

What is the Value of Joint Processing of Pilots and Data in Block-Fading Channels?

Nihar Jindal
University of Minnesota
Minneapolis, MN 55455
Email: nihar@umn.edu

Angel Lozano
Universitat Pompeu Fabra
Barcelona 08003, Spain
Email: angel.lozano@upf.edu

Thomas L. Marzetta
Bell Labs (Alcatel-Lucent)
Murray Hill, NJ 07974, USA
Email: tlm@research.bell-labs.com

Abstract—The spectral efficiency achievable with joint processing of pilot and data symbol observations is compared with that achievable through the conventional (separate) approach of first estimating the channel on the basis of the pilot symbols alone, and subsequently detecting the data symbols. Studied on the basis of a mutual information lower bound, joint processing is found to provide a non-negligible advantage relative to separate processing, particularly for fast fading. It is shown that, regardless of the fading rate, only a very small number of pilot symbols (at most one per transmit antenna and per channel coherence interval) should be transmitted if joint processing is allowed.

I. INTRODUCTION

Pilot symbols (a.k.a. training or reference symbols) are an inherent part of virtually every wireless system. Motivated by this prevalence, the spectral efficiency achievable when coherently detecting data with the assistance of pilots has been the object of much analysis (e.g., [1]–[5]). A large fraction of such work has focused on the spectral efficiency achievable with Gaussian inputs under the assumption that the fading channel is estimated on the basis of the pilot observations and then, using such estimate as it were the true channel, the data is detected. Although suboptimal, such separate processing reflects the operating conditions of existing systems.

In this paper, we move beyond this approach and quantify the advantage of jointly processing pilot and data observations when Gaussian codebooks are utilized. Since the general mutual information expression is intractable, we rely on lower bounds to the achievable spectral efficiency. These bounds allow assessing the optimum number of pilot symbols under such joint processing, and also quantify the minimum improvement in spectral efficiency that joint processing brings about relative to separate processing.

Although there has been prior work on receiver design for joint processing (e.g., [6]–[8]), to the best of our knowledge there is not yet a general understanding of the conditions (in terms of signal-to-noise ratio, fading rate, and antenna configurations) in which joint processing provides a substantial improvement. Given that joint processing is more complex than separate processing, such a quantification appears very useful.

As a starting point, a simple block-fading ergodic channel model is considered. Section II restricts itself to scalar channels, from which many of the insights can already be derived. The generalization to MIMO (multiple-input multiple-output) follows in Section III.

II. SISO

A. Channel Model

Let H represent a discrete-time scalar fading channel. Under block Rayleigh-fading, the channel is drawn from a zero-mean complex Gaussian distribution at the beginning of each block and it then remains constant for the T symbols composing the block, where T corresponds to the coherence time/bandwidth. This process is repeated for every block in an IID (independent identically distributed) fashion. A total of τ pilot symbols are inserted within each block leaving $T - \tau$ symbols available for data.

During the transmission of pilot symbols,

$$\mathbf{y}_p = \sqrt{\text{SNR}} H + \mathbf{n}_p \quad (1)$$

where the received signal, \mathbf{y}_p , and the noise, \mathbf{n}_p , are τ -dimensional vectors. The entries of \mathbf{n}_p are IID zero-mean unit-variance complex Gaussian. The channel satisfies $\mathbb{E}[|H|^2] = 1$ and thus SNR indicates the average signal-to-noise ratio. During the transmission of data symbols

$$\mathbf{y}_d = \sqrt{\text{SNR}} H \mathbf{x} + \mathbf{n}_d \quad (2)$$

where \mathbf{y}_d , \mathbf{n}_d , and the transmitted data \mathbf{x} , are all $(T - \tau)$ -dimensional. The noise \mathbf{n}_d is independent of \mathbf{n}_p but it abides by the same distribution. As argued in the Introduction, the entries of \mathbf{x} are IID zero-mean unit-variance complex Gaussian. Each transmitted codeword spans a large number of fading blocks, which endows ergodic quantities with operational meaning.

