

Article

Information Rates for Channels with Fading, Side Information and Adaptive Codewords

Gerhard Kramer 

School of Computation, Information and Technology, Technical University of Munich (TUM),
80333 Munich, Germany; gerhard.kramer@tum.de

Abstract: Generalized mutual information (GMI) is used to compute achievable rates for fading channels with various types of channel state information at the transmitter (CSIT) and receiver (CSIR). The GMI is based on variations of auxiliary channel models with additive white Gaussian noise (AWGN) and circularly-symmetric complex Gaussian inputs. One variation uses reverse channel models with minimum mean square error (MMSE) estimates that give the largest rates but are challenging to optimize. A second variation uses forward channel models with linear MMSE estimates that are easier to optimize. Both model classes are applied to channels where the receiver is unaware of the CSIT and for which adaptive codewords achieve capacity. The forward model inputs are chosen as linear functions of the adaptive codeword's entries to simplify the analysis. For scalar channels, the maximum GMI is then achieved by a conventional codebook, where the amplitude and phase of each channel symbol are modified based on the CSIT. The GMI increases by partitioning the channel output alphabet and using a different auxiliary model for each partition subset. The partitioning also helps to determine the capacity scaling at high and low signal-to-noise ratios. A class of power control policies is described for partial CSIR, including a MMSE policy for full CSIT. Several examples of fading channels with AWGN illustrate the theory, focusing on on-off fading and Rayleigh fading. The capacity results generalize to block fading channels with in-block feedback, including capacity expressions in terms of mutual and directed information.

Keywords: capacity; channel state information; directed information; fading; feedback; generalized mutual information; side information



Citation: Kramer, G. Information Rates for Channels with Fading, Side Information and Adaptive Codewords. *Entropy* **2023**, *25*, 728. <https://doi.org/10.3390/e25050728>

Academic Editors: Luca Barletta and Alex Dytso

Received: 3 March 2023

Revised: 26 March 2023

Accepted: 22 April 2023

Published: 27 April 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The capacity of fading channels is a topic of interest in wireless communications [1–4]. Fading refers to model variations over time, frequency, and space. A common approach to track fading is to insert pilot symbols into transmit symbol strings, have receivers estimate fading parameters via the pilot symbols, and have the receivers share their estimated channel state information (CSI) with the transmitters. The CSI available at the receiver (CSIR) and transmitter (CSIT) may be different and imperfect.

Information-theoretic studies on fading channels distinguish between average (ergodic) and outage capacity, causal and non-causal CSI, symbol and rate-limited CSI, and different qualities of CSIR and CSIT that are coarsely categorized as no, perfect, or partial. We refer to [5] for a review of the literature up to 2008. We here focus exclusively on average capacity and causal CSIT as introduced in [6]. Codes for such CSIT, or more generally for noisy feedback [7], are based on Shannon strategies, also called codetrees ([8], Chapter 9.4), or adaptive codewords ([9], Section 4.1). (The term “adaptive codeword” was suggested to the author by J. L. Massey.) Adaptive codewords are usually implemented by a conventional codebook and by modifying the codeword symbols as a function of the CSIT. This approach is optimal for some channels [10] and will be our main interest.

1.1. Block Fading

A model that accounts for the different time scales of data transmission (e.g., nanoseconds) and channel variations (e.g., milliseconds) is block fading [11,12]. Such fading has the channel parameters constant within blocks of L symbols and varying across blocks. A basic setup is as follows.

- The fading is described by a state process S_{H1}, S_{H2}, \dots independent of the transmitter messages and channel noise. The subscript “ H ” emphasizes that the states S_{Hi} may be hidden from the transceivers.
- Each receiver sees a state process S_{R1}, S_{R2}, \dots where S_{Ri} is a noisy function of S_{Hi} for all i .
- Each transmitter sees a state process S_{T1}, S_{T2}, \dots where S_{Ti} is a noisy function of S_{Hi} for all i .

The state processes may be modeled as memoryless [11,12] or governed by a Markov chain [13–21]. The memoryless models are particular cases of Shannon’s model [6]. For scalar channels, S_{Hi} is usually a complex number H_i . Similarly, for vector or multi-input, multi-output (MIMO) channels with M - and N -dimensional inputs and outputs, respectively, S_{Hi} is a $N \times M$ matrix \mathbf{H}_i .

Consider, for example, a point-to-point channel with block-fading and complex-alphabet inputs $X_{i\ell}$ and outputs

$$Y_{i\ell} = H_i X_{i\ell} + Z_{i\ell} \quad (1)$$

where the index $i, i = 1, \dots, n$, enumerates the blocks and the index $\ell, \ell = 1, \dots, L$, enumerates the symbols of each block. The additive white Gaussian noise (AWGN) Z_{11}, Z_{12}, \dots is a sequence of independent and identically distributed (i.i.d.) random variables that have a common circularly-symmetric complex Gaussian (CSCG) distribution.

1.2. CSI and In-Block Feedback

The motivation for modeling CSI as independent of the messages is simplicity. If one uses only pilot symbols to estimate the H_i in (1), for example, then the independence is valid, and the capacity analysis may be tractable. However, to improve performance, one can implement data and parameter estimation jointly, and one can actively adjust the transmit symbols $X_{i\ell}$ using past received symbols $Y_{ik}, k = 1, \dots, \ell - 1$, if in-block feedback is available. (Across-block feedback does not increase capacity if the state processes are memoryless; see ([22], Remark 16).) An information theory for such feedback was developed in [22], where a challenge is that code design is based on adaptive codewords that are more sophisticated than conventional codewords.

For example, suppose the CSIR is $S_{Ri} = H_i$. Then, one might expect that CSCG signaling is optimal, and the capacity is an average of $\log(1 + \text{SNR})$ terms, where SNR is a signal-to-noise ratio. However, this simplification is based on constraints, e.g., that the CSIT is a function of the CSIR and that the $X_{i\ell}$ cannot influence the CSIT. The former constraint can be realistic, e.g., if the receiver quantizes a pilot-based estimate of H_i and sends the quantization bits to the transmitter via a low-latency and reliable feedback link. On the other hand, the latter constraint is unrealistic in general.

1.3. Auxiliary Models

This paper’s primary motivation is to further develop information theory for adaptive codewords. To gain insight, it is helpful to have achievable rates with $\log(1 + \text{SNR})$ terms. A common approach to obtain such expressions is to lower bound the channel mutual information $I(X; Y)$ as follows.

Suppose X is continuous and consider two conditional densities: the density $p(x|y)$ and an auxiliary density $q(x|y)$. We will refer to such densities as reverse models; similarly,

$p(y|x)$ and $q(y|x)$ are called forward models. One may write the differential entropy of X given Y as

$$h(X|Y) = E[-\log p(X|Y)] = \underbrace{E[-\log q(X|Y)]}_{\text{average cross-entropy}} - \underbrace{E\left[\log \frac{p(X|Y)}{q(X|Y)}\right]}_{\text{average divergence}} \geq 0 \tag{2}$$

where the first expectation in (2) is an average cross-entropy, and the second is an average informational divergence, which is non-negative. Several criteria affect the choice of $q(x|y)$: the cross-entropy should be simple enough to admit theoretical or numerical analysis, e.g., by Monte Carlo simulation; the cross-entropy should be close to $h(X|Y)$; and the cross-entropy should suggest suitable transmitter and receiver structures.

We illustrate how reverse and forward auxiliary models have been applied to bound mutual information. Assume that $E[X] = E[Y] = 0$ for simplicity.

Reverse Model: Consider the reverse density that models X, Y as jointly CSCG:

$$q(x|y) = \frac{1}{\pi\sigma_L^2} \exp\left(-|x - \hat{x}_L|^2/\sigma_L^2\right) \tag{3}$$

where $\hat{X}_L = (E[XY^*]/E[|Y|^2])Y$ and

$$\sigma_L^2 = E\left[|X - \hat{X}_L|^2\right] = E\left[|X|^2\right] - \frac{|E[XY^*]|^2}{E[|Y|^2]} \tag{4}$$

is the mean square error (MSE) of the estimate \hat{X}_L . In fact, \hat{X}_L is the linear estimate with the minimum MSE (MMSE), and σ_L^2 is the linear MMSE (LMMSE) which is independent of $Y = y$; see Section 2.5. The bound in (2) gives

$$h(X|Y) \leq \log\left(\pi e \sigma_L^2\right). \tag{5}$$

Thus, if X is CSCG, then we have the desired form

$$I(X; Y) = h(X) - h(X|Y) \geq \log\left(1 + \frac{|h|^2 E[|X|^2]}{\sigma^2}\right) \tag{6}$$

where the parameters h and σ^2 are

$$h = \frac{E[XY^*]}{E[|X|^2]}, \quad \sigma^2 = E\left[|Y - hX|^2\right]. \tag{7}$$

The bound (6) is apparently due to Pinsker [23–25] and is widely used in the literature; see e.g., [18,26–38]. The bound is usually related to channels $p(y|x)$ with additive noise but (2)–(6) show that it applies generally. The extension to vector channels is given in Section 2.7 below.

Forward Model: A more flexible approach is to choose the reverse density as

$$q(x|y) = \frac{p(x)q(y|x)^s}{q(y)} \tag{8}$$

where $q(y|x)$ is a forward auxiliary model (not necessarily a density), $s \geq 0$ is a parameter to be optimized, and

$$q(y) = \int_{\mathbb{C}} p(x) q(y|x)^s dx. \tag{9}$$

Inserting (8) into (2) we compute

$$I(X; Y) \geq \max_{s \geq 0} \mathbb{E} \left[\log \frac{q(Y|X)^s}{q(Y)} \right]. \tag{10}$$

The right-hand side (RHS) of (10) is called a generalized mutual information (GMI) [39,40] and has been applied to problems in information theory [41], wireless communications [42–51], and fiber-optic communications [52–61]. For example, the bounds (6) and (10) are the same if $s = 1$ and

$$q(y|x) = \exp\left(-|y - hx|^2/\sigma^2\right) \tag{11}$$

where h and σ^2 are given by (7). Note that (11) is not a density unless $\sigma^2 = 1/\pi$ but $q(x|y)$ is a density. (We require $q(x|y)$ to be a density to apply the divergence bound in (2).)

We compare the two approaches. The bound (5) is simple to apply and works well since the choices (7) give the maximal GMI for CSCG X ; see Proposition 1 below. However, there are limitations: one must use continuous X , the auxiliary model $q(y|x)$ is fixed as (11), and the bound does not show how to design the receiver. Instead, the GMI applies to continuous/discrete/mixed X and has an operational interpretation: the receiver uses $q(y|x)$ rather than $p(y|x)$ to decode. The framework of such mismatched receivers appeared in ([62], Exercise 5.22); see also [63].

1.4. Refined Auxiliary Models

The two approaches above can be refined in several ways, and we review selected variations in the literature.

Reverse Models: The model $q(x|y)$ can be different for each $Y = y$, e.g., one may choose X as Gaussian with mean $\mathbb{E}[X|Y = y]$ and variance

$$\text{Var}[X|Y = y] = \mathbb{E}[|X|^2|Y = y] - |\mathbb{E}[X|Y = y]|^2 \tag{12}$$

and where

$$q(x|y) = \frac{1}{\pi \text{Var}[X|Y = y]} \exp\left(-\frac{|x - \mathbb{E}[X|Y = y]|^2}{\text{Var}[X|Y = y]}\right). \tag{13}$$

Inserting (13) in (2) we have the bound

$$h(X|Y) \leq \mathbb{E}[\log(\pi e \text{Var}[X|Y])] \tag{14}$$

which improves (5) in general, since $\text{Var}[X|Y = y]$ is the MMSE of X given the event $Y = y$. In other words, we have $\text{Var}[X|Y = y] \leq \sigma_L^2$ for all $Y = y$ and the following bound improves (6) for CSCG X :

$$I(X; Y) \geq \mathbb{E} \left[\log \frac{\mathbb{E}[|X|^2]}{\text{Var}[X|Y]} \right]. \tag{15}$$

In fact, the bound (15) was derived in ([50], Section III.B) by optimizing the GMI in (10) over all forward models of the form

$$q(y|x) = \exp\left(-|\tilde{g}_y - \tilde{f}_y x|^2\right) \tag{16}$$

where \tilde{f}_y, \tilde{g}_y depend on y ; see also [47–49]. We provide a simple proof. By inserting (16) into (8) and (9), absorbing the s parameter in \tilde{f}_y and \tilde{g}_y , and completing squares, one can equivalently optimize over all reverse densities of the form

$$q(x|y) = \exp\left(-|g_y - f_y x|^2 + h_y\right) \tag{17}$$

where $|f_y|^2 = \pi e^{h_y}$ so that $q(x|y)$ is a density. We next bound the cross-entropy as

$$\begin{aligned} E[-\log q(X|Y = y)] &= E\left[|g_y/f_y - X|^2\right] |f_y|^2 - h_y \\ &\geq \text{Var}[X|Y = y] \pi e^{h_y} - h_y \end{aligned} \tag{18}$$

with equality if $g_y/f_y = E[X|Y = y]$; see Section 2.5. The RHS of (18) is minimized by $\text{Var}[X|Y = y] \pi e^{h_y} = 1$, so the best choice for f_y, g_y, h_y gives the bound (14).

Remark 1. The model (16) uses generalized nearest-neighbor decoding, improving the rules proposed in [42–44]. The authors of [50] pointed out that (6) and (15) use the LMMSE and MMSE, respectively; see ([50], Equation (87)).

Remark 2. A corresponding forward model can be based on (8) and (13), namely

$$q(y|x)^s = \frac{q(x|y)}{p(x)} \Rightarrow q(y) = 1. \tag{19}$$

Remark 3. The RHS of (15) has a more complicated form than the RHS of (6) due to the outer expectation and conditional variance, and this makes optimizing X challenging when there is CSIR and CSIT. Also, if $p(y|x)$ is known, then it seems sensible to numerically compute $p(y)$ and $I(X; Y)$ directly, e.g., via Monte Carlo or numerical integration.

Remark 4. Decoding rules for discrete X can be based on decision theory as well as estimation theory; see ([64], Equation (11)).

Forward Models: Refinements of (11) appear in the optical fiber literature where the non-linear Schrödinger equation describes wave propagation [52]. Such channels exhibit complicated interactions of attenuation, dispersion, nonlinearity, and noise, and the channel density is too challenging to compute. One thus resorts to capacity lower bounds based on GMI and Monte Carlo simulation. The simplest models are memoryless, and they work well if chosen carefully. For example, the paper [52] used auxiliary models of the form

$$q(y|x) = \exp\left(-|y - hx|^2/\sigma_{|x|}^2\right) \tag{20}$$

where h accounts for attenuation and self-phase modulation, and where the noise variance $\sigma_{|x|}^2$ depends on $|x|$. Also, X was chosen to have concentric rings rather than a CSCG density. Subsequent papers applied progressively more sophisticated models with memory to better approximate the actual channel; see [53–59]. However, the rate gains over the model (20) are minor ($\approx 12\%$) for 1000 km links, and the newer models do not suggest practical receiver structures.

A related application is short-reach fiber-optic systems that use direct detection (DD) receivers [65] with photodiodes. The paper [60] showed that sampling faster than the symbol rate increases the DD capacity. However, spectrally efficient filtering gives the channel a long memory, motivating auxiliary models $q(y|x)$ with reduced memory to simplify GMI computations [61,66]. More generally, one may use channel-shortening filters [67–69] to increase the GMI.

Remark 5. The ultimate GMI is $I(X; Y)$, and one can compute this quantity numerically for the channels considered in this paper. We are motivated to focus on forward auxiliary models $q(y|x)$ to understand how to improve information rates for more complex channels. For instance, simple

$q(y|x)$ let one understand properties of optimal codes, see Lemma 3, and they suggest explicit power control policies, see Theorem 2.

Remark 6. The paper [37] (see also ([2], Equation (3.3.45)) and ([70], Equation (6))) derives two capacity lower bounds for massive MIMO channels. These bounds are designed for problems where the fading parameters have small variance so that, in effect, σ^2 in (7) is small. We will instead encounter cases where σ^2 grows in proportion to $E[|X|^2]$ and the RHS of (6) quickly saturates as $E[|X|^2]$ grows; see Remark 20.

