $\mathbf{h}_{\text{nz}}^{(k)}$ still vary independently over k .

PAT scheme, when used with $S + 1$ pilots and PASD, yields an achievable rate $\mathcal{R}^{\text{PASD}}$ that obeys $\lim_{\rho \rightarrow \infty} \frac{\mathcal{R}^{\text{PASD}}(\rho)}{\log_2 \rho} = 1 - \frac{S+1}{N}$.

Proof: For this channel, we use PASE with the metric $\frac{1}{K} \sum_{k=1}^K \|\mathbf{e}_{p,i}^{(k)}\|^2$ in place of the metric $\|\mathbf{e}_{p,i}^{(k)}\|^2$ from (57). With this modification, we obtain an error probability upper-bound analogous to (66), but where the F-distributed random variable has parameters $(2K, 2K(S + 1))$. In particular,

$$\Pr \left\{ \exists i \neq i_{\text{true}} \text{ s.t. } \sum_{k=1}^K \|\mathbf{e}_{p,i}^{(k)}\|^2 < \sum_{k=1}^K \|\mathbf{e}_{p,i_{\text{true}}}^{(k)}\|^2 \right\} \\ \leq \sum_{i \neq i_{\text{true}}} \Pr \left\{ \frac{\sum_{k=1}^K |g_{i,0}^{(k)}|^2}{\sum_{k=1}^K \|\boldsymbol{\nu}_p^{(k)}\|^2} < \frac{S}{(\sigma_{i,0}^{(\min)})^2 \rho N} \right\}. \quad (67)$$

For an F-distributed random variable with parameters $(2K, 2K(S + 1))$, the value of the cdf at any fixed point decreases with K . Thus, by choosing a suitably large K , we can make the PASE support-detection error arbitrarily small at any SNR $\rho > 0$. The result of this lemma then follows from Lemma 3. ■

V. CONCLUSION

In this paper, we considered the problem of communicating reliably over frequency-selective block-fading channels whose impulse responses are sparse and whose realizations are unknown to both transmitter and receiver, but whose statistics are known. In particular, we considered discrete-time channel impulse responses with length L and sparsity exactly $S \leq L$, whose support and coefficients remain fixed over blocks of $N > L$ channel uses but change independently from block to block.

Assuming that the non-zero coefficients and noise are both Gaussian, we first established that the ergodic noncoherent channel capacity $\mathcal{C}_{\text{sparse}}(\rho)$ obeys $\lim_{\rho \rightarrow \infty} \frac{\mathcal{C}_{\text{sparse}}(\rho)}{\log_2 \rho} = 1 - \frac{S}{N}$ for any L . Then, we shifted our focus to pilot-aided transmission (PAT), where we constructed a PAT scheme and a so-called data-aided support decoder (DASD) that together enable communication with arbitrarily small error probability using only S pilots per fading block. Furthermore, we showed that the achievable rate $\mathcal{R}^{\text{DASD}}(\rho)$ of this pair exhibits the optimal pre-log factor, i.e., $\lim_{\rho \rightarrow \infty} \frac{\mathcal{R}^{\text{DASD}}(\rho)}{\log_2 \rho} = 1 - \frac{S}{N}$. The use of S pilots can be contrasted with ‘‘compressed OFDM channel sensing,’’ for which $\mathcal{O}(S \ln^5 N)$ pilots are known to suffice for accurate channel estimation (with high probability) in the presence of noise, and for which $2S$ pilots are known to be necessary and sufficient for perfect channel estimation in the absence of noise.

Due to the complexity of DASD, we also proposed a simpler pilot-aided support decoder (PASD) that requires only $S + 1$ pilots per fading block. For PASD, the ϵ -achievable rate $\mathcal{R}_\epsilon^{\text{PASD}}(\rho)$ obeys, for any $\epsilon > 0$, $\lim_{\rho \rightarrow \infty} \frac{\mathcal{R}_\epsilon^{\text{PASD}}(\rho)}{\log_2 \rho} = 1 - \frac{S+1}{N}$ with the previously considered channel, and its achievable rate $\mathcal{R}^{\text{PASD}}(\rho)$ obeys $\lim_{\rho \rightarrow \infty} \frac{\mathcal{R}^{\text{PASD}}(\rho)}{\log_2 \rho} = 1 - \frac{S+1}{N}$ when the sparsity pattern of the block-fading channel remains fixed over fading blocks. We note that, in recent work [11], [12], the authors have proposed a loopy belief propagation based joint channel estimation and

decoding scheme, with complexity $\mathcal{O}(KLN)$, that shows empirical performance that matches the anticipated pre-log factor of $1 - \frac{S}{N}$.