B. Perfect CSI

If the receiver is provided with perfect CSI (channel-state information), Gaussian codebooks are capacity-achieving and the ergodic capacity, in bits/s/Hz, equals

$$C(\text{SNR}) = \mathbb{E} [\log_2 (1 + \text{SNR} |H|^2)] \quad (3)$$

$$= e^{1/\text{SNR}} E_1 \left(\frac{1}{\text{SNR}} \right) \log_2 e \quad (4)$$

where $E_k(\cdot)$ is the exponential integral of order k . For compactness, $C(\text{SNR})$ is often abbreviated as C .

C. Separated Processing of Pilots and Data

If the receiver uses the pilot observations, \mathbf{y}_p , to first produce an MMSE estimate of the channel, \hat{H} , and then performs nearest-neighbor decoding while treating \hat{H} as if it were H , the maximum spectral efficiency is [5]

$$I_S = \max_{\tau: 1 \leq \tau < T} \left\{ \left(1 - \frac{\tau}{T}\right) C(\text{SNR}_{\text{eff}}) \right\} \quad (5)$$

with

$$\text{SNR}_{\text{eff}} = \frac{\text{SNR}(1 - \text{MMSE})}{1 + \text{SNR} \cdot \text{MMSE}} \quad (6)$$

and $\text{MMSE} = \mathbb{E}[|H - \hat{H}|^2] = 1/(1 + \text{SNR}\tau)$. The maximization in (5) must be computed numerically as no closed form exists.

D. Spectral Efficiency Lower Bounds for Joint Processing

In the general case, the receiver decodes the data based upon \mathbf{y}_p and \mathbf{y}_d without any constraints on how these observations are used. The per-symbol mutual information $I(\mathbf{x}; \mathbf{y}_p, \mathbf{y}_d)/T$ is the maximum achievable spectral efficiency and is achieved by a maximum-likelihood decoder based on the true channel description $p(\mathbf{y}_p, \mathbf{y}_d|\mathbf{x})$. Since the expression for this mutual information is intractable, we instead utilize the following lower bound.

Theorem 1 *The ergodic spectral efficiency in bits/s/Hz when τ pilot symbols and $(T - \tau)$ complex Gaussian data symbols are transmitted on every fading block and jointly processed at the receiver satisfies*

$$\frac{1}{T} I(\mathbf{x}; \mathbf{y}_p, \mathbf{y}_d) \geq I_{J_1} \geq I_{J_2} \quad (7)$$

where

$$I_{J_1} = \left(1 - \frac{\tau}{T}\right) C - \frac{\log_2 e}{T} e^{\tau+1/\text{SNR}} \sum_{k=1}^{T-\tau} E_k\left(\tau + \frac{1}{\text{SNR}}\right) \quad (8)$$

and

$$I_{J_2} = \left(1 - \frac{\tau}{T}\right) C - \frac{1}{T} \log_2 \left(\frac{1 + \text{SNR} T}{1 + \text{SNR} \tau} \right). \quad (9)$$

Proof: See Appendix A.

The bound I_{J_1} (or, more precisely, its MIMO form given in Section III) was first derived in [4]. However, it was not given as in (8) but rather left as an expectation over the distribution of \mathbf{x} . As shown in the Appendix, where we provide an alternative derivation, this expectation can be expressed in closed form using the results of [9].

When no pilots are transmitted ($\tau = 0$), I_{J_1} reduces to the bound given for data-only transmission in [10].

E. Optimization of Number of Pilot Symbols

An initial assessment of the optimum number of pilot symbols can be made on the basis of I_{J_2} , whose maximization w.r.t. τ reduces to maximizing the concave function $\log_2(1 + \text{SNR}\tau) - \tau C$. By relaxing τ to a continuous value, the optimum number of pilots is

$$\tau^* = \frac{\log_2 e}{C} - \frac{1}{\text{SNR}} \quad (10)$$

which satisfies $0 \leq \tau^* \leq 1$. This points to τ^* being, when restricted to integers, either 0 or 1. Furthermore, $C < \log_2(1 + \text{SNR})$ (by Jensen's) implying $\tau^* = 1$.