1.5. Organization

This paper is organized as follows. Section 2 defines notation and reviews basic results. Section 3 develops two results for the GMI of scalar auxiliary models with AWGN:

- Proposition 1 in Section 3.1 states a known result, namely that the RHS of (6) is the maximum GMI for the AWGN auxiliary model (11) and a CSCG X .
- Lemma 1 in Section 3.2 generalizes Proposition 1 by partitioning the channel output alphabet into K subsets, $K \geq 1$. We use $K = 2$ to establish capacity properties at high and low SNR.

Sections 4 and 5 apply the GMI to channels with CSIT and CSIR.

- Section 4.3 treats adaptive codewords and develops structural properties of their optimal distribution.
- Lemma 2 in Section 4.4 generalizes Proposition 1 to MIMO channels and adaptive codewords. The receiver models each transmit symbol as a weighted sum of the entries of the corresponding adaptive symbol.
- Lemma 3 in Section 4.5 states that the maximum GMI for scalar channels, an AWGN auxiliary model, adaptive codewords with jointly CSCG entries, and $K = 1$ is achieved by using a conventional codebook where each symbol is modified based on the CSIT.
- Lemma 4 in Section 4.6 extends Lemma 3 to MIMO channels, including diagonal or parallel channels.
- Theorem 1 in Section 5.1 generalizes Lemma 3 to include CSIR; we use this result several times in Section 6.
- Lemma 5 in Section 5.3 generalizes Lemmas 1 and 2 by partitioning the channel output alphabet.

Sections 6–8 apply the GMI to fading channels with AWGN and illustrate the theory for on-off and Rayleigh fading.

- Lemma 6 in Section 6 gives a general capacity upper bound.
- Section 6.5 introduces a class of power control policies for full CSIT. Theorem 2 develops the optimal policy with an MMSE form.
- Theorem 3 in Section 6.6 provides a quadratic waterfilling expression for the GMI with partial CSIR.

Section 9 develops theory for block fading channels with in-block feedback (or in-block CSIT) that is a function of the CSIR and past channel inputs and outputs.

- Theorem 4 in Section 9.2 generalizes Lemma 4 to MIMO block fading channels;
- Section 9.3 develops capacity expressions in terms of directed information;
- Section 9.4 specializes the capacity to fading channels with AWGN and delayed CSIR;
- Proposition 3 generalizes Proposition 2 to channels with special CSIR and CSIT.

Section 10 concludes the paper. Finally, Appendices A–G provide results on special functions, GMI calculations, and proofs.

2. Preliminaries

2.1. Basic Notation

Let $1(\cdot)$ be the indicator function that takes on the value 1 if its argument is true and 0 otherwise. Let $\delta(\cdot)$ be the Dirac generalized function with $\int_{\mathcal{X}} \delta(x)f(x)dx = f(0) \cdot 1(0 \in \mathcal{X})$.

For $x \in \mathbb{R}$, define $(x)^+ = \max(0, x)$. The complex-conjugate, absolute value, and phase of $x \in \mathbb{C}$ are written as x^* , $|x|$, and $\arg(x)$, respectively. We write $j = \sqrt{-1}$ and $\bar{\epsilon} = 1 - \epsilon$.

Sets are written with calligraphic font, e.g., $\mathcal{S} = \{1, \dots, n\}$ and the cardinality of \mathcal{S} is $|\mathcal{S}|$. The complement of \mathcal{S} in \mathcal{T} is \mathcal{S}^c where \mathcal{T} is understood from the context.

2.2. Vectors and Matrices

Column vectors are written as $\underline{x} = [x_1, \dots, x_M]^T$ where M is the dimension, and T denotes transposition. The complex-conjugate transpose (or Hermitian) of \underline{x} is written as \underline{x}^\dagger . The Euclidean norm of \underline{x} is $\|\underline{x}\|$. Matrices are written with bold letters such as \mathbf{A} . The letter \mathbf{I} denotes the identity matrix. The determinant and trace of a square matrix \mathbf{A} are written as $\det \mathbf{A}$ and $\text{tr} \mathbf{A}$, respectively.

A singular value decomposition (SVD) is $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger$ where \mathbf{U} and \mathbf{V} are unitary matrices and $\mathbf{\Sigma}$ is a rectangular diagonal matrix with the singular values of \mathbf{A} on the diagonal. The square matrix \mathbf{A} is positive semi-definite if $\underline{x}^\dagger \mathbf{A} \underline{x} \geq 0$ for all \underline{x} . The notation $\mathbf{A} \preceq \mathbf{B}$ means that $\mathbf{B} - \mathbf{A}$ is positive semi-definite. Similarly, \mathbf{A} is positive definite if $\underline{x}^\dagger \mathbf{A} \underline{x} > 0$ for all \underline{x} , and we write $\mathbf{A} \prec \mathbf{B}$ if $\mathbf{B} - \mathbf{A}$ is positive definite.

2.3. Random Variables

Random variables are written with uppercase letters, such as X , and their realizations with lowercase letters, such as x . We write the distribution of discrete X with alphabet $\mathcal{X} = \{0, \dots, n - 1\}$ as $P_X = [P_X(0), \dots, P_X(n - 1)]$. The density of a real- or complex-valued X is written as p_X . Mixed discrete-continuous distributions are written using mixtures of densities and Dirac- δ functions.

Conditional distributions and densities are written as $P_{X|Y}$ and $p_{X|Y}$, respectively. We usually drop subscripts if the argument is a lowercase version of the random variable, e.g., we write $p(y|x)$ for $p_{Y|X}(y|x)$. One exception is that we consistently write the distributions $P_{S_R}(\cdot)$ and $P_{S_T}(\cdot)$ of the CSIR and CSIT with the subscript to avoid confusion with power notation.

2.4. Second-Order Statistics

The expectation and variance of the complex-valued random variable X are $E[X]$ and $\text{Var}[X] = E[|X - E[X]|^2]$, respectively. The correlation coefficient of X_1 and X_2 is $\rho = E[U_1 U_2^*]$ where

$$U_i = (X_i - E[X_i]) / \sqrt{\text{Var}[X_i]}$$

for $i = 1, 2$. We say that X_1 and X_2 are fully correlated if $\rho = e^{j\phi}$ for some real ϕ . Conditional expectation and variance are written as $E[X|A = a]$ and

$$\text{Var}[X|A = a] = E[(X - E[X])(X - E[X])^* | A = a].$$

The expressions $E[X|A]$, $\text{Var}[X|A]$ are random variables that take on the values $E[X|A = a]$, $\text{Var}[X|A = a]$ if $A = a$.

The expectation and covariance matrix of the random column vector $\underline{X} = [X_1, \dots, X_M]^T$ are $E[\underline{X}]$ and $\mathbf{Q}_{\underline{X}} = E[(\underline{X} - E[\underline{X}])(\underline{X} - E[\underline{X}])^\dagger]$, respectively. We write $\mathbf{Q}_{\underline{X}, \underline{Y}}$ for the covariance matrix of the stacked vector $[\underline{X}^T \underline{Y}^T]^T$. We write $\mathbf{Q}_{\underline{X}|\underline{Y}=y}$ for the covariance matrix of \underline{X} conditioned on the event $\underline{Y} = y$. $\mathbf{Q}_{\underline{X}|\underline{Y}}$ is a random matrix that takes on the matrix value $\mathbf{Q}_{\underline{X}|\underline{Y}=y}$ when $\underline{Y} = y$.

We often consider CSCG random variables and vectors. A CSCG \underline{X} has density

$$p(\underline{x}) = \frac{\exp(-\underline{x}^\dagger \mathbf{Q}_{\underline{X}}^{-1} \underline{x})}{\pi^M \det \mathbf{Q}_{\underline{X}}}$$

and we write $\underline{X} \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q}_{\underline{X}})$.

2.5. MMSE and LMMSE Estimation

Assume that $E[\underline{X}] = E[\underline{Y}] = \underline{0}$. The MMSE estimate of \underline{X} given the event $\underline{Y} = \underline{y}$ is the vector $\hat{\underline{X}}(\underline{y})$ that minimizes

$$E \left[\left\| \underline{X} - \hat{\underline{X}}(\underline{y}) \right\|^2 \middle| \underline{Y} = \underline{y} \right].$$

Direct analysis gives ([71], Chapter 4)

$$\hat{\underline{X}}(\underline{y}) = E \left[\underline{X} | \underline{Y} = \underline{y} \right] \tag{21}$$

$$E \left[\left\| \underline{X} - \hat{\underline{X}} \right\|^2 \right] = E \left[\left\| \underline{X} \right\|^2 \right] - E \left[\left\| \hat{\underline{X}} \right\|^2 \right] \tag{22}$$

$$\mathbf{Q}_{\underline{X} - \hat{\underline{X}}} = \mathbf{Q}_{\underline{X}} - \mathbf{Q}_{\hat{\underline{X}}} \tag{23}$$

$$E \left[(\underline{X} - \hat{\underline{X}}) \underline{Y}^\dagger \right] = \mathbf{0} \tag{24}$$

where the last identity is called the orthogonality principle.

The LMMSE estimate of \underline{X} given \underline{Y} with invertible $\mathbf{Q}_{\underline{Y}}$ is the vector $\hat{\underline{X}}_L = \mathbf{C} \underline{Y}$ where \mathbf{C} is chosen to minimize $E \left[\left\| \underline{X} - \hat{\underline{X}}_L \right\|^2 \right]$. We compute

$$\hat{\underline{X}}_L = E \left[\underline{X} \underline{Y}^\dagger \right] \mathbf{Q}_{\underline{Y}}^{-1} \underline{Y} \tag{25}$$

and we also have the properties (22)–(24) with $\hat{\underline{X}}$ replaced by $\hat{\underline{X}}_L$. Moreover, if \underline{X} and \underline{Y} are jointly CSCG, then the MMSE and LMMSE estimators coincide, and the orthogonality principle (24) implies that the error $\underline{X} - \hat{\underline{X}}$ is independent of \underline{Y} , i.e., we have

$$\begin{aligned} E \left[(\underline{X} - \hat{\underline{X}}) (\underline{X} - \hat{\underline{X}})^\dagger \middle| \underline{Y} = \underline{y} \right] &= E \left[\underline{X} \underline{X}^\dagger \middle| \underline{Y} = \underline{y} \right] - E \left[\underline{X} \underline{Y}^\dagger \right] \mathbf{Q}_{\underline{Y}}^{-1} \underline{y} \underline{y}^\dagger \mathbf{Q}_{\underline{Y}}^{-1} E \left[\underline{X} \underline{Y}^\dagger \right]^\dagger \\ &= \mathbf{Q}_{\underline{X}} - \mathbf{Q}_{\hat{\underline{X}}}. \end{aligned} \tag{26}$$

2.6. Entropy, Divergence, and Information

Entropies of random vectors with densities p are written as

$$h(\underline{X}) = E[-\log p(\underline{X})], \quad h(\underline{X}|\underline{Y}) = E[-\log p(\underline{X}|\underline{Y})]$$

where we use logarithms to the base e for analysis. The informational divergence of the densities p and q is

$$D(p||q) = E \left[\log \frac{p(\underline{X})}{q(\underline{X})} \right]$$

and $D(p||q) \geq 0$ with equality if and only if $p = q$ almost everywhere. The mutual information of \underline{X} and \underline{Y} is

$$I(\underline{X}; \underline{Y}) = D(p(\underline{X}, \underline{Y}) || p(\underline{X}) p(\underline{Y})) = E \left[\log \frac{p(\underline{Y}|\underline{X})}{p(\underline{Y})} \right].$$

The average mutual information of \underline{X} and \underline{Y} conditioned on \underline{Z} is $I(\underline{X}; \underline{Y}|\underline{Z})$. We write strings as $X^L = (X_1, X_2, \dots, X_L)$ and use the directed information notation (see [9,72])

$$I(X^L \rightarrow Y^L | Z) = \sum_{\ell=1}^L I(X^\ell; Y_\ell | Y^{\ell-1}, Z) \tag{27}$$

$$I(X^L \rightarrow Y^L || Z^L | W) = \sum_{\ell=1}^L I(X^\ell; Y_\ell | Y^{\ell-1}, Z^\ell, W) \tag{28}$$

where $Y_0 = 0$.

2.7. Entropy and Information Bounds

The expression (2) applies to random vectors. Choosing $q(x|y)$ as the conditional density where the $\underline{X}, \underline{Y}$ are modeled as jointly CSCG we obtain a generalization of (5):

$$h(\underline{X}|\underline{Y}) \leq \log \frac{\det(\pi e \mathbf{Q}_{\underline{X}|\underline{Y}})}{\det(\pi e \mathbf{Q}_{\underline{Y}})} = \log \det\left(\pi e \left\{ \mathbf{Q}_{\underline{X}} - \mathbb{E}[\underline{X}\underline{Y}^\dagger] \mathbf{Q}_{\underline{Y}}^{-1} \mathbb{E}[\underline{Y}\underline{X}^\dagger] \right\}\right). \tag{29}$$

The vector generalization of (6) for CSCG \underline{X} is

$$\begin{aligned} I(\underline{X}; \underline{Y}) &= h(\underline{X}) - h(\underline{X}|\underline{Y}) \\ &\geq \log \det\left(\left(\mathbf{Q}_{\underline{X}} - \mathbb{E}[\underline{X}\underline{Y}^\dagger] \mathbf{Q}_{\underline{Y}}^{-1} \mathbb{E}[\underline{Y}\underline{X}^\dagger]\right)^{-1} \mathbf{Q}_{\underline{X}}\right) \\ &\stackrel{(a)}{=} \log \det\left(\mathbf{I} + \mathbf{Q}_{\underline{Z}}^{-1} \mathbf{H} \mathbf{Q}_{\underline{X}}^{-1} \mathbf{H}^\dagger\right) \end{aligned} \tag{30}$$

where (cf. (7))

$$\mathbf{H} = \mathbb{E}[\underline{Y}\underline{X}^\dagger] \mathbf{Q}_{\underline{X}}^{-1}, \quad \mathbf{Q}_{\underline{Z}} = \mathbf{Q}_{\underline{Y}} - \mathbf{H} \mathbf{Q}_{\underline{X}} \mathbf{H}^\dagger \tag{31}$$

and step (a) in (30) follows by the Woodbury identity

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} (\mathbf{C}^{-1} + \mathbf{DA}^{-1} \mathbf{B})^{-1} \mathbf{DA}^{-1} \tag{32}$$

and the Sylvester identity

$$\det(\mathbf{I} + \mathbf{AB}) = \det(\mathbf{I} + \mathbf{BA}). \tag{33}$$

We also have vector generalizations of (14) and (15):

$$h(\underline{X}|\underline{Y}) \leq \mathbb{E} \left[\log \det(\pi e \mathbf{Q}_{\underline{X}|\underline{Y}}) \right] \tag{34}$$

$$I(\underline{X}; \underline{Y}) \geq \mathbb{E} \left[\log \frac{\det \mathbf{Q}_{\underline{X}}}{\det \mathbf{Q}_{\underline{X}|\underline{Y}}} \right] \text{ for CSCG } \underline{X}. \tag{35}$$

2.8. Capacity and Wideband Rates

Consider the complex-alphabet AWGN channel with output $Y = X + Z$ and noise $Z \sim \mathcal{CN}(0, 1)$. The capacity with the block power constraint $\frac{1}{n} \sum_{i=1}^n |X_i|^2 \leq P$ is

$$C(P) = \max_{\mathbb{E}[|X|^2] \leq P} I(X; Y) = \log(1 + P). \tag{36}$$

The low SNR regime (small P) is known as the wideband regime [73]. For well-behaved channels such as AWGN channels, the minimum E_b/N_0 and the slope S of the capacity vs. E_b/N_0 in bits/(3 dB) at the minimum E_b/N_0 are (see ([73], Equation (35)) and ([73], Theorem 9))

$$\left. \frac{E_b}{N_0} \right|_{\min} = \frac{\log 2}{C'(0)'}, \quad S = \frac{2[C'(0)]^2}{-C''(0)} \tag{37}$$

where $C'(P)$ and $C''(P)$ are the first and second derivatives of $C(P)$ (measured in nats) with respect to P , respectively. For example, the wideband derivatives for (36) are $C'(0) = 1$ and $C''(0) = -1$ so that the wideband values (37) are

$$\left. \frac{E_b}{N_0} \right|_{\min} = \log 2, \quad S = 2. \tag{38}$$

The minimal E_b/N_0 is usually stated in decibels, for example $10 \log_{10}(\log 2) = -1.59$ dB. An extension of the theory to general channels is described in ([74], Section III).