The results of this work are only a first step towards the understanding of reliable communication over sparse channels. Important open questions concern rigorous analyses of the cases that i) the inactive channel taps are not exactly zero-valued, ii) the channel has *at most* (rather than exactly) S active taps, iii) the receiver does not know the channel statistics, iv) the channel taps are correlated within and/or across blocks, and/or v) the active channel taps are non-Gaussian.

APPENDIX PROOF OF LEMMA 1

Proof: The maximum a posteriori (MAP) codeword estimate is defined as shown in (68)–(69) at the bottom of the page, where (69) results after applying Bayes rule and simplifying. Assuming that codewords are uniformly distributed over \mathcal{C} , the MAP codeword estimate reduces to the maximum likelihood estimate (70)–(72) at the bottom of the page, where the decoupling in (71) is due to independent fading and noise across

fading-blocks. Recalling that, under the hypothesis $\mathcal{L}^{(k)} = \mathcal{L}_i$, the pilot observations become

$$\mathbf{y}_p^{(k)} = \sqrt{\rho N} \mathcal{D}(\mathbf{x}_p) \mathbf{F}_{p,i} \mathbf{h}_{nz}^{(k)} + \mathbf{v}_p^{(k)}, \quad (73)$$

with $p(\mathbf{h}_{nz}^{(k)} | \mathcal{L}^{(k)} = \mathcal{L}_i) = \mathcal{CN}(\mathbf{h}_{nz}^{(k)}; \mathbf{0}, S^{-1} \mathbf{I})$, the posterior $p(\mathbf{h}_{nz}^{(k)} | \mathbf{y}_p^{(k)}, \mathbf{x}_p, \mathcal{L}^{(k)} = \mathcal{L}_i)$ is Gaussian. In particular,

$$p(\mathbf{h}_{nz}^{(k)} | \mathbf{y}_p^{(k)}, \mathbf{x}_p, \mathcal{L}^{(k)} = \mathcal{L}_i) = \mathcal{CN}(\mathbf{h}_{nz}^{(k)}; \hat{\mathbf{h}}_{nz,p,i}^{(k)}, \boldsymbol{\Sigma}_{nz,p,i}^{(k)}), \quad (74)$$

where $\hat{\mathbf{h}}_{nz,p,i}^{(k)}$ can be recognized as the \mathcal{L}_i -conditional pilot-aided MMSE estimate of $\mathbf{h}_{nz}^{(k)}$ and $\boldsymbol{\Sigma}_{nz,p,i}^{(k)}$ as its error covariance:

$$\hat{\mathbf{h}}_{nz,p,i}^{(k)} \triangleq \mathbb{E}\{\mathbf{h}_{nz}^{(k)} | \mathbf{y}_p^{(k)}, \mathbf{x}_p, \mathcal{L}^{(k)} = \mathcal{L}_i\} \quad (75)$$

$$\boldsymbol{\Sigma}_{nz,p,i}^{(k)} \triangleq \text{cov}\{\mathbf{h}_{nz}^{(k)} | \mathbf{y}_p^{(k)}, \mathbf{x}_p, \mathcal{L}^{(k)} = \mathcal{L}_i\}. \quad (76)$$

Due to the linear Gaussian model (73), the MMSE estimate $\hat{\mathbf{h}}_{nz,p,i}^{(k)}$ is a linear function of $\mathbf{y}_p^{(k)}$: See equation (77)–(79) at the bottom of the page where, for (79), we exploited the fact that

$$\hat{\mathbf{x}}_d^{\text{MAP}} = \arg \max_{\mathbf{x}_d \in \mathcal{C}} p(\mathbf{x}_d | \{\mathbf{y}_d^{(k)}\}_{k=1}^K, \{\mathbf{y}_p^{(k)}\}_{k=1}^K, \mathbf{x}_p) \quad (68)$$

$$= \arg \max_{\mathbf{x}_d \in \mathcal{C}} p(\{\mathbf{y}_d^{(k)}\}_{k=1}^K | \mathbf{x}_d, \{\mathbf{y}_p^{(k)}\}_{k=1}^K, \mathbf{x}_p) p(\mathbf{x}_d) \quad (69)$$