In order to sharpen the above assessment, we turn to the tighter I_{J_1} and consider the low- and high-power regimes separately. In the low-power regime, using

$$C = \log_2(e) (\text{SNR} - \text{SNR}^2) + \mathcal{O}(\text{SNR}^3) \quad (11)$$

and

$$e^{\tau+1/\text{SNR}} E_k\left(\tau + \frac{1}{\text{SNR}}\right) = \text{SNR} - (k + \tau) \text{SNR}^2 + \mathcal{O}(\text{SNR}^3) \quad (12)$$

it is found that maximizing I_{J_1} to second order entails maximizing the concave function $(T - \tau)(T + \tau - 1)$. Thus, the optimum is again either $\tau = 0$ or $\tau = 1$. While both values yield the same I_{J_1} to second order, an exact computation of (8) reveals that $\tau^* = 1$ for $\text{SNR} \rightarrow 0$.

In the high-power regime, using

$$e^{1/\text{SNR}} E_1(1/\text{SNR}) = \log_2 \text{SNR} - \gamma \log_2 e + \mathcal{O}\left(\frac{1}{\text{SNR}}\right) \quad (13)$$

$$e^{1/\text{SNR}} E_k(1/\text{SNR}) = \frac{1}{k-1} + \mathcal{O}\left(\frac{1}{\text{SNR}}\right), \quad k > 1, \quad (14)$$

where $\gamma = 0.5772\dots$ is the Euler-Mascheroni constant, it is found that

$$I_{J_1}|_{\tau=0} = \frac{T-1}{T} C - \frac{\log_2 e}{T} \sum_{k=1}^{T-1} \frac{1}{k} \quad (15)$$

$$I_{J_1}|_{\tau=1} = \frac{T-1}{T} C - \frac{\log_2 e}{T} \sum_{k=1}^{T-1} e \cdot E_k(1). \quad (16)$$

Since $e \cdot E_k(1) < 1/k$ strictly, $\tau = 1$ is preferable over $\tau = 0$ for $\text{SNR} \rightarrow \infty$. (For $\tau \geq 2$, I_{J_1} falls rapidly.)

Altogether, the optimum number of pilots is $\tau^* = 1$ in both the low- and high-power regimes. Setting $\tau = 0$ results in a slight loss (quantified in Section II-G), whereas $\tau \geq 2$ is decidedly suboptimal at moderate/high SNR.

Extrapolating this result to more realistic continuous-fading channels (i.e., the channel varies from symbol-to-symbol according to a random process), we can infer that, with joint processing, it is desirable to have at most roughly one pilot symbol per coherence interval.

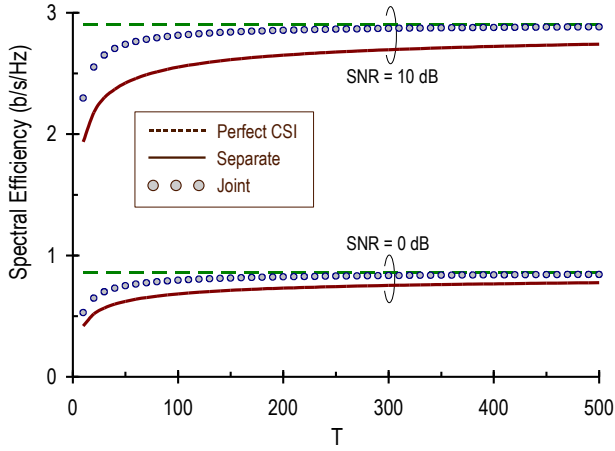


Fig. 1. Spectral Efficiency vs. T for a SISO channel at SNR = 0 dB and SNR = 10 dB. The curves correspond to C , I_S and I_{J_1} (with $\tau = 1$).

