

Capacity-Achieving Input Covariance for Single-user Multi-Antenna Channels

Antonia M. Tulino*, Angel Lozano[†] and Sergio Verdú[‡]

January 13, 2005

Abstract

We characterize the capacity-achieving input covariance for multi-antenna channels known instantaneously at the receiver and in distribution at the transmitter. Our characterization, valid for arbitrary numbers of antennas, encompasses both the eigenvectors and the eigenvalues. The eigenvectors are found for zero-mean channels with arbitrary fading profiles and a wide range of correlation and keyhole structures. For the eigenvalues, in turn, we present necessary and sufficient conditions as well as an iterative algorithm that exhibits remarkable properties: universal applicability, robustness and rapid convergence. In addition, we identify channel structures for which an isotropic input achieves capacity.

Keywords:

Channel capacity; Multi-antenna arrays; Input optimization; Fading channels; Antenna correlation; Ricean channels; Keyhole channels; Channel-state Information;

*Antonia M. Tulino is with Universita Degli Studi di Napoli, Federico II, 80125 Napoli, Italy

[†]Angel Lozano is with Bell Laboratories (Lucent Technologies), Holmdel, NJ07733, USA.

[‡]Sergio Verdú is with Princeton University, Princeton, NJ08540, USA.

I Introduction

While, in most instances of wireless communication, the receiver can accurately track the instantaneous state of the fading channel, the transmitter is often unable to perform such tracking. This is prominently true in wide-area mobile systems, where the dominant form of duplexing relies on frequency separation of uplink and downlink that renders their fading nonreciprocal. On account of this lack of reciprocity, the provision of CSI (channel state information) to the transmitter hinges on the use of feedback, which consumes resources and, more fundamentally, may incur round-trip delays nonnegligible with respect to the coherence time of the CSI being reported.

Statistical information about the channel, on the other hand, is virtually always accessible to the transmitter since the periods over which the fading process is basically stationary are several orders of magnitude larger than the duration of the fades. Moreover, the uplink and downlink statistics *are* usually reciprocal and thus statistical feedback is not only affordable, but possibly dispensable.

Altogether, the most typical operating regime in mobile systems is that in which (i) the receiver has instantaneous CSI,¹ and (ii) the transmitter has only access to its distribution. In such regime, which constitutes the focus of this paper, the input cannot be tailored to the state of the channel, but only to its distribution.

In multi-antenna channels impaired by additive Gaussian noise and with perfect CSI at the receiver, the unique capacity-achieving input is zero-mean Gaussian and thus its characterization boils down to the determination of its spatial covariance.

Unlike in the case that the CSI can be instantaneously accessed also by the transmitter, for which the capacity-achieving input covariance is well known [5, 6], for our regime of interest the structure of the capacity-achieving input covariance is only known for certain classes of channels. The earliest statement dates back to [6], where it was shown that, for channel matrices with zero-mean IID (independent identically distributed) Gaussian entries, the capacity-achieving input is isotropic. Posterior findings expanded this initial

¹This holds for signal-to-noise ratios up to some (usually very high) level, beyond which the noise becomes commensurate with the uncertainty—unavoidable in fading conditions—in the knowledge of the channel state and fully coherent reception becomes fundamentally unfeasible [1]–[4].

result in several directions:

- For correlated multi-transmit single-receive channels with zero-mean Gaussian entries, the eigenvectors of the capacity-achieving input covariance were established in [7]. This result was then extended to multiple receive antennas in [8, 9, 10]. In every case, the eigenvalues were left to numerical optimization.²
- For channels with IID Gaussian entries with arbitrary mean, the eigenvectors of the input covariance were characterized in [7, 13, 14]. (Again, the result was first proved for multi-transmit single-receive channels [7] and then extended to multiple receivers [13, 14].) Some relationships between the eigenvalues were also uncovered, but the need for a numerical optimization was not eliminated. A search procedure based on the Blahut-Arimoto algorithm has recently been given in [15].
- For the region of low signal-to-noise ratio (SNR), a complete characterization of the limiting covariance, valid for arbitrary channels, was provided in [16].

This paper unifies these various results on the eigenvectors of the input covariance and it also addresses the characterization of the corresponding eigenvalues. The specific contributions are:

- We identify the eigenvectors of the input covariance for a range of channels with arbitrary numbers of antennas. The solutions in [6]–[14] are seen to be instances of this result.
- We present necessary and sufficient conditions that implicitly characterize the eigenvalues of the input covariance. In a number of limiting scenarios, particularly at low- and high-SNR, these conditions lead to explicit solutions.
- For arbitrary SNR, an iterative algorithm is unveiled to efficiently solve for the eigenvalues. In contrast with standard numerical procedures, this algorithm has strong operational significance.

²Asymptotically in the number of antennas, this optimization is approached analytically in [11]. For the multi-transmit single-receive case, implicit necessary and sufficient conditions for the eigenvalues are given in [12].

- Finally, we determine channel structures for which the result in [6] holds, i.e., for which an isotropic input achieves capacity.

The manuscript is organized as follows. Section II introduces the key quantities and models. Section III distills the engineering significance of the eigenvectors and eigenvalues of the input covariance and reviews the various operational regimes in which their characterization is of interest. For the regime central to this paper, this characterization is addressed in Section IV. In turn, Section V identifies some channel structures on which an isotropic input achieves capacity. Sections VI and VII, finally, provide a sequence of examples and concluding remarks.

For the interested reader, results on the optimization of the input covariance under various criteria other than the capacity, a process sometimes referred to as *precoding*, can be found in [17]–[23].

II Definitions and Models

Given n_T transmit and n_R receive antennas and frequency-flat fading,³ the baseband complex model we consider is

$$\mathbf{y} = \sqrt{g} \mathbf{H} \mathbf{x} + \mathbf{n}$$

where \mathbf{x} and \mathbf{y} are the input and output vectors while \mathbf{n} is white Gaussian noise. The channel is represented by the $(n_R \times n_T)$ random matrix $\sqrt{g} \mathbf{H}$ such that

$$E[\text{Tr}\{\mathbf{H}\mathbf{H}^\dagger\}] = n_R n_T.$$

The j -th column of \mathbf{H} is noted as \mathbf{h}_j . Furthermore, to facilitate the handling of non-zero-mean channels we define $\bar{\mathbf{H}} \triangleq E[\mathbf{H}]$.

The covariance of the input, conveniently normalized, is denoted by

$$\Phi \triangleq \frac{E[\mathbf{x}\mathbf{x}^\dagger]}{\frac{1}{n_T} E[\|\mathbf{x}\|^2]}$$

³If the fading process is frequency selective, the channel can be decomposed into a number of parallel noninteracting subchannels, each experiencing approximately frequency-flat fading and having the same ergodic capacity as the aggregate channel.

where the normalization ensures that $\text{Tr}\{\Phi\}=n_T$. The input is isotropic when $\Phi=\mathbf{I}$. The ergodic capacity (in bits/s/Hz) is

$$C = \max_{\Phi:\text{Tr}\{\Phi\}=n_T} E \left[\log_2 \det \left(\mathbf{I} + \frac{\text{SNR}}{n_T} \mathbf{H}\Phi\mathbf{H}^\dagger \right) \right] \quad (1)$$

with

$$\text{SNR} \triangleq g \frac{E[\|\mathbf{x}\|^2]}{\frac{1}{n_R} E[\|\mathbf{n}\|^2]}$$

which corresponds to the average signal-to-noise ratio per receive antenna when either the input is isotropic or the channel entries are zero-mean IID.⁴

Indicating by $(\cdot)_{i,j}$ the (i,j) -th entry of a matrix, the correlation between the (i,j) -th and (i',j') -th entries of \mathbf{H} is

$$R_{\mathbf{H}}(i, j; i', j') \triangleq E [(\mathbf{H} - \bar{\mathbf{H}})_{i,j} (\mathbf{H} - \bar{\mathbf{H}})_{i',j'}^*].$$

Much of the multi-antenna literature, however, deals only with *separable* (also termed *kro-necker* or *product*) correlations [24], constrained as follows:

Definition 1 *The correlation of \mathbf{H} is said to be separable if*

$$R_{\mathbf{H}}(i, j; i', j') = (\Theta_{\mathbf{R}})_{i,i'} (\Theta_{\mathbf{T}})_{j,j'}$$

where $\Theta_{\mathbf{R}}$ and $\Theta_{\mathbf{T}}$ are $(n_{\mathbf{R}} \times n_{\mathbf{R}})$ and $(n_{\mathbf{T}} \times n_{\mathbf{T}})$ correlation matrices whose entries indicate the correlation between receive antennas and between transmit antennas, respectively.

While simple and analytically friendly, the separable correlation model has clear limitations. It usually suffices to represent the correlation that arises with spatial diversity, due to antenna proximity, but it cannot accommodate other diversity mechanisms such as those that rely on polarization or radiation pattern differences [25, 26].