$$\hat{\mathbf{x}}_d^{\text{ML}} = \arg \max_{\mathbf{x}_d \in \mathcal{C}} p(\{\mathbf{y}_d^{(k)}\}_{k=1}^K | \{\mathbf{x}_d^{(k)}\}_{k=1}^K, \{\mathbf{y}_p^{(k)}\}_{k=1}^K, \mathbf{x}_p) \quad (70)$$

$$\begin{aligned} &= \arg \max_{\mathbf{x}_d \in \mathcal{C}} \prod_{k=1}^K \sum_{i=1}^M \Pr\{\mathcal{L}^{(k)} = \mathcal{L}_i | \mathbf{x}_d^{(k)}, \mathbf{y}_p^{(k)}, \mathbf{x}_p\} \\ &\quad \times \int_{\mathbf{h}_{nz}^{(k)}} p(\mathbf{y}_d^{(k)} | \mathbf{x}_d^{(k)}, \mathbf{y}_p^{(k)}, \mathbf{x}_p, \mathbf{h}_{nz}^{(k)}, \mathcal{L}^{(k)} = \mathcal{L}_i) p(\mathbf{h}_{nz}^{(k)} | \mathbf{x}_d^{(k)}, \mathbf{y}_p^{(k)}, \mathbf{x}_p, \mathcal{L}^{(k)} = \mathcal{L}_i) \end{aligned} \quad (71)$$

$$\begin{aligned} &= \arg \max_{\mathbf{x}_d \in \mathcal{C}} \prod_{k=1}^K \sum_{i=1}^M \Pr\{\mathcal{L}^{(k)} = \mathcal{L}_i | \mathbf{y}_p^{(k)}, \mathbf{x}_p\} \\ &\quad \times \int_{\mathbf{h}_{nz}^{(k)}} p(\mathbf{y}_d^{(k)} | \mathbf{x}_d^{(k)}, \mathbf{h}_{nz}^{(k)}, \mathcal{L}^{(k)} = \mathcal{L}_i) p(\mathbf{h}_{nz}^{(k)} | \mathbf{y}_p^{(k)}, \mathbf{x}_p, \mathcal{L}^{(k)} = \mathcal{L}_i), \end{aligned} \quad (72)$$

$$\hat{\mathbf{h}}_{nz,p,i}^{(k)} = \mathbb{E}\{\mathbf{h}_{nz}^{(k)} \mathbf{y}_p^{(k)\text{H}} | \mathbf{x}_p, \mathcal{L}^{(k)} = \mathcal{L}_i\} \mathbb{E}\{\mathbf{y}_p^{(k)} \mathbf{y}_p^{(k)\text{H}} | \mathbf{x}_p, \mathcal{L}^{(k)} = \mathcal{L}_i\}^{-1} \mathbf{y}_p^{(k)} \quad (77)$$

$$= \frac{\sqrt{\rho N}}{S} \mathbf{F}_{p,i}^{\text{H}} \mathcal{D}(\mathbf{x}_p^*) (\rho \frac{N}{S} \mathcal{D}(\mathbf{x}_p) \mathbf{F}_{p,i} \mathbf{F}_{p,i}^{\text{H}} \mathcal{D}(\mathbf{x}_p^*) + \mathbf{I})^{-1} \mathbf{y}_p^{(k)} \quad (78)$$

$$= \sqrt{\frac{\rho}{N}} \mathbf{F}_{p,i}^{\text{H}} (\rho \mathbf{F}_{p,i} \mathbf{F}_{p,i}^{\text{H}} + \frac{S}{N} \mathbf{I})^{-1} \mathcal{D}(\mathbf{x}_p^*) \mathbf{y}_p^{(k)}, \quad (79)$$

$$\begin{aligned} \Sigma_{\text{nz},p,i} &= \mathbb{E}\{\mathbf{h}_{\text{nz}}^{(k)}\mathbf{h}_{\text{nz}}^{(k)\text{H}} \mid \mathcal{L}^{(k)} = \mathcal{L}_i\} - \mathbb{E}\{\mathbf{h}_{\text{nz}}^{(k)}\mathbf{y}_p^{(k)\text{H}} \mid \mathbf{x}_p, \mathcal{L}^{(k)} = \mathcal{L}_i\} \\ &\quad \times \mathbb{E}\{\mathbf{y}_p^{(k)}\mathbf{y}_p^{(k)\text{H}} \mid \mathbf{x}_p, \mathcal{L}^{(k)} = \mathcal{L}_i\}^{-1} \mathbb{E}\{\mathbf{y}_p^{(k)}\mathbf{h}_{\text{nz}}^{(k)\text{H}} \mid \mathbf{x}_p, \mathcal{L}^{(k)} = \mathcal{L}_i\} \end{aligned} \quad (80)$$