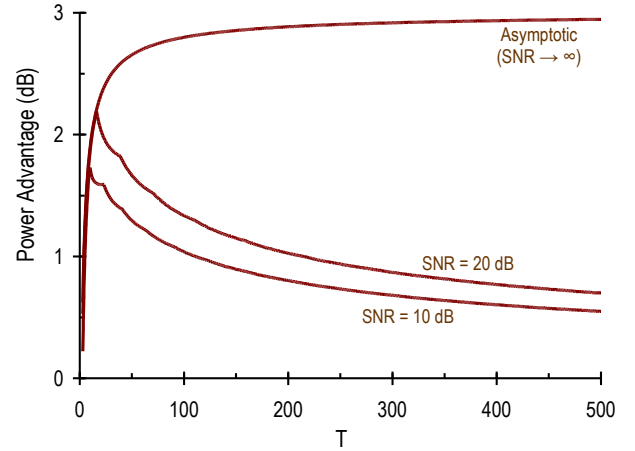


Fig. 2. Power advantage of joint relative to separate processing asymptotically (SNR $\rightarrow \infty$) and at SNR = 10 dB and SNR = 20 dB.

F. Comparison with Separate Processing of Pilots and Data

The value of joint processing is illustrated by examining how the spectral efficiency converges to the perfect-CSI capacity as the blocklength T increases. From (9), the difference between C and I_{J_2} is

$$C - I_{J_2} = \frac{\tau}{T} C + \frac{1}{T} \log_2 \left(\frac{1 + \text{SNR} T}{1 + \text{SNR} \tau} \right) \quad (17)$$

$$= \mathcal{O} \left(\frac{\log_2 T}{T} \right) \quad (18)$$

for any fixed value of τ . On the other hand, the difference between C and the spectral efficiency achievable with separate processing, I_S , vanishes only as $\mathcal{O}(1/\sqrt{T})$ [11]. This contrast is evidenced in Fig. 1.

With joint processing, as T grows the spectral efficiency converges to C even though τ is fixed because the (possibly implicit) channel estimation process can take advantage of the data symbols. On the other hand, if τ were kept fixed the spectral efficiency of the separate approach would not converge to C ; I_S converges to C only because τ is properly increased, as per (5), with T .

G. High-Power Behavior

Further insight is obtained by studying the high-power behavior of the various bounds. At high SNR, and for $\tau = 1$, the lower bounds converge absolutely to

$$I_{J_1} \rightarrow \frac{T-1}{T} \left(C - \frac{e \log_2(e) \sum_{k=1}^{T-1} E_k(1)}{T-1} \right) \quad (19)$$

$$I_{J_2} \rightarrow \frac{T-1}{T} \left(C - \frac{\log_2 T}{T-1} \right), \quad (20)$$

while, with separate processing [3],

$$I_S \rightarrow \frac{T-1}{T} (C - 1). \quad (21)$$

All the above quantities have the same pre-log factor, $(T-1)/T$, and thus the difference between the terms

inside the brackets directly gives the power penalty relative to the perfect-CSI capacity, i.e., the horizontal shift in a plot of spectral efficiency vs. SNR (dB). When the information units are bits, this horizontal shift is in 3-dB units [12].

The asymptotic difference between I_{J_1} and I_{J_2} is

$$\frac{1}{T-1} \left(\log_2 T - e \log_2(e) \sum_{k=1}^{T-1} E_k(1) \right), \quad (22)$$

in 3-dB units. This quantity decreases with T and is minute even for small values of T (e.g., 0.02 dB for $T = 10$) and thus, at high SNR, we can consider the simpler I_{J_2} with only a negligible loss in accuracy.

Based on I_{J_2} then, the asymptotic power advantage of joint processing relative to separate is

$$1 - \frac{\log_2 T}{T-1} \quad (23)$$

in 3-dB units. In Fig. 2, this quantity is plotted versus T , along with the numerically computed advantage at SNR = 10 dB and SNR = 20 dB. (The difference between the respective curves indicates that the convergence of I_S to its asymptote occurs ever more slowly as T grows.)

Using I_{J_2} and (13), it is also straightforward to compute the high-power advantage of transmitting one pilot symbol ($\tau = 1$) rather than none ($\tau = 0$) as

$$\frac{\gamma \log_2 e}{T} \quad (24)$$

in 3-dB units. For short blocks the single pilot is useful, but for larger blocklengths it makes little difference.