In this paper, we consider more general channel structures. Specifically, our analysis embraces the following representation:

Definition 2 *The channel model used primarily throughout the paper is*

$$\mathbf{H} = \bar{\mathbf{H}} + \mathbf{U}_{\mathbf{R}} \tilde{\mathbf{H}} \mathbf{U}_{\mathbf{T}}^\dagger \quad (2)$$

⁴In general, the receive signal-to-noise is $\frac{E[\|\sqrt{g}\mathbf{H}\mathbf{x}\|^2]}{E[\|\mathbf{n}\|^2]} = \frac{\text{Tr}\{E[\mathbf{H}^\dagger\mathbf{H}]\Phi\}}{n_{\mathbf{R}}n_{\mathbf{T}}}$ SNR, which depends on Φ .

where $\bar{\mathbf{H}}$ is a $(n_{\text{R}} \times n_{\text{T}})$ deterministic matrix, \mathbf{U}_{R} and \mathbf{U}_{T} are $(n_{\text{R}} \times n_{\text{R}})$ and $(n_{\text{T}} \times n_{\text{T}})$ deterministic unitary matrices and $\tilde{\mathbf{H}}$ is a $(n_{\text{R}} \times n_{\text{T}})$ random matrix with independent columns the distribution of whose entries is jointly symmetric with respect to zero, i.e., zero-mean and arbitrary in amplitude. The variances of the entries of $\tilde{\mathbf{H}}$ can be assembled into another $(n_{\text{R}} \times n_{\text{T}})$ matrix, $\tilde{\mathbf{G}}$, such that

$$(\tilde{\mathbf{G}})_{i,j} = E[|(\tilde{\mathbf{H}})_{i,j}|^2]. \quad (3)$$

While not universal, this representation encompasses most channels of interest. If the channel is zero-mean ($\bar{\mathbf{H}}=0$) and the entries of $\tilde{\mathbf{H}}$ are independent and Gaussian, then it particularizes to the model proposed in [27] and [28] and experimentally validated in the latter. Under these conditions, (2) is the Karhunen-Loève expansion of \mathbf{H} , i.e., the columns of \mathbf{U}_{R} and \mathbf{U}_{T} contain the eigenfunctions of $R_{\mathbf{H}}$ while the variances of the entries of $\tilde{\mathbf{H}}$ are [29]

$$(\tilde{\mathbf{G}})_{i,j} = \lambda_{i,j}(R_{\mathbf{H}}).$$

where $\lambda_{i,j}(\cdot)$ denotes the (i, j) -th eigenvalue.⁵ Moreover, the columns of \mathbf{U}_{R} and \mathbf{U}_{T} also correspond respectively with the eigenvectors of $E[\mathbf{H}\mathbf{H}^\dagger]$ and $E[\mathbf{H}^\dagger\mathbf{H}]$. Relevant special cases of (2) with $\bar{\mathbf{H}}=0$ and $\tilde{\mathbf{H}}$ having independent Gaussian entries include:

- If \mathbf{U}_{R} and \mathbf{U}_{T} are constrained to be Fourier matrices, then (2) yields the *virtual representation* proposed in [30].⁶ In this case, the columns of \mathbf{U}_{R} and \mathbf{U}_{T} can be interpreted as steering vectors launching and receiving energy on specific spatial directions.
- Separable correlations (cf. Definition 1) are modelled if $(\tilde{\mathbf{G}})_{i,j} = \lambda_i(\Theta_{\text{R}})\lambda_j(\Theta_{\text{T}})$ while \mathbf{U}_{R} and \mathbf{U}_{T} correspond, respectively, with the eigenvectors of Θ_{R} and Θ_{T} .

For $\bar{\mathbf{H}} \neq 0$, special cases worth highlighting are:

⁵The eigenfunctions of $R_{\mathbf{H}}$, denoted $u_k(i) v_\ell(j)$, constitute a set of complete orthonormal discrete basis functions satisfying

$$\sum_{i'} \sum_{j'} R_{\mathbf{H}}(i, j; i', j') u_k(i) v_\ell(j) = \lambda_{k,\ell}(R_{\mathbf{H}}) u_k(i) v_\ell(j)$$

which are mapped onto \mathbf{U}_{R} and \mathbf{U}_{T} via $(\mathbf{U}_{\text{R}})_{i,k} = u_k(i)$ and $(\mathbf{U}_{\text{T}})_{\ell,j} = v_\ell^*(j)$.

⁶In this representation, the independence of the entries of $\tilde{\mathbf{H}}$ is only approximate except for $n_{\text{T}}, n_{\text{R}} \rightarrow \infty$. When \mathbf{U}_{R} and \mathbf{U}_{T} are obtained through the Karhunen-Loève expansion, in contrast, the independence is always exact for any number of antennas.

- If the entries of $\tilde{\mathbf{H}}$ are IID Gaussian, \mathbf{U}_R and \mathbf{U}_T become immaterial because of the unitary invariance of the central IID Gaussian law and we obtain a standard Ricean channel with arbitrary mean $\bar{\mathbf{H}}$. The Ricean factor is given by

$$\frac{E[\|\tilde{\mathbf{H}}\|_F^2]}{\|\bar{\mathbf{H}}\|_F^2} = \frac{n_R n_T \sigma^2}{\text{Tr}\{\bar{\mathbf{H}}\bar{\mathbf{H}}^\dagger\}}$$

where $\|\cdot\|_F$ denotes the Frobenius norm and σ^2 is the variance of the entries of $\tilde{\mathbf{H}}$.

- If \mathbf{U}_R and \mathbf{U}_T are identities and the entries of $\tilde{\mathbf{H}}$ are independent, then (2) can model the use of antennas with different polarizations and/or radiation patterns, whereby the channel coefficients exhibit little correlation but possibly large disparities in their strength and distribution. In this case, $\bar{\mathbf{H}}$ and $\tilde{\mathbf{G}}$ can accommodate the crosspolar and pattern discrimination factors that determine the relative strength of the signals transmitted and received via antennas with different characteristics.

A notable exception that falls outside the representation in (2) is that of the *keyhole* (or *pinhole*) channel [31, 32]:

Definition 3 *A keyhole channel is modelled as $\mathbf{H}=\mathbf{c}_R\mathbf{c}_T^\dagger$ where \mathbf{c}_R and \mathbf{c}_T are column vectors each having independent random entries whose distribution is symmetric with respect to zero, i.e., zero-mean and arbitrary in amplitude.*

We shall also analyze this channel, a generalization of the original keyhole channel where the entries of \mathbf{c}_R and \mathbf{c}_T were zero-mean IID Gaussian [31, 32].

III Input Covariance: Eigenvectors and Eigenvalues

Let us decompose the normalized input covariance as $\Phi=\mathbf{V}\mathbf{P}\mathbf{V}^\dagger$ identifying the eigenvectors of Φ with the columns of the unitary matrix \mathbf{V} and its eigenvalues with the diagonal entries of $\mathbf{P}=\text{diag}\{p_1, p_2, \dots, p_{n_T}\}$. Both the eigenvectors and the eigenvalues have immediate engineering meaning: the former indicate the *directions* (in vector space) on which signalling takes place while the latter signify the transmit *powers* allocated onto each such

eigenvector.⁷ This is illustrated in Fig. 1, which depicts a transmit architecture that generates a signal with arbitrary covariance. For each use of the channel, a space-time encoder outputs a set of independent unit-variance Gaussian symbols, s_1, \dots, s_{n_T} , each of which is assigned a certain transmit power (which may be zero) and rotated into a certain direction before being simultaneously radiated out of the n_T transmit antennas.

Before embarking on the study of Φ for our regime of interest, it is worth reviewing its capacity-achieving forms for other relevant regimes, always under the premise of knowledge of \mathbf{H} at the receiver:

- If the transmitter has also access to \mathbf{H} instantaneously, then the columns of \mathbf{V} must equal the eigenvectors of $\mathbf{H}^\dagger \mathbf{H}$ whereas the transmit powers in \mathbf{P} can be found via waterfill on the eigenvalues of $\mathbf{H}^\dagger \mathbf{H}$ [5, 6]. This would be the regime of interest for channels that either vary slowly over time and/or where uplink and downlink are reciprocal.
- If the transmitter has access to neither \mathbf{H} nor its distribution, then the best strategy in a max-min sense is usually $\Phi = \mathbf{I}$ [33, 34].

In the remainder we focus exclusively in the regime where the transmitter has access to the distribution of \mathbf{H} , but not to \mathbf{H} itself. Clearly, the corresponding capacity is upper- and lower-bounded, respectively, by those of the above limiting regimes. As we shall see, the solutions for Φ also exhibit interesting relationships as well as some marked differences with the ones above.

IV Characterization of Φ

A Eigenvectors

We start our analysis of Φ by focusing on its eigenvectors. For zero-mean fading channels, we present the following result.

⁷More precisely, given the normalization imposed on Φ , each p_j indicates the fraction of the total power per transmit antenna allocated to each eigenvector.

Theorem 1 Consider the channel $\mathbf{H}=\mathbf{U}_R\tilde{\mathbf{H}}\mathbf{U}_T^\dagger$ with \mathbf{U}_R , \mathbf{U}_T and $\tilde{\mathbf{H}}$ as per Definition 2. The eigenvectors of the capacity-achieving input covariance are given by the columns of \mathbf{U}_T .

Proof: See Appendix A.

It is worth specializing this result to the most common zero-mean channels:

- In the case of Rayleigh-faded channels with separable correlations, the columns of \mathbf{U}_T revert to the eigenvectors of the transmit correlation matrix, Θ_T , as claimed in [8, 9, 10].⁸
- If \mathbf{U}_R and \mathbf{U}_T are identities, then Φ is diagonal and its (j, j) -th entry indicates the normalized power allocated to the j -th transmit antenna while $(\tilde{\mathbf{G}}\Phi)_{i,j}$, with $\tilde{\mathbf{G}}$ as defined in (3), represents the normalized power received by the i -th receive antenna from the j -th transmit antenna.
- If \mathbf{U}_R and \mathbf{U}_T are Fourier matrices and the entries of $\tilde{\mathbf{H}}$ are independent and zero-mean Gaussian, Theorem 1 mirrors the result independently derived in [36] for the *virtual* representation [30].

For nonzero-mean channels, in turn, we recall the result presented in [13, 14].

Theorem 2 [13, 14] Consider the channel

$$\mathbf{H} = \bar{\mathbf{H}} + \mathbf{U}_R\tilde{\mathbf{H}}\mathbf{U}_T^\dagger \quad (4)$$

where $\bar{\mathbf{H}}$, \mathbf{U}_R and \mathbf{U}_T are as per Definition 2 and the entries of $\tilde{\mathbf{H}}$ are zero-mean IID Gaussian.⁹ The eigenvectors of the capacity-achieving input covariance equal those of $\bar{\mathbf{H}}^\dagger\bar{\mathbf{H}}$.