$$= \frac{1}{S}\mathbf{I} - \frac{\rho N}{S^2}\mathbf{F}_{p,i}^{\text{H}}\mathcal{D}(\mathbf{x}_p^*)\left(\frac{\rho N}{S}\mathcal{D}(\mathbf{x}_p)\mathbf{F}_{p,i}\mathbf{F}_{p,i}^{\text{H}}\mathcal{D}(\mathbf{x}_p^*) + \mathbf{I}\right)^{-1}\mathcal{D}(\mathbf{x}_p)\mathbf{F}_{p,i} \quad (81)$$

$$= \frac{1}{S}\left(\mathbf{I} - \mathbf{F}_{p,i}^{\text{H}}(\mathbf{F}_{p,i}\mathbf{F}_{p,i}^{\text{H}} + \frac{S}{\rho N}\mathbf{I})^{-1}\mathbf{F}_{p,i}\right). \quad (82)$$

$$\begin{aligned} &\int_{\mathbf{h}_{\text{nz}}^{(k)}} p(\mathbf{y}_d^{(k)} \mid \mathbf{x}_d^{(k)}, \mathbf{h}_{\text{nz}}^{(k)}, \mathcal{L}^{(k)} = \mathcal{L}_i) p(\mathbf{h}_{\text{nz}}^{(k)} \mid \mathbf{y}_p^{(k)}, \mathbf{x}_p, \mathcal{L}^{(k)} = \mathcal{L}_i) \\ &= C \det\left(\rho N \mathbf{F}_{d,i}^{\text{H}} \mathcal{D}(\mathbf{x}_d^{(k)}) \odot \mathbf{x}_d^{(k)*} \mathbf{F}_{d,i} + \Sigma_{\text{nz},p,i}^{-1}\right)^{-1} \\ &\quad \times \exp\left(-\|\mathbf{y}_d^{(k)} - \sqrt{\rho N} \mathcal{D}(\mathbf{x}_d^{(k)}) \mathbf{F}_{d,i} \hat{\mathbf{h}}_{\text{nz},i}^{(k)}(\mathbf{x}_d^{(k)})\|^2 - \|\hat{\mathbf{h}}_{\text{nz},i}^{(k)}(\mathbf{x}_d^{(k)}) - \hat{\mathbf{h}}_{\text{nz},p,i}^{(k)}\|_{\Sigma_{\text{nz},p,i}^{-1}}^2\right), \end{aligned} \quad (83)$$

$$\begin{aligned} \hat{\mathbf{h}}_{\text{nz},i}^{(k)}(\mathbf{x}_d^{(k)}) &= \hat{\mathbf{h}}_{\text{nz},p,i}^{(k)} + \sqrt{\rho N} \Sigma_{\text{nz},p,i} \mathbf{F}_{d,i}^{\text{H}} \mathcal{D}(\mathbf{x}_d^{(k)*}) (\rho N \mathcal{D}(\mathbf{x}_d^{(k)}) \mathbf{F}_{d,i} \Sigma_{\text{nz},p,i} \mathbf{F}_{d,i}^{\text{H}} \mathcal{D}(\mathbf{x}_d^{(k)*}) + \mathbf{I})^{-1} \\ &\quad \times (\mathbf{y}_d^{(k)} - \sqrt{\rho N} \mathcal{D}(\mathbf{x}_d^{(k)}) \mathbf{F}_{d,i} \hat{\mathbf{h}}_{\text{nz},p,i}^{(k)}). \end{aligned} \quad (84)$$

\mathbf{x}_p has constant-modulus elements. Similarly, see (80)–(82) at the top of the page.

Finally, since both pdfs in (72) are Gaussian, the integral can be evaluated in closed form, reducing to (see, e.g., [23]) (83) at the top of the page where C does not depend on \mathbf{x}_d , and where $\hat{\mathbf{h}}_{\text{nz},i}^{(k)}(\mathbf{x}_d^{(k)})$ denotes the MMSE estimate of $\mathbf{h}_{\text{nz}}^{(k)}$ conditioned on the data hypothesis $\mathbf{x}_d^{(k)}$ and based on the pilot-aided prior statistics (74): See equation (84) at the top of the page. ■

ACKNOWLEDGMENT

The authors thank the anonymous reviewers for their insightful comments.