Finally, we can also quantify the distance to the true capacity of the block-fading channel. In [13], such capacity (indicated by \mathcal{C} to distinguish it from C , the capacity with perfect CSI) is shown to converge, for SNR $\rightarrow \infty$, to

$$\mathcal{C} \rightarrow \frac{T-1}{T} \left(C - \frac{1}{T-1} \log_2 \left(\frac{e^{T-1} (T-1)!}{T^{T-1}} \right) \right). \quad (25)$$

Using Stirling's approximation,

$$C \approx \frac{T-1}{T} \left(C - \frac{1}{2} \frac{\log_2 T}{T-1} \right) \quad (26)$$

for large SNR, coinciding with the high-SNR expansion of I_{J_2} save for the factor $1/2$. This indicates that the spectral efficiency with joint processing scales with the blocklength T in the same manner as the true capacity in the high-power regime. Furthermore, the power offset between I_{J_2} and the true capacity is only (approximately)

$$\frac{1}{2} \frac{\log_2 T}{T-1} \quad (27)$$

in 3-dB units. This evaluates, for instance, to 0.55 dB and 0.1 dB for $T = 10$ and $T = 100$, respectively.

III. GENERALIZATION TO MIMO

A. Channel Model

With n_T transmit and n_R receive antennas, the SISO input-output relationships in (1) and (2) become

$$\mathbf{Y}_p = \sqrt{\frac{\text{SNR}}{n_T}} \mathbf{H} \mathbf{P} + \mathbf{N}_p \quad (28)$$

$$\mathbf{Y}_d = \sqrt{\frac{\text{SNR}}{n_T}} \mathbf{H} \mathbf{X} + \mathbf{N}_d \quad (29)$$

where \mathbf{H} , \mathbf{P} , \mathbf{X} , \mathbf{N}_p and \mathbf{N}_d are, respectively, $n_R \times n_T$, $n_T \times \tau$, $n_T \times (T - \tau)$, $n_R \times \tau$ and $n_R \times (T - \tau)$. Matrices \mathbf{H} , \mathbf{X} , \mathbf{N}_p and \mathbf{N}_d have IID zero-mean unit-variance complex Gaussian entries while \mathbf{P} must satisfy power constraint $\text{Tr}\{\mathbf{P}\mathbf{P}^\dagger\} \leq n_T \tau$.

B. Perfect CSI

For notational convenience, define $C_{t,r}$ as the function

$$C_{t,r}(\rho) = \mathbb{E} \left[\log_2 \det \left(\mathbf{I} + \frac{\rho}{t} \mathbf{Z} \mathbf{Z}^\dagger \right) \right] \quad (30)$$

where \mathbf{Z} is an $r \times t$ matrix with IID zero-mean unit-variance complex Gaussian entries. The MIMO perfect-CSI capacity with n_T transmit and n_R receive antennas at SNR equals $C_{n_T, n_R}(\text{SNR})$.

C. Separated Processing of Pilots and Data

The SISO expressions for I_S in Section II-C apply verbatim with T , τ , and $C(\cdot)$ replaced, respectively, by T/n_T , $\bar{\tau} = \tau/n_T$, and $C_{n_T, n_R}(\cdot)$.

D. Spectral Efficiency Lower Bounds for Joint Processing

In the MIMO case, we allow for the possibility of either no pilot symbols ($\tau = 0$) or of at least one pilot symbol per antenna ($\tau \geq n_T$).