If \mathbf{U}_R and \mathbf{U}_T coincide respectively with the left and right singular vectors of $\bar{\mathbf{H}}$, then the IID Gaussian condition on $\tilde{\mathbf{H}}$ can be relaxed to that of independent entries with symmetric distribution with respect to zero.

Next, we turn our attention to the keyhole channel.

⁸This same set of eigenvectors is shown to maximize the outage capacity [35] and to minimize the pairwise error probability [17].

⁹Although, in this case, \mathbf{U}_R and \mathbf{U}_T are immaterial because the distribution of $\tilde{\mathbf{H}}$ is unitarily invariant, they are retained to render (4) cohesive with the primary model.

Theorem 3 Consider the channel

$$\mathbf{H} = \mathbf{c}_R \mathbf{c}_T^\dagger$$

where \mathbf{c}_R and \mathbf{c}_T are as per Definition 3. The capacity-achieving input covariance is diagonal.

Proof: See Appendix A.

It can be inferred from their respective proofs that the scope of Theorems 1–3 is somewhat broader than their claims in the following sense: the corresponding capacity-achieving eigenvectors ensure the highest achievable rate regardless of the choice of eigenvalues. This is not necessarily true for other channels.

In the low-SNR asymptote, it is possible to transcend the channel models used in Theorems 1–3 and find universal expressions for the eigenvectors of Φ . Given an arbitrary channel \mathbf{H} , for $\text{SNR} \rightarrow 0$ signalling should take place exclusively on the maximal-eigenvalue eigenspace of $E[\mathbf{H}^\dagger \mathbf{H}]$ for first- and second-order optimality [16, Theorem 12].

Relating the solutions conveyed by Theorems 1–3 it becomes clear that, in every case, the eigenvectors of Φ coincide with those of $E[\mathbf{H}^\dagger \mathbf{H}]$. Furthermore, as has just been pointed out, this set of eigenvectors always enables the appropriate covariance at low SNR. We therefore see that, for a wide range of channels with symmetrically distributed entries, the eigenvectors of its capacity-achieving input covariance equal those of $E[\mathbf{H}^\dagger \mathbf{H}]$.¹⁰ This covariance structure is particularly gratifying when contrasted with its counterpart for the regime where \mathbf{H} is known by the transmitter, in which case Φ must diagonalize $\mathbf{H}^\dagger \mathbf{H}$. For those channels to which this structure applies, when only the distribution is known then Φ should *on average* diagonalize $\mathbf{H}^\dagger \mathbf{H}$.

B Eigenvalues

Having analyzed the eigenvectors of Φ , we now shift our focus towards its eigenvalues, i.e., the transmit powers on each of those eigenvectors. Recall the decomposition $\Phi = \mathbf{V} \mathbf{P} \mathbf{V}^\dagger$ where \mathbf{V} contains the capacity-achieving eigenvectors while $\mathbf{P} = \text{diag}\{p_1, p_2, \dots, p_{n_T}\}$, normalized such that $\text{Tr}\{\mathbf{P}\} = n_T$, holds the corresponding transmit powers. The characteriza-

¹⁰This solution is nonetheless not universal: counter-examples can be found if, for instance, the distribution of the entries is not symmetric [37].

tion of \mathbf{P} provided in this section is valid for any arbitrary channel distribution for which \mathbf{V} is known.

Define a rotated version of the channel, $\hat{\mathbf{H}}=\mathbf{H}\mathbf{V}$, whose j -th column is denoted by $\hat{\mathbf{h}}_j$. As shown in Appendix B, the capacity-achieving \mathbf{P} must satisfy the following set of necessary and sufficient conditions:

$$\frac{1}{n_R}E \left[\text{Tr} \left\{ \left(\mathbf{I} + \text{SNR} \hat{\mathbf{h}}_j \hat{\mathbf{h}}_j^\dagger \right) \left(\mathbf{I} + \frac{\text{SNR}}{n_T} \hat{\mathbf{H}} \mathbf{P} \hat{\mathbf{H}}^\dagger \right)^{-1} \right\} \right] = 1 \quad \text{if } p_j > 0$$

$$\leq 1 \quad \text{if } p_j = 0 \quad (5)$$

A unique set of transmit powers exists that satisfies (5). Since the corresponding parallel channels are not orthogonal,¹¹ this solution does not correspond to a waterfill on any statistical measure of the channel. The following observations can be made:

- The inequality in (5) indicates that, below a certain SNR, some of the transmit powers may be zero.
- For $\text{SNR} \rightarrow 0$, in fact, first-order optimality requires that transmit power be allocated only to the diagonal entry of \mathbf{P} that corresponds with the maximal diagonal value of $E[\hat{\mathbf{H}}^\dagger \hat{\mathbf{H}}]$. If the multiplicity of such maximal value is plural, equal-power allocation to the corresponding entries of \mathbf{P} ensures also second-order optimality [16, Theorem 12].
- For $\text{SNR} \rightarrow \infty$, if $\hat{\mathbf{H}}^\dagger \hat{\mathbf{H}}$ is nonsingular with probability 1 then zero-order optimality [38] requires that the power allocation be uniform (see Appendix B). In many channels of interest, such as Ricean and Rayleigh-faded with nonsingular correlations, this is the case whenever $n_T \leq n_R$. If $n_T > n_R$ or in other classes of channels, the power allocation at high SNR need not be uniform [15, 38].

Beyond these asymptotes, the transmit powers satisfying (5) cannot, in general, be found explicitly. In the remainder of this section we derive—directly from (5)—an iterative power allocation algorithm.

In order to formulate this algorithm, it is useful to introduce the MMSE (minimum mean-square error) on the linear estimation of the signal, s_j , transmitted along the j -th signalling

¹¹This is in contrast with the regime where \mathbf{H} is known by the transmitter, in which case parallel orthogonal channels can be created and the power allocation does reduce to a waterfill [6].

eigenvector. Defining

$$\mathbf{B}_j \triangleq \left(\mathbf{I} + \frac{\text{SNR}}{n_T} \hat{\mathbf{H}}_j \mathbf{P}_j \hat{\mathbf{H}}_j^\dagger \right)^{-1} \quad (6)$$

where $\hat{\mathbf{H}}_j$ indicates the matrix obtained by removing from $\hat{\mathbf{H}}$ the j -th column whereas \mathbf{P}_j indicates the diagonal matrix obtained by removing from \mathbf{P} the j -th row and j -th column, such MMSE is [39]

$$\text{MMSE}_j = \frac{1}{1 + p_j \frac{\text{SNR}}{n_T} \hat{\mathbf{h}}_j^\dagger \mathbf{B}_j \hat{\mathbf{h}}_j} \quad (7)$$

which, because of the unit variance of s_j , is restricted to the interval $[0, 1]$. The useful signal recovered along the j -th eigenvector is given by $1 - \text{MMSE}_j$ and thus the corresponding SINR (signal-to-interference-and-noise ratio) equals $\frac{1}{\text{MMSE}_j} - 1$. The expectation of (7) with respect to $\hat{\mathbf{H}}$, in turn, is denoted by

$$\overline{\text{MMSE}}_j \triangleq E[\text{MMSE}_j]. \quad (8)$$

In addition, it can be verified that

$$\text{Tr} \left\{ \left(\mathbf{I} + \frac{\text{SNR}}{n_T} \hat{\mathbf{H}} \mathbf{P} \hat{\mathbf{H}}^\dagger \right)^{-1} \right\} = \sum_{\ell=1}^{n_T} \text{MMSE}_\ell + n_R - n_T. \quad (9)$$

Using (6)–(9), the conditions in (5) can be rewritten as

$$p_j = 0 \quad \frac{\text{SNR}}{n_T} E \left[\hat{\mathbf{h}}_j^\dagger \mathbf{B}_j \hat{\mathbf{h}}_j \right] \leq \frac{1}{n_T} \sum_{\ell=1}^{n_T} (1 - \overline{\text{MMSE}}_\ell) \quad (10)$$

$$p_j = \frac{1 - \overline{\text{MMSE}}_j}{\frac{1}{n_T} \sum_{\ell=1}^{n_T} (1 - \overline{\text{MMSE}}_\ell)} \quad \text{otherwise} \quad (11)$$

The term

$$\frac{\text{SNR}}{n_T} E \left[\hat{\mathbf{h}}_j^\dagger \mathbf{B}_j \hat{\mathbf{h}}_j \right] \quad (12)$$

in (10) can be interpreted as the expected SINR per unit power exhibited by the signal on the j -th eigenvector with the other powers, p_ℓ for $\ell \neq j$, at their optimal values. This is a measure of the average signalling level achievable on the j -th eigenvector. When such average signalling level is below some threshold, no power should be allocated to that eigenvector. On the other hand, the fraction of the available power allocated to each signalling eigenvector whose average level is above threshold should equal the fraction of average signal recovered from that eigenvector by an MMSE receiver. Hence, those signalling eigenvectors from which a stronger average signal level can be recovered should be allocated a larger share of the power budget.

The conditions in (11) constitute a set of coupled equations that begets an iterative approach. In order to accommodate the iterative nature of the resulting algorithm, we use $(\cdot)^{(k)}$ to index the succession of values taken by each of the quantities involved.