Remark 7. A useful method is flash signaling, where one sends with zero energy most of the time. In particular, we will consider the CSCG flash density

$$p(x) = (1 - p) \delta(x) + p \frac{e^{-|x|^2/(P/p)}}{\pi(P/p)} \tag{39}$$

where $0 < p \leq 1$ so that the average power is $E[|X|^2] = P$. Note that flash signaling is defined in ([73], Definition 2) as a family of distributions satisfying a particular property as $P \rightarrow 0$. We use the terminology informally.

2.9. Uniformly-Spaced Quantizer

Consider a uniformly-spaced scalar quantizer $q_u(\cdot)$ with B bits, domain $[0, \infty)$, and reconstruction points

$$s \in \{\Delta/2, 3\Delta/2, \dots, \Delta/2 + (2^B - 1)\Delta\}$$

where $\Delta > 0$. The quantization intervals are

$$\mathcal{I}(s) = \begin{cases} \left[s - \frac{\Delta}{2}, s + \frac{\Delta}{2} \right), & s \neq s_{\max} \\ \left[s - \frac{\Delta}{2}, \infty \right), & s = s_{\max} \end{cases}$$

where $s_{\max} = \Delta/2 + (2^B - 1)\Delta$. We will consider $B = 0, 1, \infty$. For $B = \infty$ we choose $q_u(x) = x$.

Suppose one applies the quantizer to the non-negative random variable G with density $p(g)$ to obtain $S_T = q_u(G)$. Let P_{S_T} and $P_{S_T|G}$ be the probability mass functions of S_T without and with conditioning on G , respectively. We have

$$P_{S_T|G}(s|g) = 1(g \in \mathcal{I}(s)), \quad P_{S_T}(s) = \int_{g \in \mathcal{I}(s)} p(g) dg \tag{40}$$

and using Bayes' rule, we obtain

$$p(g|s) = \begin{cases} p(g)/P_{S_T}(s), & g \in \mathcal{I}(s) \\ 0, & \text{else.} \end{cases} \tag{41}$$

3. Generalized Mutual Information

We re-derive the GMI in the usual way, where one starts with the forward model $q(y|x)$ rather than the reverse density $q(x|y)$ in (8). Consider the joint density $p(x, y)$ and define $q(y)$ as in (9) for $s \geq 0$. Note that neither $q(y|x)$ nor $q(y)$ must be densities. The GMI is defined in [39] to be $\max_{s \geq 0} I_s(X; Y)$ where (see the RHS of (10))

$$I_s(X; Y) = E \left[\log \frac{q(Y|X)^s}{q(Y)} \right] \tag{42}$$

and where the expectation is with respect to $p(x, y)$. The GMI is a lower bound on the mutual information since

$$I_s(X; Y) = I(X; Y) - D(p_{X,Y} \| p_Y q_{X|Y}). \tag{43}$$

Moreover, by using Gallager's derivation of error exponents, but without modifying his "s" variable, the GMI $I_s(X; Y)$ is achievable with a mismatched decoder that uses $q(y|x)$ for its decoding metric [39].

3.1. AWGN Forward Model with CSCG Inputs

A natural metric is based on the AWGN auxiliary channel $Y_a = hX + Z$ where h is a channel parameter and $Z \sim \mathcal{CN}(0, \sigma^2)$ is independent of X , i.e., we have the auxiliary model (here a density)

$$q(y|x) = \frac{1}{\pi\sigma^2} \exp\left(-|y - hx|^2/\sigma^2\right) \tag{44}$$

where h and σ^2 are to be optimized. A natural input is $X \sim \mathcal{CN}(0, P)$ so that (9) is

$$q(y) = \frac{\pi\sigma^2/s}{(\pi\sigma^2)^s} \cdot \frac{\exp\left(\frac{-|y|^2}{\sigma^2/s + |h|^2P}\right)}{\pi(\sigma^2/s + |h|^2P)}. \tag{45}$$

We have the following result, see [43] that considers channels of the form (1) and ([47], Proposition 1) that considers general $p(y|x)$.

Proposition 1. *The maximum GMI (42) for the channel $p(y|x)$, a CSCG input X with variance $P > 0$, and the auxiliary model (44) with $\sigma^2 > 0$ is*

$$I_1(X; Y) = \log\left(1 + \frac{|\tilde{h}|^2P}{\tilde{\sigma}^2}\right) \tag{46}$$

where $s = 1$ and (cf. (7))

$$\tilde{h} = \mathbb{E}[YX^*]/P \tag{47}$$

$$\tilde{\sigma}^2 = \mathbb{E}[|Y - \tilde{h}X|^2] = \mathbb{E}[|Y|^2] - |\tilde{h}|^2P. \tag{48}$$

The expectations are with respect to the actual density $p(x, y)$.

Proof. The GMI (42) for the model (44) is

$$I_s(X; Y) = \log\left(1 + \frac{|h|^2P}{\sigma^2/s}\right) + \frac{\mathbb{E}[|Y|^2]}{\sigma^2/s + |h|^2P} - \frac{\mathbb{E}[|Y - hX|^2]}{\sigma^2/s}. \tag{49}$$

Since (49) depends only on the ratio σ^2/s one may as well set $s = 1$. Thus, choosing $h = \tilde{h}$ and $\sigma^2 = \tilde{\sigma}^2$ gives (46).

Next, consider $Y_a = \tilde{h}X + \tilde{Z}$ where $\tilde{Z} \sim \mathcal{CN}(0, \tilde{\sigma}^2)$ is independent of X . We have

$$\mathbb{E}[|Y_a|^2] = \mathbb{E}[|Y|^2] \tag{50}$$

$$\mathbb{E}[|Y_a - \tilde{h}X|^2] = \mathbb{E}[|Y - \tilde{h}X|^2]. \tag{51}$$

In other words, the second-order statistics for the two channels with outputs Y (the actual channel output) and Y_a are the same. But the GMI (46) is the mutual information $I(X; Y_a)$. Using (43) and (49), for any s, h and σ^2 we have

$$I(X; Y_a) = \log\left(1 + \frac{|\tilde{h}|^2P}{\tilde{\sigma}^2}\right) \geq I_s(X; Y_a) = I_s(X; Y) \tag{52}$$

and equality holds if $h = \tilde{h}$ and $\sigma^2/s = \tilde{\sigma}^2$. \square

Remark 8. *The rate (46) is the same as the RHS of (6).*

Remark 9. *Proposition 1 generalizes to vector models and adaptive input symbols; see Section 4.4.*

Remark 10. The estimate \tilde{h} is the MMSE estimate of h :

$$\tilde{h} = \arg \min_h \mathbb{E} \left[|Y - hX|^2 \right] \tag{53}$$

and $\tilde{\sigma}^2$ is the variance of the error. To see this, expand

$$\mathbb{E} \left[|Y - hX|^2 \right] = \mathbb{E} \left[|(Y - \tilde{h}X) + (\tilde{h} - h)X|^2 \right] = \tilde{\sigma}^2 + |\tilde{h} - h|^2 P \tag{54}$$

where the final step follows by the definition of \tilde{h} in (47).

Remark 11. Suppose that h is an estimate other than (53). Then if $\mathbb{E} [|Y|^2] > \mathbb{E} [|Y - hX|^2]$ we may choose

$$\sigma^2/s = |h|^2 P \cdot \frac{\mathbb{E} [|Y - hX|^2]}{\mathbb{E} [|Y|^2] - \mathbb{E} [|Y - hX|^2]} \tag{55}$$

and the GMI (49) simplifies to

$$I_s(X; Y) = \log \left(\frac{\mathbb{E} [|Y|^2]}{\mathbb{E} [|Y - hX|^2]} \right). \tag{56}$$

Remark 12. The LM rate (for “lower bound to the mismatch capacity”) improves the GMI for some $q(y|x)$ [40,75]. The LM rate replaces $q(y|x)$ with $q(y|x)e^{t(x)/s}$ for some function $t(\cdot)$ and permits optimizing s and $t(\cdot)$; see ([41], Section 2.3.2). For example, if $p(y|x)$ has the form $q(y|x)s e^{t(x)}$ then the LM rate can be larger than the GMI; see [76,77].

3.2. CSIR and K-Partitions

We consider two generalizations of Proposition 1. The first is for channels with a state S_R known at the receiver but not at the transmitter. The second expands the class of CSCG auxiliary models. The motivation is to obtain more precise models under partial CSIR, especially to better deal with channels at high SNR and with high rates. We here consider discrete S_R and later extend to continuous S_R .

CSIR: Consider the average GMI

$$I_1(X; Y|S_R) = \sum_{s_R} P_{S_R}(s_R) I_1(X; Y|S_R = s_R) \tag{57}$$

where $I_1(X; Y|S_R = s_R)$ is the usual GMI where all densities are conditioned on $S_R = s_R$. The parameters (47) and (48) for the event $S_R = s_R$ are now

$$\tilde{h}(s_R) = \frac{\mathbb{E} [YX^* | S_R = s_R]}{\mathbb{E} [|X|^2 | S_R = s_R]} \tag{58}$$

$$\tilde{\sigma}^2(s_R) = \mathbb{E} \left[|Y - \tilde{h}(s_R)X|^2 | S_R = s_R \right]. \tag{59}$$

The GMI (57) is thus

$$I_1(X; Y|S_R) = \sum_{s_R} P_{S_R}(s_R) \log \left(1 + \frac{|\tilde{h}(s_R)|^2 P}{\tilde{\sigma}^2(s_R)} \right). \tag{60}$$

K-Partitions: Let $\{\mathcal{Y}_k : k = 1, \dots, K\}$ be a K-partition of \mathcal{Y} and define the auxiliary model

$$q(y|x) = \frac{1}{\pi\sigma_k^2} e^{-|y-h_kx|^2/\sigma_k^2}, \quad y \in \mathcal{Y}_k. \tag{61}$$

Observe that $q(y|x)$ is not necessarily a density. We choose $X \sim \mathcal{CN}(0, P)$ so that (9) becomes (cf. (45))

$$q(y) = \frac{\pi\sigma_k^2/s}{(\pi\sigma_k^2)^s} \cdot \frac{\exp\left(\frac{-|y|^2}{\sigma_k^2/s + |h_k|^2P}\right)}{\pi(\sigma_k^2/s + |h_k|^2P)}, \quad y \in \mathcal{Y}_k. \tag{62}$$

Define the events $\mathcal{E}_k = \{Y \in \mathcal{Y}_k\}$ for $k = 1, \dots, K$. We have

$$I_s(X; Y) = \sum_{k=1}^K \Pr[\mathcal{E}_k] \cdot \mathbb{E}\left[\log \frac{q(Y|X)^s}{q(Y)} \middle| \mathcal{E}_k\right] \tag{63}$$

and inserting (61) and (62) we have the following lemma.

Lemma 1. *The GMI (42) for the channel $p(y|x)$, $s = 1$, a CSCG input X with variance P , and the auxiliary model (61) is (see (49))*

$$I_1(X; Y) = \sum_{k=1}^K \Pr[\mathcal{E}_k] \left[\log\left(1 + \frac{|h_k|^2P}{\sigma_k^2}\right) + \frac{\mathbb{E}[|Y|^2|\mathcal{E}_k]}{\sigma_k^2 + |h_k|^2P} - \frac{\mathbb{E}[|Y - h_kX|^2|\mathcal{E}_k]}{\sigma_k^2} \right]. \tag{64}$$

Remark 13. *K-partitioning formally includes (57) as a special case by including S_R as part of the receiver’s “overall” channel output $\tilde{Y} = [Y, S_R]$. For example, one can partition $\tilde{\mathcal{Y}}$ as $\{\tilde{\mathcal{Y}}_{s_R} : s_R \in \mathcal{S}_R\}$ where $\tilde{\mathcal{Y}}_{s_R} = \mathcal{Y} \times \{s_R\}$.*

Remark 14. *The models (16) and (61) suggest building receivers based on adaptive Gaussian statistics. However, we are motivated to introduce (61) to prove capacity scaling results. For this purpose, we will use $K = 2$ with the partition*

$$\mathcal{E}_1 = \{|Y|^2 < t_R\}, \quad \mathcal{E}_2 = \{|Y|^2 \geq t_R\} \tag{65}$$

and $h_1 = 0, \sigma_1^2 = 1$. The GMI (64) thus has only the $k = 2$ term and it remains to choose h_2, σ_2^2 , and t_R .

Remark 15. *One can generalize Lemma 1 and partition $\mathcal{X} \times \mathcal{Y}$ rather than \mathcal{Y} only. However, the $q(y)$ in (62) might not have a CSCG form.*

Remark 16. *Define $P_k = \mathbb{E}[|X|^2|\mathcal{E}_k]$ and choose the LMMSE auxiliary models with*

$$h_k = \mathbb{E}[YX^*|\mathcal{E}_k]/P_k \tag{66}$$

$$\sigma_k^2 = \mathbb{E}[|Y - h_kX|^2|\mathcal{E}_k] = \mathbb{E}[|Y|^2|\mathcal{E}_k] - |h_k|^2P_k \tag{67}$$

for $k = 1, \dots, K$. The expression (64) is then

$$I_1(X; Y) = \sum_{k=1}^K \Pr[\mathcal{E}_k] \left[\log\left(1 + \frac{|h_k|^2P}{\mathbb{E}[|Y|^2|\mathcal{E}_k] - |h_k|^2P_k}\right) - \frac{|h_k|^2(P - P_k)}{\mathbb{E}[|Y|^2|\mathcal{E}_k] + |h_k|^2(P - P_k)} \right]. \tag{68}$$

Remark 17. *The LMMSE-based GMI (68) reduces to the GMI of Proposition 1 by choosing the trivial partition with $K = 1$ and $\mathcal{Y}_1 = \mathcal{Y}$. However, the GMI (68) may not be optimal for $K \geq 2$. What can be said is that the phase of h_k in (64) should be the same as the phase of $\mathbb{E}[YX^*|\mathcal{E}_k]$ for all k . We thus have K two-dimensional optimization problems, one for each pair $(|h_k|, \sigma_k^2)$, $k = 1, \dots, K$.*

Remark 18. *Suppose we choose a different auxiliary model for each $Y = y$, i.e., consider $K \rightarrow \infty$. The reverse density GMI uses the auxiliary model (19) which gives the RHS of (15):*

$$I_1(X; Y) = \int_{\mathbb{C}} p(y) \log \frac{P}{\text{Var}[X|Y=y]} dy. \tag{69}$$

Instead, the suboptimal (68) is the complicated expression

$$I_1(X; Y) = \int_{\mathbb{C}} p(y) \left[\log \left(1 + \frac{|\mathbb{E}[X|Y=y]|^2(P/P_y)}{\text{Var}[X|Y=y]} \right) - \frac{|\mathbb{E}[X|Y=y]|^2(P/P_y - 1)}{\text{Var}[X|Y=y] + |\mathbb{E}[X|Y=y]|^2(P/P_y)} \right] dy. \tag{70}$$

where $P_y = \mathbb{E}[|X|^2|Y=y]$. We show how to compute these GMIs in Appendix C.

3.3. Example: On-Off Fading

Consider the channel $Y = HX + Z$ where H, X, Z are mutually independent, $P_H(0) = P_H(\sqrt{2}) = 1/2$, and $Z \sim \mathcal{CN}(0, 1)$. The channel exhibits particularly simple fading, giving basic insight into more realistic fading models. We consider two basic scenarios: full CSIR and no CSIR.