Theorem 2 *Let $\tau = 0$ or $\tau \geq n_T$. The ergodic spectral efficiency in bits/s/Hz when τ pilot symbols and $(T - \tau)$ complex Gaussian data symbols are transmitted on every fading block and jointly processed at the receiver satisfies*

$$\frac{1}{T} I(\mathbf{X}; \mathbf{Y}_p, \mathbf{Y}_d) \geq I_{J_1} \geq I_{J_2} \quad (31)$$

where

$$I_{J_1} = \left(1 - \frac{\tau}{T} \right) C_{n_T, n_R}(\text{SNR}) - \frac{n_R}{T} C_{n_T, T-\tau} \left(\frac{\text{SNR}}{1 + \frac{\text{SNR}}{n_T} \frac{\tau}{T}} \right) \quad (32)$$

and

$$I_{J_2} = \left(1 - \frac{\tau}{T} \right) C_{n_T, n_R}(\text{SNR}) - \frac{n_T n_R}{T} \log_2 \left(\frac{1 + \text{SNR} \frac{T}{n_T}}{1 + \text{SNR} \frac{\tau}{n_T}} \right) \quad (33)$$

Proof: See Appendix B.

As a by-product of the proof, we show that I_{J_1} is maximized when the pilot matrix \mathbf{P} satisfies

$$\mathbf{P} \mathbf{P}^\dagger = \tau \mathbf{I} \quad (34)$$

which coincides with the optimality condition derived in [3] for the case of separate processing.

Henceforth, we shall focus on the case $n_T = n_R$.

Corollary 1 *If $n_T = n_R = n$, then*

$$\frac{I_{J_2}}{n} = \left(1 - \frac{\tau/n}{T/n} \right) \frac{C_{n,n}(\text{SNR})}{n} - \frac{1}{T/n} \log_2 \left(\frac{1 + \text{SNR} T/n}{1 + \text{SNR} \tau/n} \right) \quad (35)$$

which coincides with its SISO counterpart in (9) only with an effective fading blocklength of T/n , an effective number of pilot symbols of τ/n , and C replaced by $C_{n,n}/n$.

E. Optimization of Number of Pilot Symbols

In the low-power regime, the number of pilot symbols can be optimized on the basis of I_{J_1} . Using

$$C_{t,r}(\rho) = r \log_2(e) \left(\rho - \frac{t+r}{2t} \rho^2 \right) + \mathcal{O}(\rho^3) \quad (36)$$

it is found that maximizing I_{J_1} to second order requires maximizing the concave function $(T - \tau)(T + \tau - n_R)$. This implies that either $\tau = 0$ or $\tau = n$ is optimal, and the two are indistinguishable to second order.

Drawing parallels with its SISO counterpart, the maximization of I_{J_2} w.r.t. to τ is equivalent to the maximization of $\log_2(1 + \text{SNR} \bar{\tau}) - \bar{\tau} C_{n,n}/n$ w.r.t. $\bar{\tau} = \tau/n$. Hence,

$$\bar{\tau}^* = \frac{\log_2 e}{C_{n,n}/n} - \frac{1}{\text{SNR}} \quad (37)$$

if $\bar{\tau}$ is relaxed to continuous values. This quantity is below unity whenever $C_{n,n}/n \geq \log_2 e$, which implies that the optimum number of pilots is either 0 or n . Since $C_{n,n}/n \leq \log_2(1 + \text{SNR})$, $\tau = n$ is preferred over $\tau = 0$.

F. High-Power Behavior

Because I_{J_2} and I_S mirror their SISO counterparts, the asymptotic power advantage (in 3-dB units) of joint relative to separate processing for MIMO is the SISO advantage for an effective blocklength of T/n , i.e.,

$$1 - \frac{\log_2(T/n)}{T/n - 1} \quad (38)$$

APPENDIX A

By the chain rule, the mutual information with perfect receiver knowledge of H expands as $I(\mathbf{x}; \mathbf{y}_p, \mathbf{y}_d, H) = I(\mathbf{x}; \mathbf{y}_p, \mathbf{y}_d) + I(\mathbf{x}; H | \mathbf{y}_p, \mathbf{y}_d)$. Thus,

$$I(\mathbf{x}; \mathbf{y}_p, \mathbf{y}_d) = I(\mathbf{x}; \mathbf{y}_p, \mathbf{y}_d, H) - I(\mathbf{x}; H | \mathbf{y}_p, \mathbf{y}_d) \quad (39)$$

$$= I(\mathbf{x}; \mathbf{y}_p, \mathbf{y}_d, H) - h(H | \mathbf{y}_p, \mathbf{y}_d) + h(H | \mathbf{y}_p, \mathbf{y}_d, \mathbf{x}) \quad (40)$$

$$\geq I(\mathbf{x}; \mathbf{y}_p, \mathbf{y}_d, H) - h(H | \mathbf{y}_p) + h(H | \mathbf{y}_p, \mathbf{y}_d, \mathbf{x}) \quad (41)$$

where $h(\cdot)$ denotes differential entropy and (41) holds because conditioning reduces entropy.