Algorithm 1 Initialize $\mathbf{P}^{(0)}$ to best guess with $\text{Tr}\{\mathbf{P}^{(0)}\} > 0$. (If no prior information, set $\mathbf{P}^{(0)} = \mathbf{I}$). Follow the steps:

1. Iterate until solution settles:

$$p_j^{(k+1)} = \frac{1 - \overline{\text{MMSE}}_j^{(k)}}{\frac{1}{n_T} \sum_{\ell=1}^{n_T} (1 - \overline{\text{MMSE}}_\ell^{(k)})} \quad j = 1, \dots, n_T$$

2. If (10) is satisfied for every j such that p_j has converged to zero, allocation is completed. Otherwise, set $p_j = 0$ for j corresponding to the lowest value of (12) and repeat steps 1 and 2.

Note that Step 1 adjusts $\text{Tr}\{\mathbf{P}^{(k)}\} = n_T$ for $k > 0$ even if $\text{Tr}\{\mathbf{P}^{(0)}\} \neq n_T$. Hence, the total transmitted power is held at the correct value throughout the iterations as long as the initial powers are not identically zero. Notice also that, if a particular power is initialized to zero, it remains at zero indefinitely. Thus, except for those that are known to be zero, the powers within $\mathbf{P}^{(0)}$ should be strictly positive on the first pass.

As will be illustrated in Section VI through a sequence of examples, the iterations converge rapidly to the sought fixed-point set of powers.

V When does an Isotropic Input Achieve Capacity ?

There are relevant channel structures—beyond the IID zero-mean Gaussian case reported in [6]—for which an isotropic input achieves capacity. To characterize some of these channels, we shall make use of the following:

Definition 4 [40] A $(n_R \times n_T)$ matrix \mathbf{B} taking values in $\mathcal{B} \subset \mathbb{R}^+$ is column-regular if the entries of every column exhibit the same empirical distribution, i.e.,

$$\frac{1}{n_R} \sum_{i=1}^{n_R} 1\{(\mathbf{B})_{i,j} < \xi\}$$

does not depend on j , with $1\{\cdot\}$ the indicator function.

A broad characterization is encapsulated in the following result:

Proposition 1 Consider an arbitrary channel \mathbf{H} whose capacity-achieving input covariance is $\Phi = \mathbf{V}\mathbf{P}\mathbf{V}^\dagger$. If the joint distribution of the entries of $\hat{\mathbf{H}} = \mathbf{H}\mathbf{V}$ is invariant to a cyclic shift of its columns, then $\Phi = \mathbf{I}$ achieves capacity.

Proof: See Appendix C.

Application of this result to the channels in Theorems 1–3 yields physical insight:

- Consider first the zero-mean channel $\mathbf{H} = \mathbf{U}_R \tilde{\mathbf{H}} \mathbf{U}_T^\dagger$ with the columns of $\tilde{\mathbf{H}}$ required only to be zero-mean and independent. An isotropic input achieves capacity if the columns of $\tilde{\mathbf{H}}$ are marginally identically distributed. If the channel entries are jointly Gaussian, then it suffices that the variance matrix, $\tilde{\mathbf{G}}$, be column-regular as per Definition 4. If, further, the correlation is separable, the result holds if only the transmit antennas are uncorrelated ($\Theta_T = \mathbf{I}$) regardless of the receive correlations, Θ_R , as observed in [9].
- In the case of a standard Ricean channel, $\mathbf{H} = \bar{\mathbf{H}} + \mathbf{U}_R \tilde{\mathbf{H}} \mathbf{U}_T^\dagger$ with $\tilde{\mathbf{H}}$ having IID Gaussian entries, a sufficient condition is that the singular values of $\bar{\mathbf{H}}$ be identical.¹²
- Finally, in a keyhole channel, $\mathbf{H} = \mathbf{c}_R \mathbf{c}_T^\dagger$, an isotropic input attains capacity whenever the entries of \mathbf{c}_T are identically distributed.

VI Examples

In this section, we illustrate the convergence Algorithm 1 through a sequence of examples. In every case, the expectations in (8) are computed as averages of 10000 independent channel realizations.

Let us start with a Rayleigh-faded channel with separable correlations:

Example 1 Consider a 3-antenna transmit array, linear with 1-wavelength antenna spacing, and a broadside (truncated) Gaussian power azimuth spectrum with 2° root-mean-square spread. The corresponding transmit correlation is $(\Theta_T)_{i,j} \approx e^{-0.05(i-j)^2}$ [41]. Further consider $n_R = 4$ uncorrelated

¹²For arbitrary standard Ricean channels, the badness of an isotropic input is bounded in [15].

receive antennas, i.e., $\Theta_R = \mathbf{I}$. Signalling over the eigenvectors of Θ_T , as dictated by Theorem 1, the convergence at SNR = -3 dB and SNR = 5 dB is depicted in Fig. 2.

Observe the anticipated low-SNR behavior being manifest at -3 dB, where an individual eigenvector is allocated the entire transmit power. Furthermore, because of the strong transmit correlation, one of the eigenvectors remains inactive even at SNR = 5 dB.

Next, we evaluate a standard Ricean channel:

Example 2 Let $n_T = 3$ and $n_R = 2$ with

$$\mathbf{H} = \frac{1}{\sqrt{2}}\bar{\mathbf{H}} + \frac{1}{\sqrt{2}}\tilde{\mathbf{H}}$$

where $(\bar{\mathbf{H}})_{i,j} = 1$ while the entries of $\tilde{\mathbf{H}}$ are zero-mean IID Gaussian with unit variance. The Ricean factor is thus 0 dB. As per Theorem 2, the signalling eigenvectors are taken to be those of $\bar{\mathbf{H}}^\dagger \bar{\mathbf{H}}$. The convergence at SNR = 5 dB is portrayed in Fig. 3.

It is shown in [7, 13] that, in a standard Ricean channel, the powers transmitted on the eigenvectors corresponding to the zero-eigenvalue eigenspace of $\bar{\mathbf{H}}^\dagger \bar{\mathbf{H}}$ (in this example, p_2 and p_3) ought to be equal. This is corroborated in Fig. 3, where not only are their final values equal, but their convergence is simultaneous.

Finally, we home in on a polarization diversity channel conforming to Proposition 1.

Example 3 Consider $n_T = 4$, with two such antennas vertically polarized and the other two horizontally polarized, and $n_R = 2$ where one the antennas is vertically polarized and the other one horizontally polarized. Let the cross-polar discrimination (ratio of the power gain between co-polarized and cross-polarized antennas) be 7 dB. Given the low correlation between orthogonal polarizations, the entries of the channel matrix are modelled as independent and Rayleigh-faded with variances

$$\tilde{\mathbf{G}} = \frac{1}{3} \begin{pmatrix} 5 & 1 & 5 & 1 \\ 1 & 5 & 1 & 5 \end{pmatrix}$$

Since $\tilde{\mathbf{G}}$ is column-regular, capacity is achieved with $\Phi = \mathbf{I}$. The performance of Algorithm 1 at SNR = 3 dB, depicted in Fig. 4, shows rapid convergence to this solution. For emphasis, the algorithm is initialized with $\mathbf{P}^{(0)} = \text{diag}\{2.8, 0.6, 0.4, 0.2\}$, a strongly nonuniform allocation.

VII Conclusions

A broad characterization of the capacity-achieving input covariance for multi-antenna channels known instantaneously at the receiver while only in distribution at the transmitter has been presented. Unlike most previous contributions, our analysis has comprised both the eigenvectors (signalling directions) and the eigenvalues (transmission powers).

In terms of the eigenvectors, our analysis has expanded and unified previous results stating that, for a wide class of channel structures, the capacity-achieving eigenvectors should diagonalize the channel (or, more precisely, the matrix $\mathbf{H}^\dagger\mathbf{H}$) in an average manner. This finding is particularly appealing for it mirrors, in a statistical sense, the solution in the regime where the channel is known instantaneously by the transmitter, where the channel is instantaneously diagonalized.

Regarding the transmit powers, in contrast, the solution is not—despite several such claims in the literature—a statistical extension of the waterfill encountered when the transmitter has CSI. Because of the lack of orthogonality between the corresponding parallel channels, part of the power transmitted on each eigenvector spills as interference onto the other ones and thus the powers are mutually coupled beyond their sum constraint. What we have shown is that, rather than obtained via waterfill, the fraction of available transmit power allocated to each signalling eigenvector should equal the fraction of average signal power recovered, by an MMSE receiver, from that eigenvector.¹³ Although this solution does yield some behaviors that are reminiscent of a waterfill, it is in general quite distinct. Particularly noteworthy are the limiting power allocations at low and high SNR:

- At low SNR, concentrating power on the strongest eigenvector(s) is the capacity-achieving policy.
- At high SNR, where a MMSE receiver behaves in zero-forcing mode, the allocation is sensitive to the ratio between the number of transmit and receive antennas. When the number of receivers equals or exceeds the number of transmitters, a zero-forcer can extract the signal from each eigenvector while completely removing the interference

¹³This adds to the body of connections between channel capacity and MMSE estimation (see, e.g., [42, 43]). A fundamental relationship underlying these connections has been recently uncovered [44].

arising from the other ones. The powers thus decouple and the resulting policy is a uniform power allocation. With more transmitters than receivers, however, this is not the case.

Although in general not explicit, our solution for the transmit powers leads rather straightforwardly to an iterative algorithm, far more alluring than a standard numerical procedure, whose main features are:

- It has immediate operational significance. The iterative computation of the powers requires only average mean-square errors, which in many receiver implementations are readily available as a by-product.
- It is universally applicable. Given an arbitrary channel and a set of signalling eigenvectors, it finds the power allocation that maximizes the rate attained by a Gaussian input signalling thereon. Applied to the capacity-achieving eigenvectors, it finds the capacity-achieving power allocation.
- It requires no step-size or parameter adjustments.
- It is very robust in terms of initial conditions, which are only required to be nonzero.
- It exhibits very rapid convergence. This might render it suitable for implementation and use in time-varying environments, where the channel distribution itself is subject to slow large-scale variations that require tracking.