Full CSIR: Suppose $S_R = H$ and

$$q(y|x, h) = p(y|x, h) = \frac{1}{\pi\sigma^2} e^{-|y-hx|^2/\sigma^2} \tag{71}$$

which corresponds to having (58) and (59) as

$$\tilde{h}(0) = 0, \quad \tilde{h}(\sqrt{2}) = \sqrt{2}, \quad \tilde{\sigma}^2(0) = \sigma^2(\sqrt{2}) = 1. \tag{72}$$

The GMI (60) with $X \sim \mathcal{CN}(0, P)$ thus gives the capacity

$$C(P) = \frac{1}{2} \log(1 + 2P). \tag{73}$$

The wideband values (37) are

$$\left. \frac{E_b}{N_0} \right|_{\min} = \log 2, \quad S = 1. \tag{74}$$

Compared with (38), the minimal E_b/N_0 is the same as without fading, namely -1.59 dB. However, fading reduces the capacity slope S ; see the dashed curve in Figure 1.

No CSIR: Suppose $S_R = 0$ and $X \sim \mathcal{CN}(0, P)$ and consider the densities

$$p(y|x) = \frac{e^{-|y|^2}}{2\pi} + \frac{e^{-|y-\sqrt{2}x|^2}}{2\pi} \tag{75}$$

$$p(y) = \frac{e^{-|y|^2}}{2\pi} + \frac{e^{-|y|^2/(1+2P)}}{2\pi(1+2P)}. \tag{76}$$

The mutual information can be computed by numerical integration or by Monte Carlo integration:

$$I(X; Y) \approx \frac{1}{N} \sum_{i=1}^N \log \frac{p_{Y|X}(y_i|x_i)}{p_Y(y_i)} \tag{77}$$

where the RHS of (77) converges to $I(X; Y)$ for long strings x^N, y^N sampled from $p(x, y)$. The results for $X \sim \mathcal{CN}(0, P)$ are shown in Figure 1 as the curve labeled “ $I(X; Y)$ Gauss”.

Next, Proposition 1 gives $h = 1/\sqrt{2}$, $\sigma^2 = 1 + P/2$, and

$$I_1(X; Y) = \log\left(1 + \frac{P}{2 + P}\right). \tag{78}$$

The wideband values (37) are

$$\frac{E_b}{N_0}\Big|_{\min} = \log 4, \quad S = 2/3 \tag{79}$$

so the minimal E_b/N_0 is 1.42 dB and the capacity slope S has decreased further. Moreover, the rate saturates at large SNR at 1 bit per channel use.

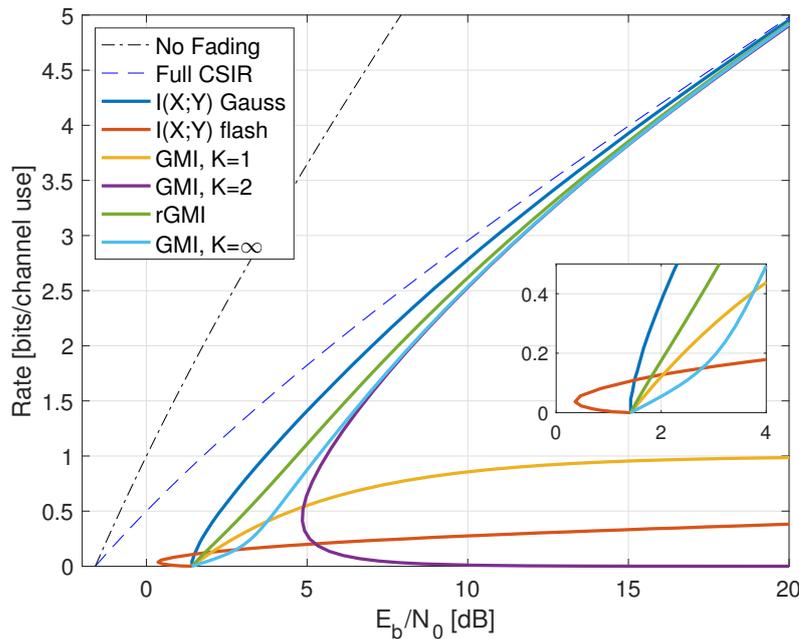


Figure 1. Rates for on-off fading with $S_R = 0$. The curve “Full CSIR” refers to $S_R = H$ and is a capacity upper bound. Flash signaling uses $p = 0.05$; the GMI for the $K = 2$ partition uses the threshold $t_R = P^{0.4} + 3$.

The “ $I(X; Y)$ Gauss” curve in Figure 1 suggests that the no-CSIR capacity approaches the full-CSIR capacity for large SNR. To prove this, consider the $K = 2$ partition specified in Remark 14 with $h_1 = 0$, $h_2 = \sqrt{2}$, and $\sigma_2^2 = 1$. Since we are not using LMMSE auxiliary models, we must compute the GMI using the general expression (64), which is

$$I_1(X; Y) = \Pr[\mathcal{E}_2] \left[\log(1 + 2P) + \frac{E[|Y|^2|\mathcal{E}_2]}{1 + 2P} - E\left[|Y - \sqrt{2}X|^2|\mathcal{E}_2\right] \right]. \tag{80}$$

In Appendix B.1, we show that choosing $t_R = P^{\lambda_R} + b$ where $0 < \lambda_R < 1$ and b is a real constant makes all terms behave as desired as P increases:

$$\Pr[\mathcal{E}_2] \rightarrow 1/2, \quad \frac{E[|Y|^2|\mathcal{E}_2]}{1 + 2P} \rightarrow 1, \quad E\left[|Y - \sqrt{2}X|^2|\mathcal{E}_2\right] \rightarrow 1. \tag{81}$$

The GMI (80) of Lemma 1 thus gives the maximal value (73) for large P :

$$\lim_{P \rightarrow \infty} \left[\frac{1}{2} \log(1 + 2P) - I_1(X; Y) \right] = 0. \tag{82}$$

Figure 1 shows the behavior of $I_1(X; Y)$ for $K = 2$, $\lambda_R = 0.4$, and $b = 3$. Effectively, at large SNR, the receiver can estimate H accurately, and one approaches the full-CSIR capacity.

Remark 19. For on-off fading, one may compute $I(X; Y)$ directly and use the densities (75) and (76) to decode. Nevertheless, the partitioning of Lemma 1 helps prove the capacity scaling (82).

Consider next the reverse density GMI (69) and the forward model GMI (70). Appendix C.1 shows how to compute $E[X|Y = y]$, $E[|X|^2|Y = y]$, and $\text{Var}[X|Y = y]$, and Figure 1 plots the GMIs as the curves labeled “rGMI” and “GMI, $K = \infty$ ”, respectively. The rGMI curve gives the best possible rates for AWGN auxiliary models, as shown in Section 1.4. The results also show that the large- K GMI (70) is worse than the $K = 1$ GMI at low SNR but better than the $K = 2$ GMI of Remark 14.

Finally, the curve labeled “ $I(X; Y)$ Gauss” in Figure 1 suggests that the minimal E_b/N_0 is 1.42 dB even for the capacity-achieving distribution. However, we know from ([73], Theorem 1) that flash signaling (39) can approach the minimal E_b/N_0 of -1.59 dB. For example, the flash rates $I(X; Y)$ with $p = 0.05$ are plotted in Figure 1. Unfortunately, the wideband slope is $S = 0$ ([73], Theorem 17), and one requires very large flash powers (very small p) to approach -1.59 dB.

Remark 20. As stated in Remark 6, the paper [37] (see also [2,70]) derives two capacity lower bounds. These bounds are the same for our problem, and they are derived using the following steps (see ([37], Lemmas 3 and 4)):

$$\begin{aligned} I(X; Y) &= I(X, S_H; Y) - I(S_H; Y|X) \\ &\geq I(X; Y|S_H) - I(S_H; Y|X). \end{aligned} \tag{83}$$

Now consider $Y = HX + Z$ where H, X, Z are mutually independent, $S_H = H$, $\text{Var}[Z] = 1$, and $X \sim \mathcal{CN}(0, P)$. We have

$$I(X; Y|H) \geq E[\log(1 + |H|^2P)] \tag{84}$$

$$\begin{aligned} I(H; Y|X) &= h(Y|X) - h(Z) \\ &\leq \log(\pi e(1 + \text{Var}[H]P)) - h(Z) \end{aligned} \tag{85}$$

where (84) and (85) follow by (5), in the latter case with the roles of X and Y reversed. The bound (85) works well if $\text{Var}[H]$ is small, as for massive MIMO with “channel hardening”. However, for our on-off fading model, the bound (83) is

$$\begin{aligned} I(X; Y) &\geq E[\log(1 + |H|^2P)] - \log(1 + \text{Var}[H]P) \\ &= \frac{1}{2} \log(1 + 2P) - \log(1 + P/2) \end{aligned} \tag{86}$$

which is worse than the $K = 1$ and $K = \infty$ GMIs and is not shown in Figure 1.

4. Channels with CSIT

This section studies Shannon’s channel with side information, or state, known causally at the transmitter [5,6]. We begin by treating general channels and then focus mainly on complex-alphabet channels. The capacity expression has a random variable A that is either a list (for discrete-alphabet states) or a function (for continuous-alphabet states). We refer to A as an adaptive symbol of an adaptive codeword.

4.1. Model

The problem is specified by the functional dependence graph (FDG) in Figure 2. The model has a message M , a CSIT string S_T^n , and a noise string Z^n . The variables M, S_T^n, Z^n are mutually statistically independent, and S_T^n and Z^n are strings of i.i.d. random

variables with the same distributions as S_T and Z , respectively. S_T^n is available causally at the transmitter, i.e., the channel input $X_i, i = 1, \dots, n$, is a function of M and the sub-string S_T^i . The receiver sees the channel outputs

$$Y_i = f(X_i, S_{Ti}, Z_i) \tag{87}$$

for some function $f(\cdot)$ and $i = 1, 2, \dots, n$.

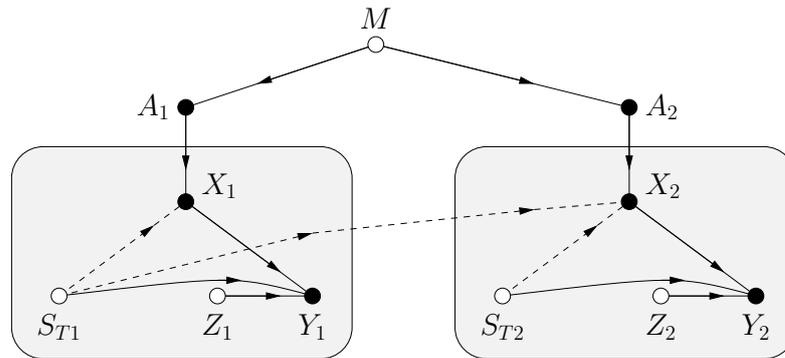


Figure 2. FDG for $n = 2$ uses of a channel with CSIT. Open nodes represent statistically independent random variables, and filled nodes represent random variables that are functions of their parent variables. Dashed lines represent the CSIT influence on X^n .

Each A_i represents a list of possible choices of X_i at time i . More precisely, suppose that S_T has alphabet $\mathcal{S}_T = \{0, 1, \dots, \nu - 1\}$ and define the adaptive symbol

$$A = (X(0), \dots, X(\nu - 1))$$

whose entries have alphabet \mathcal{X} . Here $S_T = s_T$ means that $X(s_T)$ is transmitted, i.e., we have $X = X(S_T)$. If S_T has a continuous alphabet, we make A a function rather than a list, and we may again write $X = X(S_T)$. Some authors therefore write A as $X(\cdot)$. (Shannon in [6] denoted our A and X as the respective X and x .)

Remark 21. The conventional choice for A if $\mathcal{X} = \mathbb{C}$ is

$$A = \left(\sqrt{P(0)} e^{j\phi(0)}, \dots, \sqrt{P(\nu - 1)} e^{j\phi(\nu - 1)} \right) \cdot U \tag{88}$$

where U has $E[|U|^2] = 1$, $P(s_T) = E[|X(s_T)|^2]$, and $\phi(s_T)$ is a phase shift. The interpretation is that U represents the symbol of a conventional codebook without CSIT, and these symbols are scaled and rotated. In other words, one separates the message-carrying U from an adaptation due to S_T via

$$X = \sqrt{P(S_T)} e^{j\phi(S_T)} U. \tag{89}$$

Remark 22. One may define the channel by the functional relation (87), by $p(y|a)$, or by $p(y|x, s_T)$; see Shannon’s emphasis in ([6], Theorem); see ([22], Remark 3). We generally prefer to use $p(y|a)$ since we interpret A as a channel input.

Remark 23. One can add feedback and let X_i be a function of (M, S_T^i, Y^{i-1}) , but feedback does not increase the capacity if the state and noise processes are memoryless ([22], Section V).

Remark 24. The model (87) permits block fading and MIMO transmission by choosing X_i and Y_i as vectors [11,78].

4.2. Capacity

The capacity of the model under study is (see [6])

$$C = \max_A I(A; Y) \tag{90}$$

where $A - [S_T, X] - Y$ forms a Markov chain. One may limit attention to A with cardinality $|\mathcal{A}|$ satisfying (see ([22], Equation (56)), [79], ([80], Theorem 1))

$$|\mathcal{A}| \leq \min(|\mathcal{Y}|, 1 + |S_T|(|\mathcal{X}| - 1)). \tag{91}$$

As usual, for the cost function $c(x, y)$ and the average block cost constraint

$$\frac{1}{n} \sum_{i=1}^n E[c(X_i, Y_i)] \leq P \tag{92}$$

the unconstrained maximization in (90) becomes a constrained maximization over the A for which $E[c(X, Y)] \leq P$. Also, a simple upper bound on the capacity is

$$\begin{aligned} C(P) &\leq \max_{A: E[c(X, Y)] \leq P} I(A; Y, S_T) \\ &\stackrel{(a)}{=} \max_{X(S_T): E[c(X(S_T), Y)] \leq P} I(X; Y | S_T) \end{aligned} \tag{93}$$

where step (a) follows by the independence of A and S_T . This bound is tight if the receiver knows S_T .

Remark 25. The chain rule for mutual information gives

$$I(A; Y) = I(X(0) \dots X(v - 1); Y) \tag{94}$$

$$= \sum_{s_T=0}^{v-1} I(X(s_T); Y | X(0), \dots, X(s_T - 1)). \tag{95}$$

The RHS of (94) suggests treating the channel as a multi-input, single-output (MISO) channel, and the expression (95) suggests using multi-level coding with multi-stage decoding [81]. For example, one may use polar coded modulation [82–84] with Honda-Yamamoto shaping [85,86].

Remark 26. For $\mathcal{X} = \mathbb{C}$ and the conventional adaptive symbol (88), we compute $I(A; Y) = I(U; Y)$ and

$$C(P) = \max_{P(S_T), \phi(S_T): E[c(X(S_T), Y)] \leq P} I(U; Y). \tag{96}$$

4.3. Structure of the Optimal Input Distribution

Let \mathcal{A} be the alphabet of A and let $\mathcal{X} = \mathbb{C}$, i.e., we have $\mathcal{A} = \mathbb{C}^v$ for discrete S_T . Consider the expansions

$$p(y|a) = \sum_{s_T} P_{S_T}(s_T) p(y|x(s_T), s_T) \tag{97}$$

$$\begin{aligned} p(y) &= \int_{\mathcal{A}} p(a) p(y|a) da \\ &= \sum_{s_T} P_{S_T}(s_T) \int_{\mathbb{C}} p(x(s_T)) p(y|x(s_T), s_T) dx(s_T). \end{aligned} \tag{98}$$

Observe that $p(y)$, and hence $h(Y)$, depends only on the marginals $p(x(s_T))$ of A ; see ([80], Section III). So define the set of densities having the same marginals as A :

$$\mathcal{P}(A) = \{p(\tilde{a}) : p(\tilde{x}(s_T)) = p(x(s_T)) \text{ for all } s_T \in \mathcal{S}_T\}.$$

This set is convex, since for any $p^{(1)}(a), p^{(2)}(a) \in \mathcal{P}(A)$ and $0 \leq \lambda \leq 1$ we have

$$\lambda p^{(1)}(a) + (1 - \lambda)p^{(2)}(a) \in \mathcal{P}(A). \tag{99}$$

Moreover, for fixed $p(y)$, the expression $I(A; Y)$ is a convex function of $p(a|y)$, and $p(a|y) = p(a)p(y|a)/p(y)$ is a linear function of $p(a)$. Maximizing $I(A; Y)$ over $\mathcal{P}(A)$ is thus the same as minimizing the concave function $h(Y|A)$ over the convex set $\mathcal{P}(A)$. An optimal $p(a)$ is thus an extreme of $\mathcal{P}(A)$. Some properties of such extremes are developed in [87,88].