REFERENCES

- [1] R. J.-M. Cramer, R. A. Scholtz, and M. Z. Win, "Evaluation of an ultra-wide-band propagation channel," *IEEE Trans. Antennas Propagat.*, vol. 50, pp. 561–570, May 2002.
- [2] J. C. Preisig and G. Deane, "Surface wave focusing and acoustic communications in the surf zone," *J. Acoust. Soc. Amer.*, vol. 116, pp. 2067–2080, Oct. 2004.
- [3] A. F. Molisch, "Ultrawideband propagation channels—Theory, measurement, and modeling," *IEEE Trans. Veh. Technol.*, vol. 54, pp. 1528–1545, Sep. 2005.
- [4] 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, Jun. 2010.
- [5] P. Schniter, "A message-passing receiver for BICM-OFDM over unknown clustered-sparse channels," *IEEE J. Sel. Topics Signal Process. (Special Issue on Soft Detection for Wireless Transmission)* (see also arXiv:1101.4724), to be published.
- [6] A. F. Molisch, K. Balakrishnan, C.-C. Chong, S. Emami, A. Fort, J. Karedal, J. Kunisch, H. Schantz, U. Schuster, and K. Siwiak, IEEE 802.15.4a Channel Model—Final Report, Tech. Rep., Document IEEE 802.1504-0062-02-004a, 2005.
- [7] 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.*, vol. 52, no. 9, pp. 2572–2583, Sep. 2004.
- [8] Y. Liang and V. Veeravalli, "Capacity of noncoherent time-selective Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3095–3110, Dec. 2004.
- [9] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. New York: Cambridge Univ. Press, 2005.
- [10] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [11] P. Schniter, "Joint estimation and decoding for sparse channels via relaxed belief propagation," in *Proc. Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, Nov. 2010.
- [12] P. Schniter, "Belief-propagation-based joint channel estimation and decoding for spectrally efficient communication over unknown sparse channels," *Phys. Commun. (Special Issue on Compressive Sensing in Communications)* (see also arXiv:1012.4519), to be published.
- [13] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
- [14] M. Clyde and E. I. George, "Model uncertainty," *Statist. Sci.*, vol. 19, no. 1, pp. 81–94, 2004.
- [15] H. Weingarten, Y. Steinberg, and S. S. Shamai (Shitz), "Gaussian codes and weighted nearest neighbor decoding in fading multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1665–1686, Aug. 2004.
- [16] P. Steinhagen and H. Lenstra, "Chebotarev and his density theorem," *Math. Intell.*, vol. 18, no. 2, pp. 26–37, 1996.
- [17] T. Tao, "An uncertainty principle for cyclic groups of prime order," *Math. Res. Letter*, vol. 12, pp. 121–127, 2005.
- [18] A. P. Kannu and P. Schniter, "On the spectral efficiency of noncoherent doubly selective block-fading channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2829–2844, Jun. 2010.
- [19] S. B. Wicker, *Error Control Systems for Digital Communication and Storage*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
- [20] H. El Gamal, G. Caire, and M. O. Damen, "The MIMO ARQ channel: Diversity-multiplexing-delay tradeoff," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3601–3621, Aug. 2006.
- [21] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 1289–1306, Sep. 2006.
- [22] J. A. Tropp, "On the conditioning of random subdictionaries," *Appl. Comput. Harmonic Anal.*, vol. 25, pp. 1–24, Mar. 2008.
- [23] D. J. Reader and W. G. Cowley, "Blind maximum likelihood sequence detection over fast fading channels," in *Proc. Eur. Signal Process. Conf.*, Trieste, Italy, Sep. 1996.

Arun Pachai Kannu received the M.S. and Ph.D. degrees in Electrical Engineering from The Ohio State University in 2004 and 2007, respectively. From 2007 to 2009, he was a Senior Engineer in Qualcomm Inc., San Diego, CA. He is currently an Assistant Professor in the Department of Electrical Engineering, Indian Institute of Technology, Madras.

Philip Schniter (SM'05) received the B.S. and M.S. degrees in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign in 1992 and 1993, respectively. From 1993 to 1996 he was employed by Tektronix Inc. in Beaverton, OR as a systems engineer, and in 2000, he received the Ph.D. degree in Electrical Engineering from Cornell University in Ithaca, NY. Subsequently, he joined the Department of Electrical and Computer Engineering at The Ohio State University in Columbus, OH, where he is now an Associate Professor and a member of the Information Processing Systems (IPS) Lab. In 2003, he received the National Science Foundation CAREER Award, and in 2008–2009 he was a visiting professor at Eurecom (Sophia Antipolis, France) and Supélec (Gif-sur-Yvette, France).