Besides fast convergence, real-time tracking would demand robustness concerning the calculation of the required averages. The coherence distance of the channel distribution typically ranges between a few meters and a few tens of meters [45]. At frequencies of 2–5 GHz, where the fade duration is on the order of a few cm, this coherence distance may expose no more than a few hundred independent channel realizations whereby those averages must be extracted. In order to assess the performance of Algorithm 1 in such conditions, we revisit Example 1 with the expectations computed, this time, as averages of 100 (instead of 10000) independent channel realizations. The convergence, displayed in Fig. 5 for both $\text{SNR}=-3$ dB and $\text{SNR}=5$ dB, is almost undistinguishable from that in Fig. 1.

To finalize, some additional design lessons emerging from our analysis:

- Isotropic inputs are adequate if the columns of the channel matrix, or of any rotated version thereof, are independent and balanced in terms of power.
- In the widely used separable correlation model, this requirement reverts simply to lack of correlation at the transmitter, irrespective of the receive correlation. Moreover, even in the presence of transmit correlation, the receive correlation has little (if any) influence on the capacity-achieving input: it has no impact on the signalling eigenvectors and it plays no role on the low- and high-SNR transmit power allocation.¹⁴
- A proper structuring of the input usually increases the supported rates markedly at low SNR. The advantage tends to be more modest at moderate and high SNR, where an isotropic input can often attain a hefty fraction of the capacity.

Acknowledgements

The highly constructive comments made by the reviewers are gratefully acknowledged.

Appendix

A Proof of Theorems 1 and 3

We start by proving Theorem 1 using a technique similar to the one introduced in [14]. Recall that $\Phi = \mathbf{V}\mathbf{P}\mathbf{V}^\dagger$. In order to prove that achieving capacity requires $\mathbf{V} = \mathbf{U}_T$, we need to show that the capacity-achieving covariance for the rotated channel $\mathbf{H}\mathbf{U}_T$ is diagonal. In terms of this rotated channel, the mutual information (in bits/s/Hz) is

$$\mathcal{I}(\text{SNR}, \mathbf{P}) = E \left[\log_2 \det \left(\mathbf{I} + \frac{\text{SNR}}{n_T} \tilde{\mathbf{H}}\mathbf{P}\tilde{\mathbf{H}}^\dagger \right) \right] \quad (13)$$

where \mathbf{U}_R turns out to be immaterial. We want to show that nonzero off-diagonal entries in \mathbf{P} can only reduce $\mathcal{I}(\text{SNR}, \mathbf{P})$. To that end, define $\mathbf{\Pi}_\ell$ as a diagonal matrix all of whose

¹⁴Its lack of influence on the capacity-achieving input notwithstanding, receive correlation may strongly affect the ensuing capacity.

diagonal entries are 1 except for the (ℓ, ℓ) -th entry, which is -1 . The entries of $\mathbf{\Pi}_\ell \mathbf{P} \mathbf{\Pi}_\ell^\dagger$ equal those of \mathbf{P} except for the off-diagonals in the ℓ -th row and ℓ -th column, whose sign is reversed. Then, $\text{Tr}\{\mathbf{\Pi}_\ell \mathbf{P} \mathbf{\Pi}_\ell^\dagger\} = \text{Tr}\{\mathbf{P}\}$ and

$$\begin{aligned} \mathcal{I}(\text{SNR}, \mathbf{\Pi}_\ell \mathbf{P} \mathbf{\Pi}_\ell^\dagger) &= E \left[\log_2 \det \left(\mathbf{I} + \frac{\text{SNR}}{n_T} (\tilde{\mathbf{H}} \mathbf{\Pi}_\ell) \mathbf{P} (\tilde{\mathbf{H}} \mathbf{\Pi}_\ell)^\dagger \right) \right] \\ &= E \left[\log_2 \det \left(\mathbf{I} + \frac{\text{SNR}}{n_T} \tilde{\mathbf{H}} \mathbf{P} \tilde{\mathbf{H}}^\dagger \right) \right] \\ &= \mathcal{I}(\text{SNR}, \mathbf{P}) \end{aligned} \quad (14)$$

where (14) follows from the fact that, since the columns of $\tilde{\mathbf{H}}$ are independent and their distribution symmetric, reversing the sign of the ℓ -th column does not alter the distribution. The matrix $\frac{1}{2}(\mathbf{P} + \mathbf{\Pi}_\ell \mathbf{P} \mathbf{\Pi}_\ell^\dagger)$ has entries equal to those of \mathbf{P} except for the off-diagonals in the ℓ -th row and ℓ -th column, which are zero. Invoking Jensen's inequality,

$$\begin{aligned} \mathcal{I} \left(\text{SNR}, \frac{1}{2}(\mathbf{P} + \mathbf{\Pi}_\ell \mathbf{P} \mathbf{\Pi}_\ell^\dagger) \right) &\geq \frac{\mathcal{I}(\text{SNR}, \mathbf{P}) + \mathcal{I}(\text{SNR}, \mathbf{\Pi}_\ell \mathbf{P} \mathbf{\Pi}_\ell^\dagger)}{2} \\ &= \mathcal{I}(\text{SNR}, \mathbf{P}) \end{aligned}$$

Hence, nulling the off-diagonal entries of any column and corresponding row of \mathbf{P} can only increase $\mathcal{I}(\text{SNR}, \mathbf{P})$. Repeating the same process n_T times, we find that (13) is maximized when \mathbf{P} is indeed diagonal.

Theorem 3 can be proved analogously. In this case, we need to show that maximizing

$$\mathcal{I}(\text{SNR}, \mathbf{P}) = E \left[\log_2 \det \left(\mathbf{I} + \frac{\text{SNR}}{n_T} \mathbf{c}_R \mathbf{c}_T^\dagger \mathbf{P} \mathbf{c}_T \mathbf{c}_R^\dagger \right) \right] \quad (15)$$

requires that \mathbf{P} be diagonal. Since the distribution of the vector \mathbf{c}_T is not altered by the sign reversal of its ℓ -th entry,

$$\mathcal{I}(\text{SNR}, \mathbf{P}) = \mathcal{I}(\text{SNR}, \mathbf{\Pi}_\ell \mathbf{P} \mathbf{\Pi}_\ell^\dagger)$$

and, from Jensen's inequality,

$$\mathcal{I} \left(\text{SNR}, \frac{1}{2}(\mathbf{P} + \mathbf{\Pi}_\ell \mathbf{P} \mathbf{\Pi}_\ell^\dagger) \right) \geq \mathcal{I}(\text{SNR}, \mathbf{P})$$

confirming that nulling the off-diagonal entries of any column and row of \mathbf{P} increases $\mathcal{I}(\text{SNR}, \mathbf{P})$. Repeating the process n_T times, we find that \mathbf{P} must be diagonal.

B Derivation of Eq. (5)

Let \mathbf{P} be the diagonal matrix that maximizes the strictly concave function (expressed, for convenience, in nats/s/Hz)

$$\mathcal{I}(\mathbf{P}_d) = E \left[\log_e \det \left(\mathbf{I} + \frac{\text{SNR}}{n_T} \hat{\mathbf{H}} \mathbf{P}_d \hat{\mathbf{H}}^\dagger \right) \right] \quad (16)$$

over the convex set of diagonal positive semidefinite matrices \mathbf{P}_d such that $\text{Tr}\{\mathbf{P}_d\}=n_T$. This maximum is characterized by a set of Kuhn-Tucker conditions, which we derive from first principles [46, Section 15.2]. Specifically, we impose that the derivative of (16) in the direction from \mathbf{P} to any alternative matrix \mathbf{P}_d be negative. Letting

$$\mathbf{P}_\mu = (1 - \mu)\mathbf{P} + \mu\mathbf{P}_d$$

for $0 \leq \mu \leq 1$, the one-side derivative of (16) with respect to μ at $\mu=0^+$ is

$$\frac{d}{d\mu} \mathcal{I}(\mathbf{P}_\mu) = E \left[\text{Tr} \left\{ \left(\mathbf{I} + \frac{\text{SNR}}{n_T} \hat{\mathbf{H}} \mathbf{P}_d \hat{\mathbf{H}}^\dagger \right) \left(\mathbf{I} + \frac{\text{SNR}}{n_T} \hat{\mathbf{H}} \mathbf{P} \hat{\mathbf{H}}^\dagger \right)^{-1} - \mathbf{I} \right\} \right]$$

and, therefore, we impose that

$$E \left[\text{Tr} \left\{ \left(\mathbf{I} + \frac{\text{SNR}}{n_T} \hat{\mathbf{H}} \mathbf{P}_d \hat{\mathbf{H}}^\dagger \right) \left(\mathbf{I} + \frac{\text{SNR}}{n_T} \hat{\mathbf{H}} \mathbf{P} \hat{\mathbf{H}}^\dagger \right)^{-1} - \mathbf{I} \right\} \right] \leq 0 \quad (17)$$

for every \mathbf{P}_d in the set. Since the left-hand side of (17) is affine on \mathbf{P}_d , it suffices to impose it on the extreme points of the set. Moreover, the line connecting the j -th extreme point ($p_j=n_T$, $p_\ell=0$ for $\ell \neq j$) with \mathbf{P} can be extended beyond \mathbf{P} if and only if the optimum p_j is strictly positive, in which case the derivative at \mathbf{P} vanishes and (17) is a strict equality. Otherwise, if the optimum p_j is zero, (17) remains an inequality. With these considerations, the necessary and sufficient conditions in (5) are readily obtained from (17).