The signal-to-noise ratio when estimating H on the basis of \mathbf{y}_p is $\text{SNR} \tau$. Thus, $H | \mathbf{y}_p$ is conditionally Gaussian with variance $1/(1 + \text{SNR} \tau)$ and therefore

$$h(H | \mathbf{y}_p) = \log_2(\pi e) - \log_2(1 + \text{SNR} \tau). \quad (42)$$

In turn, the signal-to-noise ratio when estimating H on the basis of $(\mathbf{y}_p, \mathbf{y}_d)$, conditioned on \mathbf{x}_d , is $\text{SNR} \tau + \text{SNR} \sum_{k=1}^{T-\tau} |x_k|^2$ and thus

$$h(H | \mathbf{y}_p, \mathbf{y}_d, \mathbf{x}) = -\mathbb{E} \left[\log_2 \left(1 + \text{SNR} \tau + \text{SNR} \sum_{k=1}^{T-\tau} |x_k|^2 \right) \right] + \log_2(\pi e). \quad (43)$$

Using $I(\mathbf{x}; \mathbf{y}_p, \mathbf{y}_d, H) = (T - \tau) C$, plugging (42) and (43) into (41), and scaling all the terms by $1/T$,

$$I_{J_1} = \left(1 - \frac{\tau}{T}\right) C - \frac{1}{T} \mathbb{E} \left[\log_2 \left(1 + \frac{\text{SNR} \sum_{k=1}^{T-\tau} |x_k|^2}{1 + \text{SNR} \tau} \right) \right]. \quad (44)$$

A closed form for the expectation in (44) is given in [9], leading directly to (8).

The subsequent lower bound, I_{J_2} , follows from application of Jensen's inequality to (44). Since $\mathbb{E}[|x_k|^2] = 1$,

$$\mathbb{E} \left[\log_2 \left(1 + \frac{\text{SNR} \sum_{k=1}^{T-\tau} |x_k|^2}{1 + \text{SNR} \tau} \right) \right] \leq \log_2 \left(1 + \frac{\text{SNR} (T - \tau)}{1 + \text{SNR} \tau} \right) \quad (45)$$

APPENDIX B

Starting at (41), we need only compute $h(\mathbf{H} | \mathbf{y}_p)$ and $h(\mathbf{H} | \mathbf{y}_p, \mathbf{y}_d, \mathbf{x})$. Because the n_R antennas are decoupled when conditioned on either \mathbf{y}_p or $(\mathbf{y}_p, \mathbf{y}_d, \mathbf{x})$, these terms can be evaluated separately for each receive antenna. From [3], the covariances of one row of \mathbf{H} conditioned on \mathbf{y}_p and on $(\mathbf{y}_p, \mathbf{y}_d, \mathbf{x})$, respectively, are

$$\mathbf{K}_{\mathbf{H} | \mathbf{y}_p} = \left(\mathbf{I} + \frac{\text{SNR}}{n_T} \mathbf{P} \mathbf{P}^\dagger \right)^{-1} \quad (46)$$

$$\mathbf{K}_{\mathbf{H} | \mathbf{y}_p, \mathbf{y}_d, \mathbf{x}} = \left(\mathbf{I} + \frac{\text{SNR}}{n_T} (\mathbf{P} \mathbf{P}^\dagger + \mathbf{X} \mathbf{X}^\dagger) \right)^{-1}. \quad (47)$$