For example, consider $|\mathcal{S}_T| = 2$ and $\mathcal{X} = \mathcal{S}_T = \{0, 1\}$, for which (91) states that at most $|\mathcal{A}| = 3$ adaptive symbols need have positive probability (and at most $|\mathcal{A}| = 2$ adaptive symbols if $|\mathcal{Y}| = 2$). Suppose the marginals have $P_{X(0)}(0) = 1/2, P_{X(1)}(0) = 3/4$ and consider the matrix notation

$$P_A = \begin{bmatrix} P_A(0,0) & P_A(0,1) \\ P_A(1,0) & P_A(1,1) \end{bmatrix}$$

where we write $P_A(x_1, x_2)$ for $P_A([x_1, x_2])$. The optimal P_A must then be one of the two extremes

$$P_A = \begin{bmatrix} 1/2 & 0 \\ 1/4 & 1/4 \end{bmatrix}, \quad P_A = \begin{bmatrix} 1/4 & 1/4 \\ 1/2 & 0 \end{bmatrix}. \tag{100}$$

For the first P_A , the codebook has the property that if $X(0) = 0$ then $X(1) = 0$ while if $X(0) = 1$ then $X(1)$ is uniformly distributed over $\mathcal{X} = \{0, 1\}$.

Next, consider $|\mathcal{S}_T| = 2$ and marginals $P_{X(0)}, P_{X(1)}$ that are uniform over $\mathcal{X} = \{0, 1, \dots, |\mathcal{X}| - 1\}$. This case was treated in detail in ([80], Section VI.A), see also [89], and we provide a different perspective. A classic theorem of Birkhoff [90] ensures that the extremes of $\mathcal{P}(A)$ are the $|\mathcal{X}|!$ distributions P_A for which the $|\mathcal{X}| \times |\mathcal{X}|$ matrix

$$P_A = \begin{bmatrix} P_A(0,0) & \dots & P_A(0,|\mathcal{X}| - 1) \\ \vdots & \ddots & \vdots \\ P_A(|\mathcal{X}| - 1,0) & \dots & P_A(|\mathcal{X}| - 1,|\mathcal{X}| - 1) \end{bmatrix}.$$

is a permutation matrix multiplied by $1/|\mathcal{X}|$. For example, for $|\mathcal{X}| = 2$ we have the two extremes

$$P_A = \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad P_A = \frac{1}{2} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{101}$$

The permutation property means that $X(s_T)$ is a function of $X(0)$, i.e., the encoding simplifies to a conventional codebook as in Remark 21 with uniformly-distributed U and a permutation $\pi_{s_T}(\cdot)$ indexed by s_T such that $X(s_T) = \pi_{s_T}(U)$. For example, for the first P_A in (101) we may choose $X(s_T) = U$, which is independent of s_T . On the other hand, for the second P_A in (101) we may choose $X(s_T) = U \oplus s_T$ where \oplus denotes addition modulo-2.

For $|\mathcal{S}_T| > 2$, the geometry of $\mathcal{P}(A)$ is more complicated; see ([80], Section VI.B). For example, consider $\mathcal{X} = \{0, 1\}$ and suppose the marginals $P_{X(s_T)}, s_T \in \mathcal{S}_T$, are all uniform. Then the extremes include P_A related to linear codes and their cosets, e.g., two extremes for $|\mathcal{S}_T| = 3$ are related to the repetition code and single parity check code:

$$P_A(a) = 1/2, a \in \{[0, 0, 0], [1, 1, 1]\}$$

$$P_A(a) = 1/4, a \in \{[0, 0, 0], [0, 1, 1], [1, 0, 1], [1, 1, 0]\}.$$

This observation motivates concatenated coding, where the message is first encoded by an outer encoder followed by an inner code that is the coset of a linear code. The transmitter then sends the entries at position S_T of the inner codewords, which are vectors of dimension $|S_T|$. We do not know if there are channels for which such codes are helpful.

4.4. Generalized Mutual Information

Consider the vector channel $p(\underline{y}|\underline{x})$ with input set $\mathcal{X} = \mathbb{C}^M$ and output set $\mathcal{Y} = \mathbb{C}^N$. The GMI for adaptive symbols is $\max_{s \geq 0} I_s(A; \underline{Y})$ where

$$I_s(A; \underline{Y}) = \mathbb{E} \left[\log \frac{q(\underline{Y}|A)^s}{q(\underline{Y})} \right] \tag{102}$$

and the expectation is with respect to $p(a, \underline{y})$. Suppose the auxiliary model is $q(\underline{y}|a)$ and define

$$q(\underline{y}) = \int_{\mathcal{A}} p(a)q(\underline{y}|a)^s da. \tag{103}$$

The GMI again provides a lower bound on the mutual information since (cf. (43))

$$I_s(A; \underline{Y}) = I(A; \underline{Y}) - D(p_{A, \underline{Y}} \| p_{\underline{Y}} q_{A|\underline{Y}}) \tag{104}$$

where $q(a|\underline{y}) = p(a)q(\underline{y}|a)^s / q(\underline{y})$ is a reverse channel density.

We next study reverse and forward models as in Sections 1.3 and 1.4. Suppose the entries $\underline{X}(s_T)$ of A are jointly CSCG.

Reverse Model: We write \underline{A} when we consider A to be a column vector that stacks the $\underline{X}(s_T)$. Consider the following reverse density motivated by (13):

$$q(\underline{a}|\underline{y}) = \frac{\exp\left(-(\underline{a} - \mathbb{E}[\underline{A}|\underline{Y} = \underline{y}])^\dagger \mathbf{Q}_{\underline{A}|\underline{Y}=\underline{y}}^{-1} (\underline{a} - \mathbb{E}[\underline{A}|\underline{Y} = \underline{y}])\right)}{\pi^{\nu M} \det \mathbf{Q}_{\underline{A}|\underline{Y}=\underline{y}}}. \tag{105}$$

A corresponding forward model is $q(\underline{y}|a) = q(\underline{a}|\underline{y}) / p(a)$ and the GMI with $s = 1$ becomes (cf. (35))

$$I_1(A; \underline{Y}) = \mathbb{E} \left[\log \frac{\det \mathbf{Q}_{\underline{A}}}{\det \mathbf{Q}_{\underline{A}|\underline{Y}}} \right]. \tag{106}$$

To simplify, one may focus on adaptive symbols as in (89):

$$\underline{X} = \mathbf{Q}_{\underline{X}(s_T)}^{1/2} \cdot \underline{U} \tag{107}$$

where $\underline{U} \sim \mathcal{CN}(\underline{0}, \mathbf{I})$ and the $\mathbf{Q}_{\underline{X}(s_T)}$ are covariance matrices. We thus have $I(A; \underline{Y}) = I(\underline{U}; \underline{Y})$ (cf. (96)) and using (105) but with \underline{A} replaced with \underline{U} we obtain

$$I_1(A; \underline{Y}) = \mathbb{E} \left[-\log \det \mathbf{Q}_{\underline{U}|\underline{Y}} \right]. \tag{108}$$

Forward Model: Perhaps the simplest forward model is $q(\underline{y}|a) = p(\underline{y}|\underline{x}(s_T))$ for some fixed value $s_T \in S_T$. One may interpret this model as having the receiver assume that $S_T = s_T$. A natural generalization of this idea is as follows: define the auxiliary vector

$$\underline{\tilde{X}} = \sum_{s_T} \mathbf{W}(s_T) \underline{X}(s_T) \tag{109}$$

where the $\mathbf{W}(s_T)$ are $M \times M$ complex matrices, i.e., $\underline{\tilde{X}}$ is a linear function of the entries of $A = [\underline{X}(s_T) : s_T \in \mathcal{S}_T]$. For example, the matrices might be chosen based on $P_{\mathcal{S}_T}(\cdot)$. However, observe that $\underline{\tilde{X}}$ is independent of S_T . Now define the auxiliary model

$$q(\underline{y}|a) = q(\underline{y}|\underline{\tilde{x}})$$

where we abuse notation by using the same $q(\cdot)$. The expression (103) becomes

$$q(\underline{y}) = \int_{\mathcal{A}} p(a) q(\underline{y}|a)^s da = \int_{\mathbb{C}} p(\underline{\tilde{x}}) q(\underline{y}|\underline{\tilde{x}})^s d\underline{\tilde{x}}. \tag{110}$$

Remark 27. We often consider \mathcal{S}_T to be a discrete set, but for CSCG channels we also consider $\mathcal{S}_T = \mathbb{C}$ so that the sum over \mathcal{S}_T in (109) is replaced by an integral over \mathbb{C} .

We now specialize further by choosing the auxiliary channel $\underline{Y}_a = \mathbf{H} \underline{\tilde{X}} + \underline{Z}$ where \mathbf{H} is an $N \times M$ complex matrix, \underline{Z} is an N -dimensional CSCG vector that is independent of $\underline{\tilde{X}}$ and has invertible covariance matrix $\mathbf{Q}_{\underline{Z}}$, and \mathbf{H} and $\mathbf{Q}_{\underline{Z}}$ are to be optimized. Further choose $A = [\underline{X}(s_T) : s_T \in \mathcal{S}_T]$ whose entries are jointly CSCG with correlation matrices

$$\mathbf{R}(s_{T1}, s_{T2}) = \mathbb{E} \left[\underline{X}(s_{T1}) \underline{X}(s_{T2})^\dagger \right].$$

Since $\underline{\tilde{X}}$ in (109) is independent of S_T , we have

$$q(\underline{y}|a) = \frac{\exp \left(-(\underline{y} - \mathbf{H} \underline{\tilde{x}})^\dagger \mathbf{Q}_{\underline{Z}}^{-1} (\underline{y} - \mathbf{H} \underline{\tilde{x}}) \right)}{\pi^N \det \mathbf{Q}_{\underline{Z}}}. \tag{111}$$

Moreover, $\underline{\tilde{X}}$ is CSCG so (110) is

$$q(\underline{y}) = \frac{\pi^N \det(\mathbf{Q}_{\underline{Z}}/s)}{(\pi^N \det \mathbf{Q}_{\underline{Z}})^s} \cdot \frac{\exp \left(-\underline{y}^\dagger (\mathbf{Q}_{\underline{Z}}/s + \mathbf{H} \mathbf{Q}_{\underline{\tilde{X}}} \mathbf{H}^\dagger)^{-1} \underline{y} \right)}{\pi^N \det (\mathbf{Q}_{\underline{Z}}/s + \mathbf{H} \mathbf{Q}_{\underline{\tilde{X}}} \mathbf{H}^\dagger)}$$

where

$$\mathbf{Q}_{\underline{\tilde{X}}} = \sum_{s_{T1}, s_{T2}} \mathbf{W}(s_{T1}) \mathbf{R}(s_{T1}, s_{T2}) \mathbf{W}(s_{T2})^\dagger.$$

We have the following generalization of Proposition 1.

Lemma 2. The maximum GMI (102) for the channel $p(\underline{y}|a)$, an adaptive vector $A = [\underline{X}(s_T) : s_T \in \mathcal{S}_T]$ that has jointly CSCG entries, an $\underline{\tilde{X}}$ as in (109) with $\mathbf{Q}_{\underline{\tilde{X}}} \succ \mathbf{0}$, and the auxiliary model (111) with $\mathbf{Q}_{\underline{Z}} \succ \mathbf{0}$ is

$$I_1(A; \underline{Y}) = \log \det \left(\mathbf{I} + \mathbf{Q}_{\underline{Z}}^{-1} \tilde{\mathbf{H}} \mathbf{Q}_{\underline{\tilde{X}}} \tilde{\mathbf{H}}^\dagger \right) \tag{112}$$

where (cf. (31))

$$\tilde{\mathbf{H}} = \mathbb{E} \left[\underline{Y} \underline{\tilde{X}}^\dagger \right] \mathbf{Q}_{\underline{\tilde{X}}}^{-1} \tag{113}$$

$$\mathbf{Q}_{\underline{\tilde{Z}}} = \mathbf{Q}_{\underline{Y}} - \tilde{\mathbf{H}} \mathbf{Q}_{\underline{\tilde{X}}} \tilde{\mathbf{H}}^\dagger. \tag{114}$$

The expectation is with respect to the actual channel with joint distribution/density $p(a, \underline{y})$.

Proof. See Appendix D. \square

Remark 28. Since $\underline{\tilde{X}}$ is a function of A , the rate (112) can alternatively be derived by using $I(A; \underline{Y}) \geq I(\underline{\tilde{X}}; \underline{Y})$ and applying the bound (30) with \underline{X} replaced with $\underline{\tilde{X}}$.

Remark 29. The estimate $\tilde{\mathbf{H}}$ is the MMSE estimate of \mathbf{H} :

$$\tilde{\mathbf{H}} = \arg \min_{\mathbf{H}} \mathbb{E} \left[\|\underline{Y} - \mathbf{H}\underline{\tilde{X}}\|^2 \right] \tag{115}$$

and $\mathbf{Q}_{\underline{\tilde{Z}}}$ is the resulting covariance matrix of the error. To see this, expand (cf. (54))

$$\begin{aligned} \mathbb{E} \left[\|\underline{Y} - \mathbf{H}\underline{\tilde{X}}\|^2 \right] &= \mathbb{E} \left[\left\| (\underline{Y} - \tilde{\mathbf{H}}\underline{\tilde{X}}) + (\tilde{\mathbf{H}} - \mathbf{H})\underline{\tilde{X}} \right\|^2 \right] \\ &= \mathbb{E} \left[\|\underline{Y} - \tilde{\mathbf{H}}\underline{\tilde{X}}\|^2 \right] + \text{tr} \left((\tilde{\mathbf{H}} - \mathbf{H})\mathbf{Q}_{\underline{\tilde{X}}}(\tilde{\mathbf{H}} - \mathbf{H})^\dagger \right) \end{aligned} \tag{116}$$

where the final step follows by the definition of $\tilde{\mathbf{H}}$ in (113).

Remark 30. Suppose that \mathbf{H} is an estimate other than (115). Generalizing (55), if $\mathbf{Q}_{\underline{Y}} \succ \mathbf{Q}_{\underline{Z}}$ we may choose

$$\mathbf{Q}_{\underline{Z}}/s = \left(\mathbf{H}\mathbf{Q}_{\underline{\tilde{X}}}\mathbf{H}^\dagger \right)^{1/2} \left(\mathbf{Q}_{\underline{Y}} - \mathbf{Q}_{\underline{Z}} \right)^{-1/2} \mathbf{Q}_{\underline{Z}} \left(\mathbf{Q}_{\underline{Y}} - \mathbf{Q}_{\underline{Z}} \right)^{-1/2} \left(\mathbf{H}\mathbf{Q}_{\underline{\tilde{X}}}\mathbf{H}^\dagger \right)^{1/2} \tag{117}$$

where

$$\mathbf{Q}_{\underline{Z}} = \mathbb{E} \left[(\underline{Y} - \mathbf{H}\underline{\tilde{X}})(\underline{Y} - \mathbf{H}\underline{\tilde{X}})^\dagger \right]. \tag{118}$$

Appendix D shows that (102) then simplifies to (cf. (56))

$$I_s(A; \underline{Y}) = \log \det \left(\mathbf{Q}_{\underline{Z}}^{-1} \mathbf{Q}_{\underline{Y}} \right). \tag{119}$$

Remark 31. The GMI (112) does not depend on the scaling of $\underline{\tilde{X}}$ since this is absorbed in $\tilde{\mathbf{H}}$. For example, one can choose the weighting matrices in (109) so that $\mathbb{E} \left[\|\underline{\tilde{X}}\|^2 \right] = P$.