Let us now restrict our attention to $n_T \leq n_R$ over channels such that $p_j > 0 \forall j$. Under these conditions, $\hat{\mathbf{H}}^\dagger \hat{\mathbf{H}}$ and \mathbf{P} are nonsingular with probability 1. (If some of the powers are zero, the same consideration can be made for the nonzero powers by removing the corresponding rows and columns within the various matrices involved.) Letting $\text{SNR} \rightarrow \infty$ in (17), the equality is satisfied if $\mathbf{P}=\mathbf{I}$. This can also be seen directly from the capacity itself since

$$\lim_{\text{SNR} \rightarrow \infty} \left(E \left[\log \det \left(\mathbf{I} + \frac{\text{SNR}}{n_T} \hat{\mathbf{H}}^\dagger \hat{\mathbf{H}} \mathbf{P} \right) \right] - n_T \log \text{SNR} \right) = E \left[\log \det \left(\frac{1}{n_T} \hat{\mathbf{H}}^\dagger \hat{\mathbf{H}} \right) \right] + \log \det \mathbf{P}$$

which, because of the concavity of the $\log \det(\cdot)$ function and the constraint on the trace of \mathbf{P} , is maximized for $\mathbf{P}=\mathbf{I}$.

C Proof of Proposition 1

For $\hat{\mathbf{H}}=\mathbf{H}\mathbf{V}$, the capacity-achieving input covariance is $\mathbf{P}=\text{diag}\{p_1, \dots, p_{n_T}\}$. Denote by $\mathbf{P}^{\{m\}}$ a cyclic shift of \mathbf{P} by m positions, i.e., another diagonal matrix such that

$$p_j^{\{m\}} = p_{j'}$$

with $j'=(j-m)_{\text{mod } n_T}$. Clearly, $\text{Tr}\{\mathbf{P}^{\{m\}}\}=n_T$ for any shift $m=1, \dots, n_T$ and

$$\frac{1}{n_T} \sum_{m=1}^{n_T} \mathbf{P}^{\{m\}} = \mathbf{I}$$

Invoking Jensen's inequality, the mutual information satisfies

$$\begin{aligned} E \left[\log_2 \det \left(\mathbf{I} + \frac{\text{SNR}}{n_T} \hat{\mathbf{H}} \hat{\mathbf{H}}^\dagger \right) \right] &\geq \frac{1}{n_T} \sum_{m=1}^{n_T} E \left[\log_2 \det \left(\mathbf{I} + \frac{\text{SNR}}{n_T} \hat{\mathbf{H}} \mathbf{P}^{\{m\}} \hat{\mathbf{H}}^\dagger \right) \right] \\ &= E \left[\log_2 \det \left(\mathbf{I} + \frac{\text{SNR}}{n_T} \hat{\mathbf{H}} \mathbf{P}^{\{m\}} \hat{\mathbf{H}}^\dagger \right) \right] \end{aligned} \quad (18)$$

where (18) holds if the joint distribution of the entries of $\hat{\mathbf{H}}$ is invariant to a cyclic shift of its columns by m positions.

References

- [1] A. Lapidoth and S. Shamai, "Fading channels: How perfect need "perfect side information" be?," *IEEE Trans. on Inform. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.
- [2] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multi-antenna systems on flat-fading channels," *IEEE Trans. on Inform. Theory, Special Issue on Space-Time Transmission, Reception, Coding and Signal Design*, 2003.
- [3] G. Taricco and M. Elia, "Capacity of fading channels with no side information," *Electronic Letters*, vol. 33, pp. 1368–1370, July 1997.
- [4] A. M. Sengupta and P. P. Mitra, "Capacity of multivariate channels with multiplicative noise: I. Random matrix techniques and large-N expansions for full transfer matrices," *LANL arXiv:physics/0010081*, Oct. 2000.
- [5] G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communications," *IEEE Trans. Commun.*, vol. 46, no. 3, 1998.
- [6] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecom*, vol. 10, pp. 585–595, Nov. 1999.

- [7] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. on Inform. Theory*, vol. 47, pp. 2632–2639, Sept. 2001.
- [8] S. A. Jafar and A. J. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Trans. on Wireless Communications*, vol. 3, no. 4, pp. 1165–1175, July 2004.
- [9] W. Rhee and J. M. Cioffi, "On the capacity of multiuser wireless channels with multiple antennas," *IEEE Trans. on Inform. Theory*, vol. 49, no. 10, pp. 2580–2595, Oct. 2003.
- [10] J. Kotecha and A. Sayeed, "On the capacity of correlated MIMO channels," *Proc. of Intern. Symp. on Inform. Theory (ISIT'03)*, July 2003.
- [11] A. L. Moustakas, S. H. Simon, and A. M. Sengupta, "MIMO capacity through correlated channels in the presence of correlated interferers and noise: a (not so) large N analysis," *IEEE Trans. on Inform. Theory*, vol. 49, no. 10, pp. 2545–2561, Oct. 2003.
- [12] E. Jorswieck and H. Boche, "Optimal transmission with imperfect channel state information at the transmit array side," *Wireless Personal Communications*, vol. 27, no. 1, pp. 33–56, Oct. 2003.
- [13] S. Venkatesan, S. H. Simon, and R. A. Valenzuela, "Capacity of a Gaussian MIMO channel with nonzero mean," *Proc. of IEEE Veh. Tech. Conf. (VTC'03), Orlando, FL*, Oct. 2003.
- [14] D. Höslı and A. Lapidoth, "The capacity of a MIMO Ricean channel is monotonic in the singular values of the mean," *5th Intern. ITG Conf. on Source and Channel Coding, Erlangen, Germany*, pp. 381–385, Jan. 2004.
- [15] D. Höslı and A. Lapidoth, "How good is an isotropic Gaussian input on a MIMO Ricean channel?," *Proc. of Intern. Symp. on Inform. Theory (ISIT'04), Chicago, IL*, July 2004.
- [16] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. on Inform. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.
- [17] H. Sampath and A. Paulraj, "Linear precoding for space-time coded systems with known fading correlations," *IEEE Commun. Letters*, vol. 6, no. 6, pp. 239–241, June 2002.
- [18] R. U. Nabar, H. Bolcskei, and A. Paulraj, "Transmit optimization for spatial multiplexing in the presence of spatial fading correlation," *IEEE GLOBECOM'01*, vol. 11, pp. 131–135, 2001.
- [19] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multi-carrier MIMO channels: a unified framework for convex optimization," *IEEE Trans. on Signal Processing*, vol. 51, no. 9, pp. 2381–2401, Sept. 2003.
- [20] S. Zhou and G. B. Giannakis, "Optimal transmitter eigen-beamforming and space-time block coding based on channel correlations," *IEEE Trans. on Inform. Theory*, vol. 49, no. 7, pp. 1673–1690, July 2003.
- [21] O. Simeone and U. Spagnolini, "Combined linear pre-equalization and BLAST equalization with channel correlation feedback," *IEEE Commun. Letters*, vol. 7, no. 10, pp. 487–489, Oct. 2003.

- [22] O. Simeone, Y. Bar-Ness, and U. Spagnolini, "Linear and nonlinear preequalization / equalization for MIMO systems with long-term channel state information at the transmitter," *IEEE Trans. on Wireless Communications*, vol. 3, no. 2, pp. 373–378, Mar. 2004.
- [23] J. Akhtar and D. Gesbert, "A closed-form precoder for spatial multiplexing over correlated MIMO channels," *IEEE GLOBECOM'03*, vol. 4, pp. 1–5, Dec. 2003.
- [24] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effects on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, Mar. 2000.
- [25] H. Ozelik, M. Herdin, W. Weichselberger, J. Wallace, and E. Bonek, "Deficiencies of the 'kro-necker' MIMO radio channel model," *IEE Electronic Letters*, vol. 39, pp. 1209–1210, Aug. 2003.
- [26] A. M. Tulino, S. Verdú, and A. Lozano, "Capacity of antenna arrays with space, polarization and pattern diversity," *Inform. Theory Workshop (ITW'03), Paris, France*, Apr. 2003.
- [27] J. H. Kotecha and A. Sayeed, "Optimal signal design for estimation of correlated MIMO channels," *IEEE Intern. Conf. in Commun. (ICC'03)*, pp. 3170–3174, May 2003.
- [28] W. Weichselberger, M. Herdin, H. Ozelik, and E. Bonek, "A stochastic MIMO channel model with joint correlation of both link ends," *IEEE Transactions on Wireless Communications (to appear)*.
- [29] W. K. Pratt, *Digital Image Processing*, New York: Wiley, 1978.
- [30] A. M. Sayeed, "Deconstructing multi-antenna channels," *IEEE Trans. on Signal Processing*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.
- [31] D. Chizhik, G. J. Foschini, M. J. Gans, and R. A. Valenzuela, "Keyholes, correlations and capacities of multi-element transmit and receive antennas," *IEEE Trans. on Wireless Commun.*, vol. 1, no. 2, pp. 361–368, Apr. 2002.
- [32] D. Gesbert, H. Bolcskei, D. Gore, and A. J. Paulraj, "MIMO wireless channels: Capacity and performance prediction," *Proc. of IEEE GLOBECOM'2000, San Francisco, CA, USA*, Dec. 2000.
- [33] D. Pérez Palomar, J. M. Cioffi, and M. A. Lagunas, "Uniform power allocation in MIMO channels: A game theoretic approach," *IEEE Trans. on Inform. Theory*, vol. 49, no. 7, pp. 1707–1727, July 2003.
- [34] S. Wei and D. Goeckel, "On the minimax robustness of the uniform transmission power strategy in MIMO systems," *IEEE Commun. Letters*, vol. 7, no. 11, pp. 523–524, Nov. 2003.
- [35] S. H. Simon and A. L. Moustakas, "Optimizing MIMO antenna systems with channel covariance feedback," *IEEE J. Select. Areas Commun.*, vol. 21, no. 3, pp. 406–417, Apr. 2003.
- [36] Y. Liang and V. Veeravalli, "Correlated MIMO Rayleigh fading channels: Capacity and optimal signalling," *37th Asilomar Conf. on Signals, Systems and Computers, Pacific Grove, CA*, Nov. 2003.