Defining $\Delta = h(\mathbf{H} | \mathbf{y}_p) - h(\mathbf{H} | \mathbf{y}_p, \mathbf{y}_d, \mathbf{x})$, we have

$$\Delta = n_R \mathbb{E} \left[\log \det \mathbf{K}_{\mathbf{H} | \mathbf{y}_p, \mathbf{y}_d, \mathbf{x}} \right] - n_R \log \det \mathbf{K}_{\mathbf{H} | \mathbf{y}_p} \quad (48)$$

$$= n_R \mathbb{E} \left[\log \det \left(\mathbf{I} + \left(\mathbf{I} + \frac{\text{SNR}}{n_T} \mathbf{P} \mathbf{P}^\dagger \right)^{-1} \frac{\text{SNR}}{n_T} \mathbf{X} \mathbf{X}^\dagger \right) \right] \quad (49)$$

To obtain I_{J_1} we must find the pilot sequence \mathbf{P} that minimizes (49). This amounts to choosing the worst-case noise covariance when the input and the channel are both spatially white. Since the distribution of \mathbf{X} is rotationally invariant, we need only consider diagonal forms for $\mathbf{P} \mathbf{P}^\dagger$. To show that the best choice is $\mathbf{P} \mathbf{P}^\dagger = \tau \mathbf{I}$, we apply the argument in [14, Sec. 4.1] to the function in (49), which is convex w.r.t. $\mathbf{P} \mathbf{P}^\dagger$. With $\mathbf{P} \mathbf{P}^\dagger = \tau \mathbf{I}$,

$$\Delta = n_R \mathbb{E} \left[\log \det \left(\mathbf{I} + \frac{\frac{\text{SNR}}{n_T} \mathbf{X} \mathbf{X}^\dagger}{1 + \frac{\text{SNR} \tau}{n_T}} \right) \right] \quad (50)$$

$$= n_R C_{n_T, T-\tau} \left(\frac{\text{SNR}}{1 + \text{SNR} \frac{\tau}{n_T}} \right). \quad (51)$$

I_{J_2} is reached by applying Jensen's inequality to (50).

REFERENCES

- [1] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inform. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.
- [2] L. Zheng and D. N. C. Tse, "Communication on the Grassman manifold: A geometric approach to the non-coherent multiple-antenna channel," *IEEE Trans. Inform. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.
- [3] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
- [4] S. Furrer and D. Dahlhaus, "Multiple-antenna signaling over fading channels with estimated channel state information: Capacity analysis," *IEEE Trans. Inform. Theory*, vol. 53, no. 6, pp. 2028–2043, Jun. 2007.
- [5] A. Lapidoth and S. Shamai, "Fading channels: How perfect need 'perfect side information' be?" *IEEE Trans. Inform. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.
- [6] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions: general model, design criteria, and signal processing," *IEEE Signal Proc. Magazine*, vol. 21, no. 6, pp. 12–25, Nov. 2004.
- [7] W. Zhang and J. N. Laneman, "How good is phase-shift keying for peak-limited Rayleigh fading channels in the low-SNR regime?" *IEEE Trans. Inform. Theory*, vol. 53, no. 1, pp. 236–251, Jan. 2007.
- [8] T. Li and O. Collins, "A successive decoding strategy for channels with memory," *Proc. of ISIT*, Sep. 2005.
- [9] H. Shin and J. H. Lee, "Capacity of multiple-antenna fading channels: Spatial fading correlation, double scattering and keyhole," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2636–2647, Oct. 2003.
- [10] M. Godavarti, T. L. Marzetta, and S. Shamai, "Capacity of a mobile multiple-antenna wireless link with isotropically random Rician fading," *IEEE Trans. Inform. Theory*, vol. 49, no. 12, pp. 3330–3334, Dec. 2003.
- [11] N. Jindal and A. Lozano, "Optimum pilot overhead in wireless communication: A unified treatment of continuous and block-fading channels," *Submitted to IEEE Trans. Wireless Comm.*, 2009.
- [12] A. Lozano, A. M. Tulino, and S. Verdú, "High-SNR power offset in multiantenna communication," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4134–4151, Dec. 2005.
- [13] B. H. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 46, Mar. 2000.
- [14] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecom*, vol. 10, pp. 585–595, Nov. 1999.