4.5. Optimal Codebooks for CSCG Forward Models

The following Lemma maximizes the GMI for scalar channels and A with CSCG entries without requiring A to have the form (89). Nevertheless, this form is optimal, and we refer to ([10], page 2013) and Section 6.4 for similar results. In the following, let $U(s_T) \sim \mathcal{CN}(0, 1)$ for all s_T .

Lemma 3. The maximum GMI (102) for the channel $p(y|a)$, an adaptive symbol A with jointly CSCG entries, the forward model (111), and with fixed $P(s_T) = \mathbb{E} \left[|X(s_T)|^2 \right]$ is

$$I_1(A; \underline{Y}) = \log \left(1 + \frac{\tilde{P}}{\mathbb{E} \left[|\underline{Y}|^2 \right] - \tilde{P}} \right) \tag{120}$$

where, writing $X(s_T) = \sqrt{P(s_T)} U(s_T)$ for all s_T , we have

$$\tilde{P} = \mathbb{E} \left[\left| \mathbb{E} \left[\underline{Y} U(s_T)^* \mid S_T \right] \right|^2 \right]. \tag{121}$$

This GMI is achieved by choosing fully-correlated symbols:

$$X(s_T) = \sqrt{P(s_T)} e^{j\phi(s_T)} U \tag{122}$$

and $\tilde{X} = c U$ for some non-zero constant c and a common $U \sim \mathcal{CN}(0, 1)$, and where

$$\phi(s_T) = -\arg\left(\mathbb{E}\left[YU(s_T)^* \mid S_T = s_T\right]\right). \tag{123}$$

Proof. See Appendix E. \square

Remark 32. The expression (121) is based on (A58) in Appendix E and can alternatively be written as $\tilde{P} = |\tilde{h}|^2 \bar{P}$ where

$$\tilde{h} = \mathbb{E}\left[Y\tilde{X}^*\right] / \bar{P}.$$

Remark 33. The power levels $P(s_T)$ may be optimized, usually under a constraint such as $\mathbb{E}[P(s_T)] \leq P$.

Remark 34. By the Cauchy-Schwarz inequality, we have

$$\mathbb{E}\left[\left|\mathbb{E}\left[YU(s_T)^* \mid S_T\right]\right|^2\right] \leq \mathbb{E}\left[|Y|^2\right].$$

Furthermore, equality holds if and only if $|YU(s_T)^*|$ is a constant for each s_T , but this case is not interesting.

4.6. Forward Model GMI for MIMO Channels

The following lemma generalizes Lemma 3 to MIMO channels without claiming a closed-form expression for the optimal GMI. Let $\underline{U}(s_T) \sim \mathcal{CN}(\underline{0}, \mathbf{I})$ for all s_T .

Lemma 4. A GMI (102) for the channel $p(y|a)$, an adaptive vector A with jointly CSCG entries, the auxiliary model (111), and with fixed $\mathbf{Q}_{\underline{X}(s_T)}$ is given by (112) that we write as

$$I_1(A; \underline{Y}) = \log\left(\frac{\det \mathbf{Q}_{\underline{Y}}}{\det(\mathbf{Q}_{\underline{Y}} - \tilde{\mathbf{D}} \tilde{\mathbf{D}}^\dagger)}\right). \tag{124}$$

where for $M \times M$ unitary $\mathbf{V}_R(s_T)$ we have

$$\tilde{\mathbf{D}} = \mathbb{E}\left[\mathbf{U}_T(s_T) \mathbf{\Sigma}(s_T) \mathbf{V}_R(s_T)^\dagger\right] \tag{125}$$

and $\mathbf{U}_T(s_T)$ and $\mathbf{\Sigma}(s_T)$ are $N \times N$ unitary and $N \times M$ rectangular diagonal matrices, respectively, of the SVD

$$\mathbb{E}\left[\underline{Y}\underline{U}(s_T)^\dagger \mid S_T = s_T\right] = \mathbf{U}_T(s_T) \mathbf{\Sigma}(s_T) \mathbf{V}_T(s_T)^\dagger \tag{126}$$

for all s_T , and the $\mathbf{V}_T(s_T)$ are $M \times M$ unitary matrices. The GMI (124) is achieved by choosing the symbols (cf. (122) and (A87) below):

$$\underline{X}(s_T) = \mathbf{Q}_{\underline{X}(s_T)}^{1/2} \mathbf{V}_T(s_T) \underline{U} \tag{127}$$

and $\tilde{X} = \mathbf{C}\underline{U}$ for some invertible $M \times M$ matrix \mathbf{C} and a common M -dimensional vector $\underline{U} \sim \mathcal{CN}(\underline{0}, \mathbf{I})$. One may maximize (124) over the unitary $\mathbf{V}_R(s_T)$.

Proof. See Appendix G. \square

Using Lemma 4, the theory for MISO channels with $N = 1$ is similar to the scalar case of Lemma 3; see Remark 35 below. However, optimizing the GMI is more difficult for $N > 1$ because one must optimize over the unitary matrices $\mathbf{V}_R(s_T)$ in (125); see Remark 36 below.

Remark 35. Consider $N = 1$ in which case one may set $\mathbf{U}_T(s_T) = \mathbf{1}$ and (126) is a $1 \times M$ vector where $\Sigma(s_T)$ has as the only non-zero singular value

$$\sigma(s_T) = \left\| \mathbb{E} \left[\underline{Y} \underline{U}(s_T)^\dagger \mid S_T = s_T \right] \right\| = \left(\sum_{m=1}^M \left| \mathbb{E} \left[Y U_m(s_T)^* \mid S_T = s_T \right] \right|^2 \right)^{1/2}. \tag{128}$$

The absolute value of the scalar (125) is maximized by choosing $\mathbf{V}_R(s_T) = \mathbf{I}$ for all s_T to obtain (cf. (121))

$$\tilde{\mathbf{D}} \tilde{\mathbf{D}}^\dagger = \mathbb{E}[\sigma(s_T)]^2. \tag{129}$$

Remark 36. Consider $M = 1$ in which case one may set $\mathbf{V}_T(s_T) = \mathbf{1}$ and (126) is a $N \times 1$ vector where $\Sigma(s_T)$ has as the only non-zero singular value

$$\sigma(s_T) = \left\| \mathbb{E} \left[\underline{Y} U(s_T)^\dagger \mid S_T = s_T \right] \right\| = \left(\sum_{n=1}^N \left| \mathbb{E} \left[Y_n U(s_T)^* \mid S_T = s_T \right] \right|^2 \right)^{1/2}. \tag{130}$$

We should now find the $\mathbf{V}_R(s_T) = e^{j\phi_R(s_T)}$ that minimize the determinant in the denominator of (124) where (see (125))

$$\tilde{\mathbf{D}} = \mathbb{E} \left[\underline{u}_T(s_T) \sigma(s_T) e^{-j\phi_R(s_T)} \right] \tag{131}$$

and where each $\underline{u}_T(s_T)$ is one of the columns of the $N \times N$ unitary matrix $\mathbf{U}_T(s_T)$.

Remark 37. Consider $M = N$ and the product channel

$$p(\underline{y} | a) = \prod_{m=1}^M p(y_m | [x_m(s_T) : s_T \in \mathcal{S}_T]) \tag{132}$$

where $x_m(s_T)$ is the m 'th entry of $\underline{x}(s_T)$. We choose $\mathbf{Q}_{\underline{X}(s_T)}$ as diagonal with diagonal entries $\sqrt{P_m(s_T)}$, $m = 1, \dots, M$. Also choosing $\mathbf{V}_R(s_T) = \mathbf{I}$ makes the matrix $\tilde{\mathbf{D}} \tilde{\mathbf{D}}^\dagger$ diagonal with the diagonal entries (cf. (121) where $M = N = 1$)

$$\left(\sum_{s_T} P_{s_T} \left| \mathbb{E} \left[Y_m U_m(s_T)^* \mid S_T = s_T \right] \right|^2 \right)^2 \tag{133}$$

for $m = 1, \dots, M$. The GMI (124) is thus (cf. (120))

$$I_1(A; \underline{Y}) = \sum_{m=1}^M \log \left(\frac{\mathbb{E}[|Y_m|^2]}{\mathbb{E}[|Y_m|^2] - \mathbb{E}[\left| \mathbb{E} \left[Y_m U_m(s_T)^* \mid S_T \right] \right|^2]} \right). \tag{134}$$

Remark 38. For general $p(\underline{y} | a)$, one might wish to choose diagonal $\mathbf{Q}_{\underline{X}(s_T)}$ and a product model

$$q(\underline{y} | a) = \prod_{m=1}^M q_m(y_m | \bar{x}_m)$$

where the $q_m(\cdot)$ are scalar AWGN channels

$$q_m(y | x) = \frac{1}{\pi \sigma_m^2} \exp \left(-|y - h_m x|^2 / \sigma_m^2 \right)$$

with possibly different h_m and σ_m^2 for each m . Consider also

$$\bar{X}_m = \sum_{s_T} w_m(s_T) X_m(s_T)$$

for some complex weights $w_m(s_T)$, i.e., \bar{X}_m is a weighted sum of entries from the list $[X_m(s_T) : s_T \in S_T]$. The maximum GMI is now the same as (134) but without requiring the actual channel to have the form (132).

Remark 39. If the actual channel is $\underline{Y} = \mathbf{H} \underline{X} + \underline{Z}$ then

$$E[\underline{Y} \underline{U}(s_T)^\dagger | S_T = s_T] = E[\mathbf{H} \underline{X}(s_T) \underline{U}(s_T)^\dagger | S_T = s_T] = E[\mathbf{H} | S_T = s_T] \mathbf{Q}_{\underline{X}(s_T)}^{1/2} \tag{135}$$

where the final step follows because $\underline{U}(S_T) - S_T - \mathbf{H}$ forms a Markov chain. The expression (135) is useful because it separates the effects of the channel and the transmitter.

Remark 40. Combining Remarks 37 and 39, suppose the actual channel is $\underline{Y} = \mathbf{H} \underline{X} + \underline{Z}$ with $M = N$ and where \mathbf{H} is diagonal with diagonal entries $H_m, m = 1, \dots, M$. The GMI (124) is then (cf. (134))

$$I_1(A; \underline{Y}) = \sum_{m=1}^M \log \left(\frac{E[|Y_m|^2]}{E[|Y_m|^2] - E\left[\left|E\left[H_m \sqrt{P_m}(S_T) \mid S_T\right]\right|^2\right]} \right) \tag{136}$$

where $E[|Y_m|^2] = 1 + E[|H_m|^2 P_m(S_T)]$.

5. Channels with CSIR and CSIT

Shannon’s model includes CSIR [11]. The FDG is shown in Figure 3 where there is a hidden state S_H , the CSIR S_R and CSIT S_T are functions of S_H , and the receiver sees the channel outputs

$$[Y_i, S_{Ri}] = [f(X_i, S_{Hi}, Z_i), S_{Ri}] \tag{137}$$

for some function $f(\cdot)$ and $i = 1, 2, \dots, n$. (By defining $S_H = [S_{H1}, Z_H]$ and calling S_{H1} the hidden channel state we can include the case where S_R and S_T are noisy functions of S_{H1} .) As before, M, S_H^n, Z^n are mutually statistically independent, and S_H^n and Z^n are i.i.d. strings of random variables with the same distributions as S_T and Z , respectively. Observe that we have changed the notation by writing Y for only part of the channel output. The new Y (without the S_R) is usually called the “channel output”.

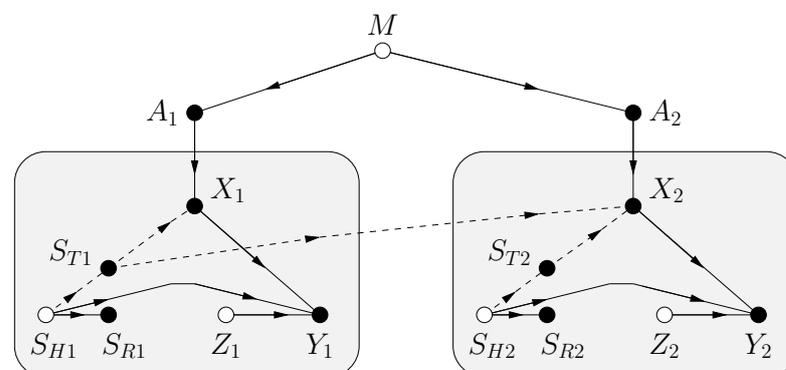


Figure 3. FDG for $n = 2$ channel uses with different CSIT and CSIR. The hidden channel state S_{Hi} permits dependent S_{Ri} and S_{Ti} .

5.1. Capacity and GMI

We begin with scalar channels for which (90) is

$$C = \max_A I(A; Y, S_R) = \max_A I(A; Y|S_R) \tag{138}$$

where A and S_R are independent.

Reverse Model: The expression (108) with the adaptive symbol (88) is

$$I_1(A; Y, S_R) = E[-\log \text{Var}[U|Y, S_R]]. \tag{139}$$

Forward Model: Consider the expansion

$$I_1(A; Y|S_R) = \int_{S_R} p(s_R) I_1(A; Y|S_R = s_R) ds_R \tag{140}$$

where $I_1(A; Y|S_R = s_R)$ is the GMI (102) with all densities conditioned on $S_R = s_R$. We choose the forward model

$$q(y|a, s_R) = \frac{1}{\pi\sigma(s_R)^2} \exp\left(-\frac{|y - h(s_R)\bar{x}(s_R)|^2}{\sigma(s_R)^2}\right). \tag{141}$$

where similar to (109) we define

$$\bar{X}(s_R) = \sum_{s_T} w(s_T, s_R) X(s_T) \tag{142}$$

for complex weights $w(s_T, s_R)$, i.e., $\bar{X}(s_R)$ is a weighted sum of entries from the list $A = [X(s_T) : s_T \in \mathcal{S}_T]$. We have the following straightforward generalization of Lemma 3.

Theorem 1. *The maximum GMI (140) for the channel $p(y|a, s_R)$, an adaptive symbol A with jointly CSCG entries, the model (141), and with fixed $P(s_T) = E[|X(s_T)|^2]$ is*

$$I_1(A; Y|S_R) = E\left[\log\left(1 + \frac{\tilde{P}(S_R)}{E[|Y|^2|S_R] - \tilde{P}(S_R)}\right)\right] \tag{143}$$

where for all $s_R \in \mathcal{S}_R$ we have

$$\tilde{P}(s_R) = E\left[\left|E\left[YU(s_T)^* \middle| S_T, S_R = s_R \right]\right|^2\right]. \tag{144}$$

Remark 41. *To establish Theorem 1, the receiver may choose $\bar{X} = \sqrt{P}U$ to be independent of s_R . Alternatively, the receiver may choose $\bar{X}(s_R) = \sqrt{E[|X|^2|S_R = s_R]}U$. Both choices give the same GMI since the expectation in (144) does not depend on the scaling of \bar{X} ; see Remark 31.*

Remark 42. *The partition idea of Lemmas 1 and 5 carries over to Theorem 1. We may generalize (143) as*

$$I_1(A; Y|S_R) = \int_{S_R} p(s_R) \sum_{k=1}^K \Pr[\mathcal{E}_k|S_R = s_R] \left[\log\left(1 + \frac{|h_k(s_R)|^2 P}{\sigma_k^2(s_R)}\right) + \frac{E[|Y|^2|\mathcal{E}_k, S_R = s_R]}{\sigma_k^2(s_R) + |h_k(s_R)|^2 P} - \frac{E\left[|Y - h_k(s_R)\sqrt{P}U\right|^2|\mathcal{E}_k, S_R = s_R\right]}{\sigma_k^2(s_R)} \right] ds_R \tag{145}$$

where the $X(s_T)$, $s_T \in \mathcal{S}_T$, are given by (122) and the $h_k(s_R)$ and $\sigma_k^2(s_R)$, $k = 1, \dots, K$, $s_R \in \mathcal{S}_R$, can be optimized.