- [37] A. Lozano, A. M. Tulino, and S. Verdú, "Multiantenna capacity: Myths and realities," *Space-Time Wireless Systems: From Array Processing to MIMO Communications* (Editors: H. Bölcskei, D. Gesbert, C. Papadias and J. Van der Veen), Cambridge University Press, 2005.
- [38] A. Lozano, A. M. Tulino, and S. Verdú, "High-SNR power offset in multi-antenna communication," *Proc. of Intern. Symp. on Inform. Theory (ISIT'04)*, Chicago, IL, July 2004.
- [39] S. Verdú, *Multiuser Detection*, Cambridge University Press, 1998.
- [40] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, 2004.
- [41] T.-S. Chu and L. J. Greenstein, "A semiempirical representation of antenna diversity gain at cellular and PCS base stations," *IEEE Trans. Commun.*, vol. 45–46, June 1997.
- [42] M. K. Varanasi and T. Guess, "Optimum decision-feedback multiuser equalization with successive decoding achieves the total capacity of the Gaussian multiple-access channel," *Asilomar Conference*, pp. 1405–1409, 1998.
- [43] S. Shamai and S. Verdú, "The impact of frequency-flat fading on the spectral efficiency of CDMA," *IEEE Trans. on Inform. Theory*, vol. 47, no. 5, May 2001.
- [44] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. on Inform. Theory (to appear)*, 2005.
- [45] M. Gudmunson, "Correlation model for shadow fading in mobile radio systems," *Electr. Letters*, vol. 27, no. 23, pp. 2145–2146, 1991.
- [46] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York, Wiley, 1990.

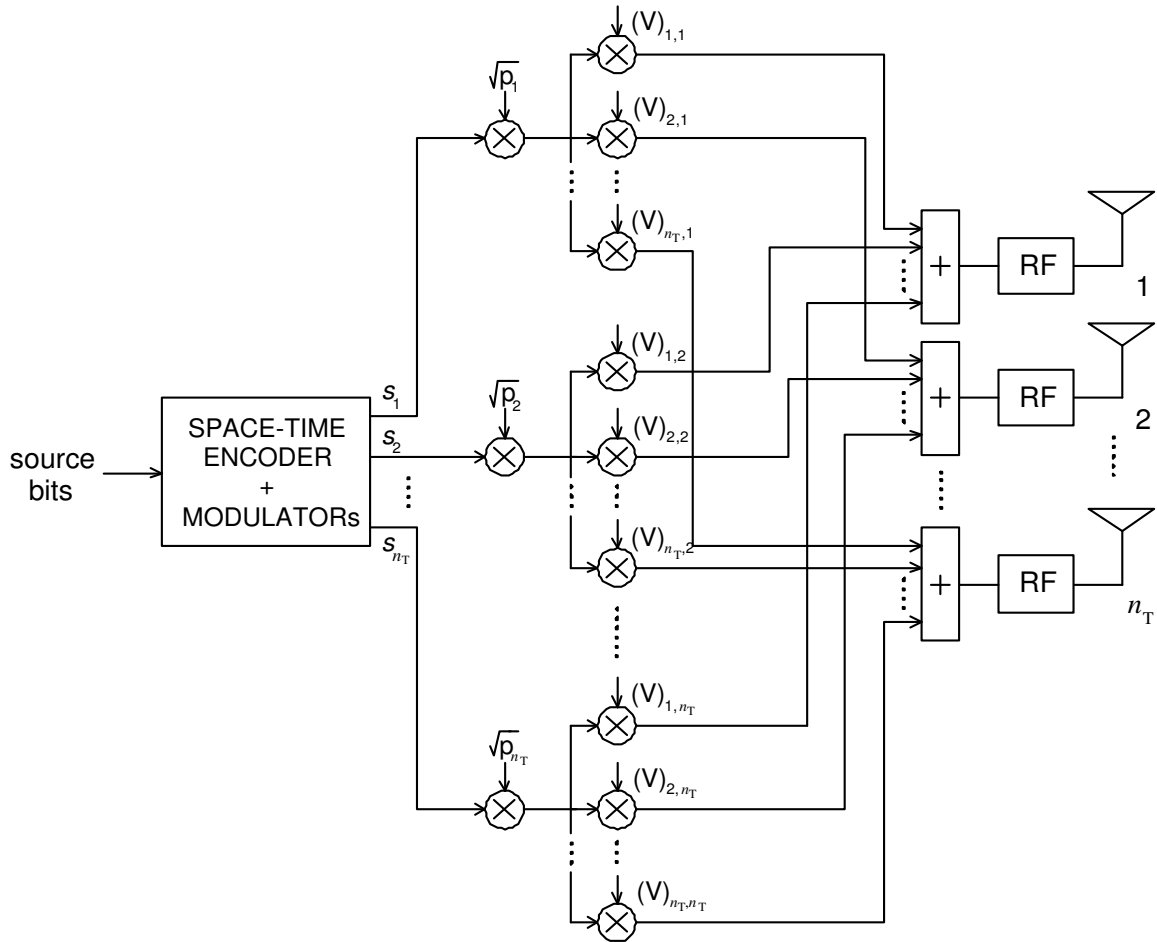


Figure 1: Transmit architecture generating a signal with spatial covariance $\Phi = \mathbf{V}\mathbf{P}\mathbf{V}^\dagger$ where \mathbf{V} is unitary and $\mathbf{P} = \text{diag}\{p_1, p_2, \dots, p_{n_T}\}$.

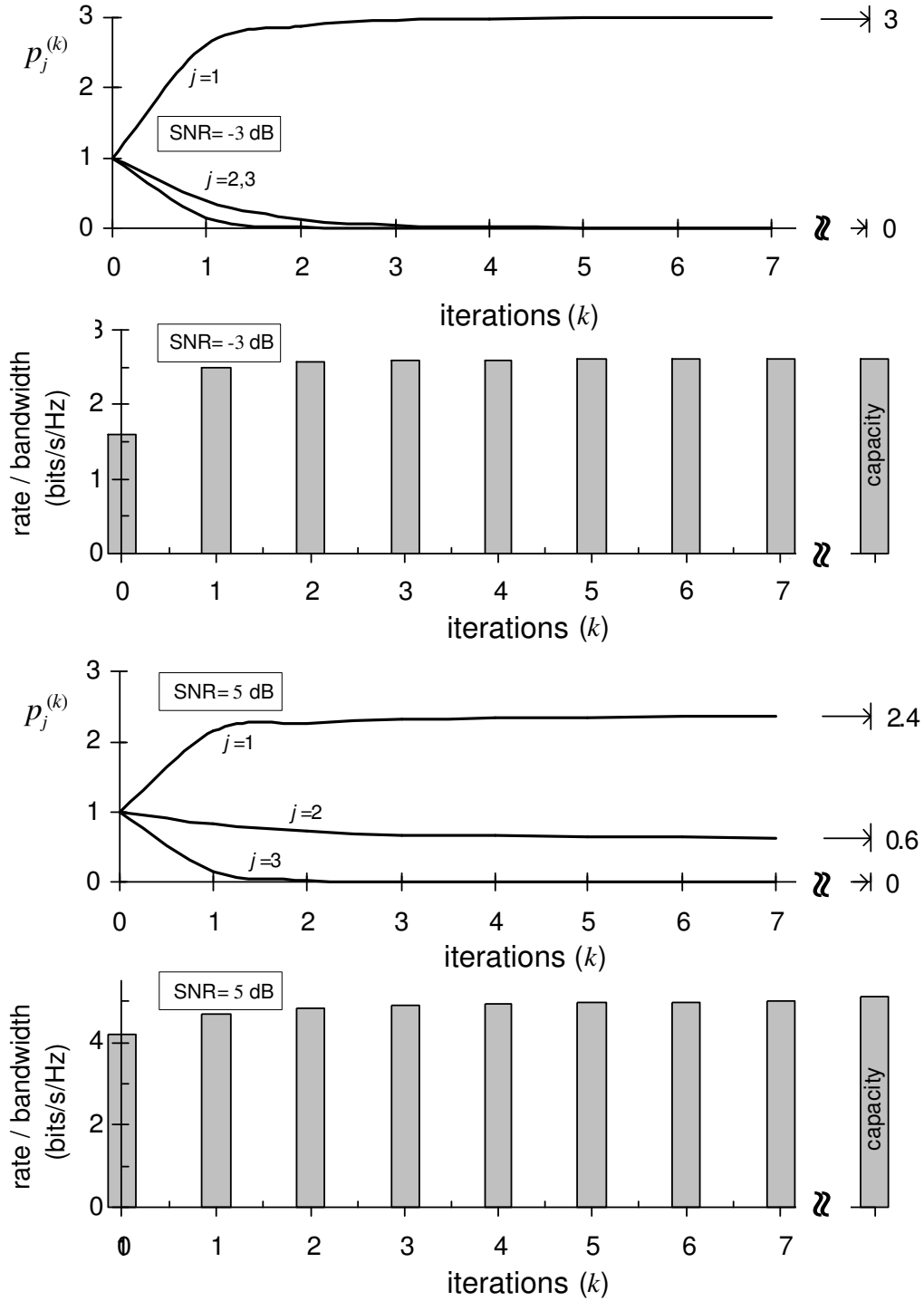


Figure 2: With $n_T=3$ and $n_R=4$ as per Example 1, values taken by $\mathbf{P}^{(k)}$ for $k=1, \dots, 7$ at SNR=-3 dB and SNR=5 dB given $\mathbf{P}^{(0)}=\mathbf{I}$. Also shown are the corresponding rates per unit bandwidth. The capacity-achieving powers are $\mathbf{P}=\text{diag}\{3, 0, 0\}$ at -3 dB and $\mathbf{P}=\text{diag}\{2.4, 0.6, 0\}$ at 5 dB.