Remark 43. One is usually interested in the optimal power control policy $P(s_T)$ under the constraint $E[P(s_T)] \leq P$. Taking the derivative of (143) with respect to $\sqrt{P(s_T)}$ and setting to zero we obtain

$$E \left[\frac{E[|Y|^2|S_R] \tilde{P}(S_R)' - \tilde{P}(S_R) E[|Y|^2|S_R]'}{E[|Y|^2|S_R] [E[|Y|^2|S_R] - \tilde{P}(S_R)]} \right] = 2\lambda \sqrt{P(s_T)} P_{S_T}(s_T) \tag{146}$$

where $\tilde{P}(S_R)'$ and $E[|Y|^2|S_R]'$ are derivatives with respect to $\sqrt{P(s_T)}$. We use (146) below to derive power control policies.

Remark 44. A related model is a compound channel where $p(y|a, s_R)$ is indexed by the parameter s_R ([91], Chapter 4). The problem is to find the maximum worst-case reliable rate if the transmitter does not know s_R . Alternatively, the transmitter must send its message to all $|\mathcal{S}_R|$ receivers indexed by $s_R \in \mathcal{S}_R$. A compound channel may thus be interpreted as a broadcast channel with a common message.

5.2. CSIT@R

An interesting specialization of Shannon’s model is when the receiver knows S_T and can determine $X(S_T)$. We refer to this scenario as CSIT@R. The model was considered in ([10], Section II) when S_T is a function of S_R . More generally, suppose S_T is a function of $[Y, S_R]$. The capacity (138) then simplifies to (see ([10], Proposition 1))

$$\begin{aligned} C &\stackrel{(a)}{=} \max_A I(A; Y, S_T | S_R) \\ &\stackrel{(b)}{=} \max_A I(X; Y | S_R, S_T) \\ &\stackrel{(c)}{=} \sum_{s_T} P_{S_T}(s_T) \left[\max_{X(s_T)} I(X(s_T); Y | S_R, S_T = s_T) \right] \end{aligned} \tag{147}$$

where step (a) follows because S_T is a function of $[Y, S_R]$; step (b) follows because A and (S_R, S_T) are independent, X is a function of $[A, S_T]$, and $A - [S_T, X] - Y$ forms a Markov chain; and step (c) follows because one may optimize $X(s_T)$ separately for each $s_T \in \mathcal{S}_T$.

As discussed in [10], a practical motivation for this model is when the CSIT is based on error-free feedback from the receiver to the transmitter. In this case, where S_T is a function of S_R , the expression (144) becomes

$$\tilde{P}(s_R) = \left| E \left[Y U(s_T)^* \mid S_R = s_R \right] \right|^2. \tag{148}$$

Remark 45. The insight that one can replace adaptive symbols A with channel inputs X when X is a function of A and past Y appeared for two-way channels in ([9], Section 4.2.3) and networks in ([22], Section V.A), ([72], Section IV.F).

5.3. MIMO Channels and K-Partitions

We consider generalizations to MIMO channels and K -partitions as in Section 3.2.

MIMO Channels: Consider the average GMI

$$I_1(A; \underline{Y} | S_R) = \int_{S_R} p(s_R) I_1(A; \underline{Y} | S_R = s_R) ds_R \tag{149}$$

and choose the parameters (113) and (114) for the event $S_R = s_R$. We have

$$\tilde{\mathbf{H}}(s_R) = \mathbb{E} \left[\underline{\mathbf{Y}} \underline{\mathbf{X}}^\dagger \mid S_R = s_R \right] \mathbb{E} \left[\underline{\mathbf{X}} \underline{\mathbf{X}}^\dagger \mid S_R = s_R \right]^{-1} \tag{150}$$

$$\mathbf{Q}_{\underline{\mathbf{Z}}}(s_R) = \mathbb{E} \left[\underline{\mathbf{Y}} \underline{\mathbf{Y}}^\dagger \mid S_R = s_R \right] - \tilde{\mathbf{H}}(s_R) \mathbb{E} \left[\underline{\mathbf{X}} \underline{\mathbf{X}}^\dagger \mid S_R = s_R \right] \tilde{\mathbf{H}}(s_R)^\dagger \tag{151}$$

and the GMI (149) is (cf. (60) and (112))

$$I_1(A; \underline{\mathbf{Y}} | S_R) = \mathbb{E} \left[\log \det \left(\mathbf{I} + \mathbf{Q}_{\underline{\mathbf{Z}}}(s_R)^{-1} \tilde{\mathbf{H}}(s_R) \mathbf{Q}_{\underline{\mathbf{X}}} \tilde{\mathbf{H}}(s_R)^\dagger \right) \right]. \tag{152}$$

K-Partitions: Let $\{\mathcal{Y}_k : k = 1, \dots, K\}$ be a K -partition of $\underline{\mathcal{Y}}$ and define the events $\mathcal{E}_k = \{\underline{\mathbf{Y}} \in \mathcal{Y}_k\}$ for $k = 1, \dots, K$. As in Remark 13, K -partitioning formally includes (149) as a special case by including S_R as part of the receiver’s “overall” channel output $\underline{\mathbf{Y}} = [\underline{\mathbf{Y}}, S_R]$. The following lemma generalizes Lemma 1.

Lemma 5. A GMI with $s = 1$ for the channel $p(\underline{\mathbf{y}}|a)$ is

$$\begin{aligned} I_1(A; \underline{\mathbf{Y}}) = \sum_{k=1}^K \Pr[\mathcal{E}_k] \left\{ \log \det \left(\mathbf{I} + \mathbf{Q}_{\underline{\mathbf{Z}}_k}^{-1} \mathbf{H}_k \mathbf{Q}_{\underline{\mathbf{X}}} \mathbf{H}_k^\dagger \right) \right. \\ \left. + \mathbb{E} \left[\underline{\mathbf{Y}}^\dagger \left(\mathbf{Q}_{\underline{\mathbf{Z}}_k} + \mathbf{H}_k \mathbf{Q}_{\underline{\mathbf{X}}} \mathbf{H}_k^\dagger \right)^{-1} \underline{\mathbf{Y}} \mid \mathcal{E}_k \right] \right. \\ \left. - \mathbb{E} \left[\left(\underline{\mathbf{Y}} - \mathbf{H}_k \underline{\mathbf{X}} \right)^\dagger \mathbf{Q}_{\underline{\mathbf{Z}}_k}^{-1} \left(\underline{\mathbf{Y}} - \mathbf{H}_k \underline{\mathbf{X}} \right) \mid \mathcal{E}_k \right] \right\} \end{aligned} \tag{153}$$

where the \mathbf{H}_k and $\mathbf{Q}_{\underline{\mathbf{Z}}_k}$, $k = 1, \dots, K$, can be optimized.

Remark 46. For scalars the GMI (153) is

$$I_1(A; Y) = \sum_{k=1}^K \Pr[\mathcal{E}_k] \left[\log \left(1 + \frac{|h_k|^2 \bar{P}}{\sigma_k^2} \right) + \frac{\mathbb{E}[|Y|^2 | \mathcal{E}_k]}{\sigma_k^2 + |h_k|^2 \bar{P}} - \frac{\mathbb{E}[|Y - h_k \bar{X}|^2 | \mathcal{E}_k]}{\sigma_k^2} \right] \tag{154}$$

which is the same as (64) except that \bar{X} , \bar{P} replace X, P . If we follow (66) and (67) then (154) becomes (68) but with

$$h_k = \mathbb{E} \left[Y \bar{X}^* \mid \mathcal{E}_k \right] / P_k, \quad P_k = \mathbb{E} \left[|\bar{X}|^2 \mid \mathcal{E}_k \right].$$

Remark 47. Consider Remark 14 and choose $K = 2$, $h_1 = 0$, $\sigma_1^2 = 1$. The GMI (154) then has only the $k = 2$ term, and it again remains to select h_2 , σ_2^2 , and t_R .

Remark 48. If we define

$$\mathbf{Q}_{\underline{\mathbf{X}}}^{(k)} = \mathbb{E} \left[\underline{\mathbf{X}} \underline{\mathbf{X}}^\dagger \mid \mathcal{E}_k \right], \quad \mathbf{Q}_{\underline{\mathbf{Y}}}^{(k)} = \mathbb{E} \left[\underline{\mathbf{Y}} \underline{\mathbf{Y}}^\dagger \mid \mathcal{E}_k \right] \tag{155}$$

and choose the LMMSE auxiliary models with

$$\mathbf{H}_k = \mathbb{E} \left[\underline{\mathbf{Y}} \underline{\mathbf{X}}^\dagger \mid \mathcal{E}_k \right] \left(\mathbf{Q}_{\underline{\mathbf{X}}}^{(k)} \right)^{-1} \tag{156}$$

$$\mathbf{Q}_{\underline{\mathbf{Z}}_k} = \mathbf{Q}_{\underline{\mathbf{Y}}}^{(k)} - \mathbf{H}_k \mathbf{Q}_{\underline{\mathbf{X}}}^{(k)} \mathbf{H}_k^\dagger \tag{157}$$

for $k = 1, \dots, K$ then the expression (153) is (cf. (68))

$$I_1(A; \underline{Y}) = \sum_{k=1}^K \Pr[\mathcal{E}_k] \left[\log \det \left(\mathbf{I} + \mathbf{Q}_{Z_k}^{-1} \mathbf{H}_k \mathbf{Q}_{\underline{X}} \mathbf{H}_k^\dagger \right) - \text{tr} \left(\left(\mathbf{Q}_{\underline{Y}}^{(k)} + \mathbf{H}_k \mathbf{D}_{\underline{X}}^{(k)} \mathbf{H}_k^\dagger \right)^{-1} \mathbf{H}_k \mathbf{D}_{\underline{X}}^{(k)} \mathbf{H}_k^\dagger \right) \right] \tag{158}$$

where $\mathbf{D}_{\underline{X}}^{(k)} = \mathbf{Q}_{\underline{X}} - \mathbf{Q}_{\underline{X}}^{(k)}$.

Remark 49. We may proceed as in Remark 18 and consider large K . These steps are given in Appendix F.

6. Fading Channels with AWGN

This section treats scalar, complex-alphabet, AWGN channels with CSIR for which the channel output is

$$[Y, S_R] = [HX + Z, S_R] \tag{159}$$

where H, A, Z are mutually independent, $E[|H|^2] = 1$, and $Z \sim \mathcal{CN}(0, 1)$. The capacity under the power constraint $E[|X|^2] \leq P$ is (cf. (138))

$$C(P) = \max_{A: E[|X|^2] \leq P} I(A; Y|S_R). \tag{160}$$

However, the optimization in (160) is often intractable, and we desire expressions with $\log(1 + \text{SNR})$ terms to gain insight. We develop three such expressions: an upper bound and two lower bounds. It will be convenient to write $G = |H|^2$.

Capacity Upper Bound: We state this bound as a lemma since we use it to prove Proposition 2 below.

Lemma 6. The capacity (160) is upper bounded as

$$C(P) \leq \max E[\log(1 + GP(S_T))] \tag{161}$$

where the maximization is over $P(S_T)$ with $E[P(S_T)] = P$.

Proof. Consider the steps

$$\begin{aligned} I(A; Y|S_R) &\leq I(A; Y, S_T, H|S_R) \\ &\stackrel{(a)}{=} I(A; Y|S_R, S_T, H) \\ &= h(Y|S_R, S_T, H) - h(Z) \\ &\stackrel{(b)}{\leq} E[\log \text{Var}[Y|S_R, S_T, H]] \end{aligned} \tag{162}$$

where step (a) is because A and $[S_R, S_T, H]$ are independent, and step (b) follows by the entropy bound

$$h(Y|B = b) \leq \log(\pi e \text{Var}[Y|B = b]) \tag{163}$$

which we applied with $B = [S_R, S_T, H]$. Finally, we compute $\text{Var}[Y|S_R, S_T, H] = 1 + GP(S_T)$. \square

Reverse Model GMI: Consider the adaptive symbol (88) and the GMI (139). We expand the variances in (139) as

$$\text{Var}[U|Y = y, S_R = s_R] = E[|U|^2|Y = y, S_R = s_R] - |E[U|Y = y, S_R = s_R]|^2.$$

Appendix C shows that one may write

$$E[U|Y = y, S_R = s_R] = \int_{\mathbb{C} \times S_T} p(h, s_T|y, s_R) \frac{h\sqrt{P(s_T)}e^{j\phi(s_T)}y}{1 + |h|^2P(s_T)} ds_T dh \quad (164)$$

and

$$E[|U|^2|Y = y, S_R = s_R] = \int_{\mathbb{C} \times S_T} p(h, s_T|y, s_R) \left(\frac{1}{1 + |h|^2P(s_T)} + \frac{|h|^2P(s_T)|y|^2}{(1 + |h|^2P(s_T))^2} \right) ds_T dh. \quad (165)$$

We use the expressions (164) and (165) to compute achievable rates by numerical integration. For example, suppose that $S_T = 0$ and $S_R = H$, i.e., we have full CSIR and no CSIT. The averaging density is then

$$p(h, s_T|y, s_R) = \delta(h - s_R) \delta(s_T)$$

and the variance simplifies to the capacity-achieving form

$$\text{Var}[U|Y = y, S_R = h] = \frac{1}{1 + |h|^2P}.$$

Forward Model GMI: A forward model GMI is given by Theorem 1 where

$$\tilde{P}(s_R) = E \left[\left| E \left[H\sqrt{P(S_T)} \middle| S_T, S_R = s_R \right] \right|^2 \right] \quad (166)$$

$$E[|Y|^2|S_R = s_R] = 1 + E[GP(S_T)|S_R = s_R] \quad (167)$$

so that (143) becomes

$$I_1(A; Y|S_R) = E \left[\log \left(1 + \frac{\tilde{P}(S_R)}{1 + E[GP(S_T)|S_R] - \tilde{P}(S_R)} \right) \right]. \quad (168)$$

Remark 50. Jensen’s inequality implies that the denominator in (168) is greater than or equal to

$$1 + \text{Var} \left[\sqrt{GP(S_T)} \middle| S_R \right]. \quad (169)$$

Equality requires that for all $S_R = s_R$ we have

$$\tilde{P}(s_R) = E \left[\sqrt{GP(S_T)} \middle| S_R = s_R \right]^2 \quad (170)$$

which is valid if H is a function of $[S_R, S_T]$, for example. However, if there is channel uncertainty after conditioning on $[S_R, S_T]$ then $\tilde{P}(s_R)$ is usually smaller than the RHS of (170).

Remark 51. Consider $S_R = H$ or $S_R = H\sqrt{P(S_T)}$. For both cases, H is a function of $[S_R, S_T]$ and the denominator in (168) is the variance (169). In fact, for $S_R = H\sqrt{P(S_T)}$, the expression (169) takes on the minimal value 1. This CSIR is thus the best possible; see Proposition 2.

Remark 52. For MIMO channels we replace (159) with

$$[Y, S_R] = [HX + Z, S_R] \tag{171}$$

where H, A, Z are mutually independent and $Z \sim \mathcal{CN}(0, I)$. One usually considers the constraint $E[\|X\|^2] \leq P$.

Remark 53. The model (171) includes block fading. For example, choosing $M = N$ and $H = H I$ gives scalar block fading. Moreover, the capacity per symbol without in-block feedback is the same as for the $M = N = 1$ case except that P is replaced with P/M ; see [11] and Section 9.

6.1. CSIR and CSIT Models

We study two classes of CSIR, as shown in Table 1. The first class has full (or “perfect”) CSIR, by which we mean either $S_R = H$ or $S_R = H\sqrt{P(S_T)}$. The motivation for studying the latter case is that it models block fading channels with long blocks where the receiver estimates $H\sqrt{P(S_T)}$ using pilot symbols, and the number of pilot symbols is much smaller than the block length [10]. Moreover, one achieves the upper bound (161), see Proposition 2 below.

We coarsely categorize the CSIT as follows:

- Full CSIT: $S_T = H$;
- CSIT@R: $S_T = q_u(G)$ where $q_u(\cdot)$ is the quantizer of Section 2.9 with $B = 0, 1, \infty$;
- Partial CSIT: S_T is not known exactly at the receiver.