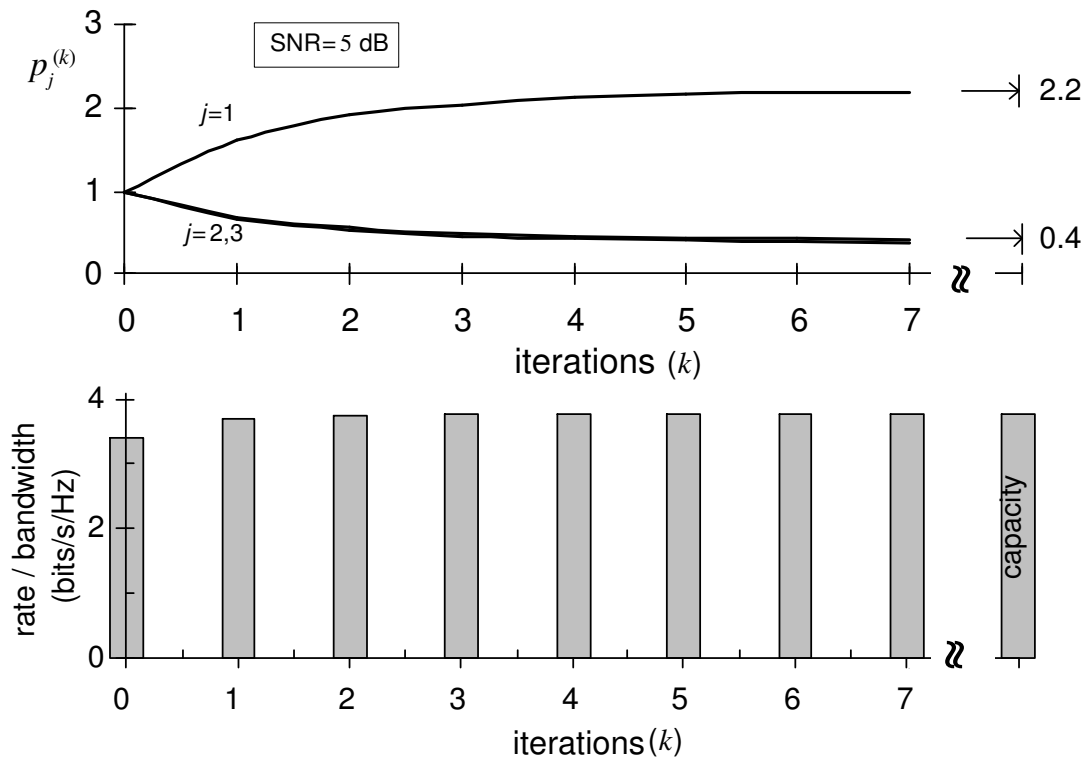


Figure 3: With $n_T=3$ and $n_R=2$ as per Example 2, values taken by $\mathbf{P}^{(k)}$ for $k=1, \dots, 7$ at SNR=5 dB given $\mathbf{P}^{(0)}=\mathbf{I}$. Also shown are the corresponding rates per unit bandwidth. The capacity-achieving powers are $\mathbf{P}=\text{diag}\{2.2, 0.4, 0.4\}$.

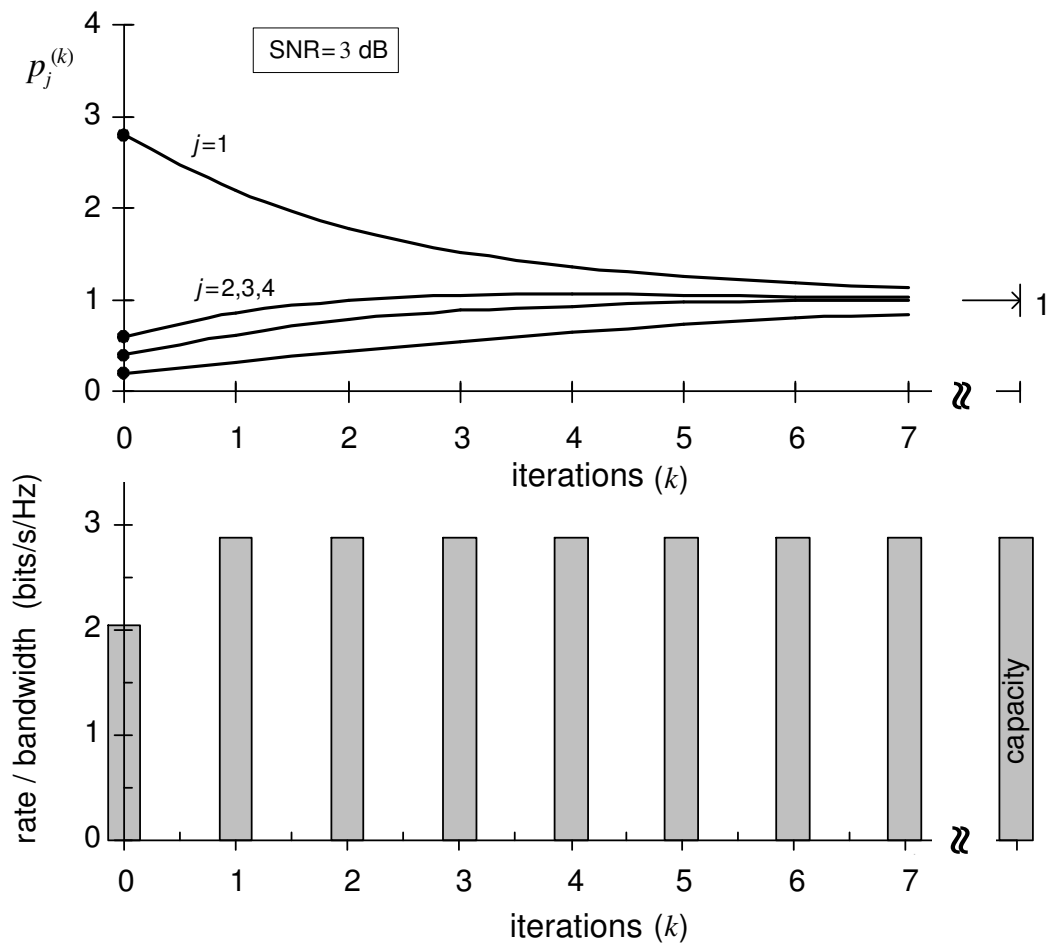


Figure 4: With $n_T=4$ and $n_R=2$ as per Example 3, values taken by $\mathbf{P}^{(k)}$ for $k=1, \dots, 7$ at SNR=3 dB with initialization $\mathbf{P}^{(0)}=\text{diag}\{2.8, 0.6, 0.4, 0.2\}$. Also shown are the corresponding rates per unit bandwidth. The capacity-achieving solution is $\mathbf{P}=\mathbf{I}$ (given by Proposition 1).

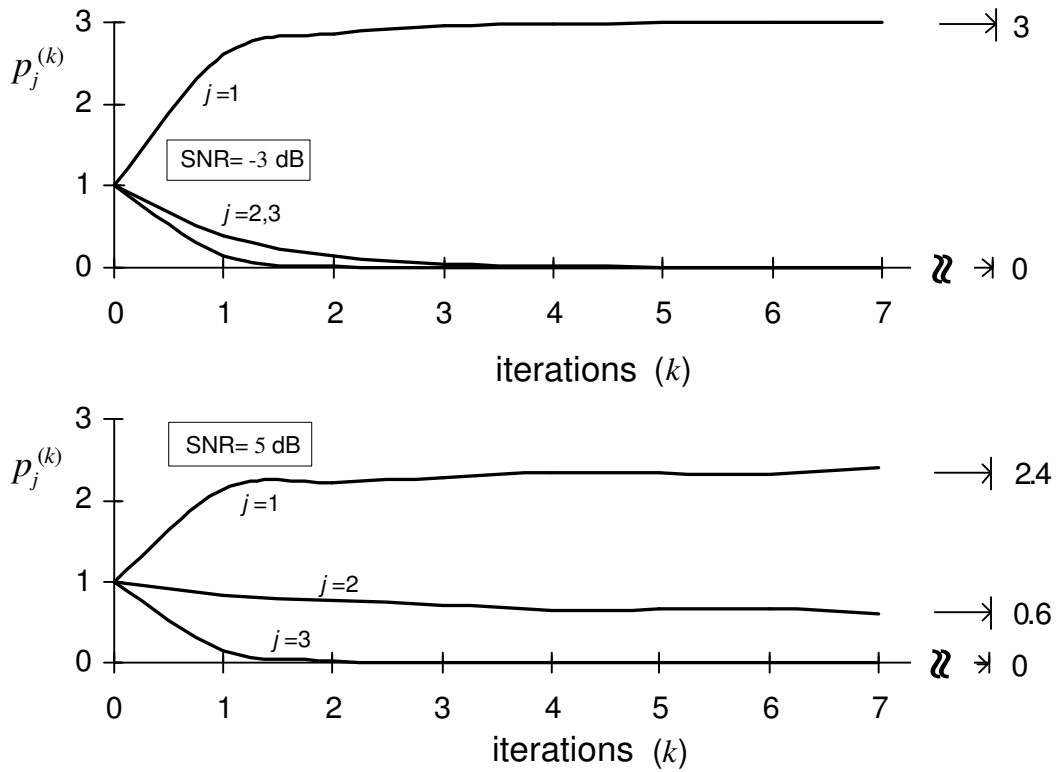


Figure 5: With $n_T=3$ and $n_R=4$ as per Example 1, values taken by $\mathbf{P}^{(k)}$ for $k=1, \dots, 7$ at SNR = -3 dB and 5 dB given $\mathbf{P}^{(0)}=\mathbf{I}$. At each iteration, the expectations are computed as averages of only 100 independent channel realizations.