The capacity of the CSIT@R models is given by $\log(1 + \text{SNR})$ expressions [10,92]; see also [93]. The partial CSIT model is interesting because achieving capacity generally requires adaptive codewords and closed-form capacity expressions are unavailable. The GMI lower bound of Theorem 1 and Remark 42 and the capacity upper bound of Lemma 6 serve as benchmarks.

Table 1. Models Studied in Section 6 (General Fading), Section 7 (On–Off Fading) and Section 8 (Rayleigh Fading).

		CSIR	
		Full	Partial/No
CSIT	Full	Section 6.3	Section 6.5
	@R	Section 6.3	Section 6.6
	Partial/No	Section 6.4	Section 6.2

The partial CSIR models have S_R being a lossy function of H . For example, a common model is based on LMMSE channel estimation with

$$H = \sqrt{\epsilon} S_R + \sqrt{1-\epsilon} Z_R \tag{172}$$

where $0 \leq \epsilon \leq 1$ and S_R, Z_R are uncorrelated. The CSIT is categorized as above, except that we consider $S_T = f_T(S_R)$ for some function $f_T(\cdot)$ rather than $S_T = q_u(G)$.

To illustrate the theory, we study two types of fading: one with discrete H and one with continuous H , namely

- Section 7: on-off fading with $P_H(0) = P_H(\sqrt{2}) = 1/2$;
- Section 8: Rayleigh fading with $H \sim \mathcal{CN}(0, 1)$.

For on-off fading we have $p(g) = \frac{1}{2}\delta(g) + \frac{1}{2}\delta(g - 2)$ and for Rayleigh fading we have $p(g) = e^{-g} \cdot 1(g \geq 0)$.

Remark 54. For channels with partial CSIR, we will study the GMI for partitions with $K = 1$ and $K = 2$. The full CSIT model has received relatively little attention in the literature, perhaps because CSIR is usually more accurate than CSIT ([5], Section 4.2.3).

6.2. No CSIR, No CSIT

Without CSIR or CSIT, the channel is a classic memoryless channel [94] for which the capacity (160) becomes the usual expression with $S_R = 0$ and $A = X$. For CSCG X and $U = X/E[|X|^2]$, the reverse and forward model GMI (139) and (168) are the respective

$$I_1(X; Y) = E[-\log \text{Var}[U|Y]] \tag{173}$$

$$I_1(X; Y) = \log\left(1 + \frac{P|E[H]|^2}{1 + P \text{Var}[H]}\right). \tag{174}$$

For example, the forward model GMI is zero if $E[H] = 0$.

6.3. Full CSIR, CSIT@R

Consider the full CSIR models with $S_R = H$ and CSIT@R. The capacity is given by $\log(1 + \text{SNR})$ expressions that we review.

First, the capacity with $B = 0$ (no CSIT) is

$$C(P) = E[\log(1 + GP)] = \int_0^\infty p(g) \log(1 + gP) dg. \tag{175}$$

The wideband derivatives are (see (37))

$$C'(0) = E[G] = 1, \quad C''(0) = -E[G^2] \tag{176}$$

so that the wideband values (37) are (see ([73], Theorem 13))

$$\frac{E_b}{N_0} \Big|_{\min} = \log 2, \quad S = \frac{2}{E[G^2]}. \tag{177}$$

The minimal E_b/N_0 is the same as without fading, namely -1.59 dB. However, Jensen's inequality gives $E[G^2] \geq E[G]^2 = 1$ with equality if and only if $G = 1$. Thus, fading reduces the capacity slope S .

More generally, the capacity with full CSIR and $S_T = q_u(G)$ is (see [10])

$$\begin{aligned} C(P) &= \max_{P(S_T): E[P(S_T)] \leq P} E[\log(1 + GP(S_T))] \\ &= \max_{P(S_T): E[P(S_T)] \leq P} \int_0^\infty p(g, s_T) \log(1 + gP(s_T)) dg ds_T. \end{aligned} \tag{178}$$

To optimize the power levels $P(s_T)$, consider the Lagrangian

$$E[\log(1 + GP(S_T))] + \lambda(P - E[P(S_T)]) \tag{179}$$

where $\lambda \geq 0$ is a Lagrange multiplier. Taking the derivative with respect to $P(s_T)$, we have

$$\lambda = E\left[\frac{G}{1 + GP(s_T)} \Big| S_T = s_T\right] = \int_0^\infty p(g|s_T) \frac{g}{1 + gP(s_T)} dg \tag{180}$$

as long as $P(s_T) \geq 0$. If this equation cannot be satisfied, choose $P(s_T) = 0$. Finally, set λ so that $E[P(S_T)] = P$.

For example, consider $B = \infty$ and $S_T = G$. We then have $p(g|s_T) = \delta(g - s_T)$ and therefore

$$P(g) = \left(\frac{1}{\lambda} - \frac{1}{g}\right)^+ \tag{181}$$

where λ is chosen so that $E[P(G)] = P$. The capacity (178) is then (see ([95], Equation (7)))

$$C(P) = \int_{\lambda}^{\infty} p(g) \log(g/\lambda) dg. \tag{182}$$

Consider now the quantizer $q_u(\cdot)$ of Section 2.9 with $B = 1$. We have two equations for λ , namely

$$\lambda = \int_0^{\Delta} \frac{p(g)}{P_{S_T}(\Delta/2)} \cdot \frac{g}{1 + gP(\Delta/2)} dg \tag{183}$$

$$\lambda = \int_{\Delta}^{\infty} \frac{p(g)}{P_{S_T}(3\Delta/2)} \cdot \frac{g}{1 + gP(3\Delta/2)} dg. \tag{184}$$

Observe the following for (183) and (184):

- both $P(\Delta/2)$ and $P(3\Delta/2)$ decrease as λ increases;
- the maximal λ permitted by (183) is $E[G|G \leq \Delta]$ which is obtained with $P(\Delta/2) = 0$;
- the maximal λ permitted by (184) is $E[G|G \geq \Delta]$ which is obtained with $P(3\Delta/2) = 0$.

Thus, if $E[G|G \geq \Delta] > E[G|G \leq \Delta]$, then at P below some threshold, we have $P(\Delta/2) = 0$ and $P(3\Delta/2) = P/P_{S_T}(3\Delta/2)$. The capacity in nats per symbol at low power and for fixed Δ is thus

$$\begin{aligned} C(P) &= \int_{\Delta}^{\infty} p(g) \log(1 + gP(3\Delta/2)) dg \\ &\approx P E[G|G \geq \Delta] - \frac{P^2}{2P_{S_T}(3\Delta/2)} E[G^2|G \geq \Delta] \end{aligned} \tag{185}$$

where we used

$$\log(1 + x) \approx x - \frac{x^2}{2}$$

for small x . The wideband values (37) are

$$\left. \frac{E_b}{N_0} \right|_{\min} = \frac{\log 2}{E[G|G \geq \Delta]} \tag{186}$$

$$S = \frac{2P_{S_T}(3\Delta/2) E[G|G \geq \Delta]^2}{E[G^2|G \geq \Delta]}. \tag{187}$$

One can thus make the minimum E_b/N_0 approach $-\infty$ if one can make $E[G|G \geq \Delta]$ as large as desired by increasing Δ .

Remark 55. Consider the MIMO model (171) with $S_R = \mathbf{H}$. Suppose the CSIT is $S_T = f_T(S_R)$ for some function $f_T(\cdot)$. The capacity (178) generalizes to

$$\begin{aligned} C(P) &= \max_{\underline{\mathbf{X}}(S_T): E[\|\underline{\mathbf{X}}(S_T)\|^2] \leq P} I(\underline{\mathbf{X}}; \mathbf{H}\underline{\mathbf{X}} + \underline{\mathbf{Z}}|\mathbf{H}, S_T) \\ &= \max_{\mathbf{Q}(S_T): E[\text{tr}(\mathbf{Q}(S_T))] \leq P} E \left[\log \det \left(\mathbf{I} + \mathbf{H}\mathbf{Q}(S_T)\mathbf{H}^\dagger \right) \right]. \end{aligned} \tag{188}$$

6.4. Full CSIR, Partial CSIT

Consider first the full CSIR $S_R = H\sqrt{P(S_T)}$ and then the less informative $S_R = H$.

$S_R = H\sqrt{P(S_T)}$: We have the following capacity result that implies this CSIR is the best possible since one can achieve the same rate as if the receiver sees both H and S_T ; see the first step of (162). We could thus have classified this model as CSIT@R.

Proposition 2 (see ([10], Proposition 3)). *The capacity of the channel (159) with $S_R = H\sqrt{P(S_T)}$ and general S_T is*

$$\begin{aligned} C(P) &= \max_{P(S_T): E[P(S_T)] \leq P} \int_{\mathbb{C}} p(s_R) \log(1 + |s_R|^2) ds_R \\ &= \max_{P(S_T): E[P(S_T)] \leq P} E[\log(1 + GP(S_T))]. \end{aligned} \tag{189}$$

Proof. Achievability follows by Theorem 1 with Remark 51. The converse is given by Lemma 6. \square

Remark 56. Proposition 2 gives an upper bound and (thus) a target rate when the receiver has partial CSIR. For example, we will use the K-partition idea of Lemma 1 (see also Remark 46) to approach the upper bound for large SNR.

Remark 57. Proposition 2 partially generalizes to block-fading channels; see Proposition 3 in Section 9.5.

$S_R = H$: The capacity is (138) with

$$I(A; Y|H) = E \left[\log \frac{p(Y|A, H)}{p(Y|H)} \right] \tag{190}$$

where $E[|X|^2] \leq P$ and where

$$p(y|a, h) = \int_{\mathbb{C}} p(s_T|h) \frac{e^{-|y-hx(s_T)|^2}}{\pi} ds_T \tag{191}$$

and

$$\begin{aligned} p(y|h) &= \int_{\mathbb{C}} p(s_T|h) \left(\int_{\mathcal{A}} p(a) p(y|a, h, s_T) da \right) ds_T \\ &= \int_{\mathbb{C}} p(s_T|h) \left(\int_{\mathbb{C}} p(x(s_T)) \frac{e^{-|y-hx(s_T)|^2}}{\pi} dx(s_T) \right) ds_T. \end{aligned} \tag{192}$$

For example, if each entry $X(s_T)$ of A is CSCG with variance $P(s_T)$ then

$$p(y|h) = \int_{\mathbb{C}} p(s_T|h) \frac{\exp\left(-\frac{|y|^2}{1+gP(s_T)}\right)}{\pi(1+gP(s_T))} ds_T. \tag{193}$$

In general, one can compute $I(A; Y|H)$ numerically by using (190)–(192), but the calculations are hampered if the integrals in (191) and (192) do not simplify.

For the reverse model GMI (139), the averaging density in (164) and (165) is here

$$p(h, s_T|y, s_R) = \delta(h - s_R) \frac{p(s_T|h) p(y|h, s_T)}{p(y|h)}. \tag{194}$$

We use numerical integration to compute the GMI.

To obtain more insight, we state the forward model rates of Theorem 1 and Remark 51 as a Corollary.

Corollary 1. *An achievable rate for the fading channels (159) with $S_R = H$ and partial CSIT is the forward model GMI*

$$I_1(A; Y|H) = E[\log(1 + \text{SNR}(H))] \tag{195}$$

where

$$\text{SNR}(h) = \frac{|h|^2 \tilde{P}_T(h)}{1 + |h|^2 \text{Var}[\sqrt{P(S_T)} | H = h]} \tag{196}$$

and

$$\tilde{P}_T(h) = \mathbb{E} \left[\sqrt{P(S_T)} | H = h \right]^2. \tag{197}$$

Remark 58. Jensen’s inequality gives

$$\tilde{P}_T(h) \leq \mathbb{E}[P(S_T) | H = h] \tag{198}$$

by the concavity of the square root. Equality holds if and only if $P(S_T)$ is a constant given $H = h$.

Remark 59. Choosing $P(s_T) = P$ for all s_T in Corollary 1 gives $\tilde{P}_T(h) = P$ for all h and the rate (195) is the capacity (175) without CSIT.

Remark 60. For large P , the $\text{SNR}(h)$ in (196) saturates unless $P(s_T)/P \rightarrow 1$ for all s_T , i.e., the high-SNR capacity is the same as the capacity without CSIT. The CSIT thus must become more accurate as P increases to improve the rate.

Remark 61. To optimize the power levels, consider (146) and

$$\tilde{P}(h)' = 2|h|^2 \sqrt{\tilde{P}_T(h)} p(s_T|h) \tag{199}$$

$$\mathbb{E}[|Y|^2 | H = h]' = 2|h|^2 \sqrt{P(s_T)} p(s_T|h). \tag{200}$$

However, the resulting equations give little insight due to the expectation over H in (146). An exception is the on-off fading case where the expectation has only one term; see (254) and (255).

6.5. Partial CSIR, Full CSIT

Suppose S_R is a (perhaps noisy) function of H ; see (172). The capacity is given by (160) for which we need to compute $p(y|a, s_R)$ and $p(y|s_R)$. The GMI with a K -partition of the output space $\mathcal{Y} \times S_R$ can be helpful for these problems. We assume that the CSIR is either $S_R = 0$ or $S_R = 1 (G \geq t)$ for some transmitter threshold t ; see [95].

Suppose that $S_T = H$. We then have

$$p(y|a, s_R) = \int_{\mathbb{C}} p(h|s_R) \frac{\exp(-|y - h x(h)|^2)}{\pi} dh$$

$$p(y|s_R) = \int_{\mathbb{C}^2} p(h|s_R) p(x(h)) \frac{\exp(-|y - h x(h)|^2)}{\pi} dx(h) dh.$$

Now select the $X(h)$ to be jointly CSCG with variances $\mathbb{E}[|X(h)|^2] = P(h)$ and correlation coefficients

$$\rho(h, h') = \frac{\mathbb{E}[X(h)X(h')^*]}{\sqrt{P(h)P(h')}}.$$

and where $\mathbb{E}[P(H)] \leq P$. We then have

$$p(y|s_R) = \int_{\mathbb{C}} p(h) \frac{e^{-|y|^2/(|h|^2 P(h)+1)}}{2\pi(|h|^2 P(h)+1)} dh.$$

As in (98), $p(y|s_R)$ and therefore $h(Y|S_R)$ depend only on the marginals $p(x(h))$ of A and not on the $\rho(h, h')$. We thus have the problem of finding the $\rho(h, h')$ that minimize

$$h(Y|S_R, A) = \int_{\mathcal{A}} p(a) h(Y|S_R, A = a) da.$$

However, we study the conventional A in (88) for simplicity.

For the reverse model GMI (139), the averaging density in (164) and (165) is (cf. (194))

$$p(h, s_T|y, s_R) = \delta(s_T - h) \frac{p(h|s_R) p(y|h, s_R)}{p(y|s_R)}. \tag{201}$$

We again use numerical integration to compute the GMI.

For the forward model GMI, consider the same model and CSCG X as in Theorem 1. Since H is a function of S_T , we use (169) in Remark 50 to write

$$I_1(A; Y|S_R) = E \left[\log \left(1 + \frac{\tilde{P}(S_R)}{1 + \text{Var} \left[\sqrt{GP(H)} | S_R \right]} \right) \right] \tag{202}$$

where (see (170))

$$\tilde{P}(s_R) = E \left[\sqrt{GP(H)} | S_R = s_R \right]^2 \tag{203}$$

$$E \left[|Y|^2 | S_R = s_R \right] = 1 + E[GP(H) | S_R = s_R]. \tag{204}$$

The transmitter compensates for the phase of H , and it remains to adjust the transmit power levels $P(h)$. We study five power control policies and two types of CSIR; see Table 2.

Table 2. Power Control Policies and Minimal SNRs.

		CSIR	
		None: $S_R = 0$	$S_R = 1(G \geq t)$
Policy	TCP	Equation (221)	Equation (226)
	TMF	Equation (222)	Equation (227)
	TCI	Equation (223)	Equation (228)
	GMI-Optimal	See Theorem 2	
	TMMSE	See Remark 64	

Heuristic Policies: The first three policies are reasonable heuristics and have the form

$$P(h) = \begin{cases} \hat{P} g^a, & g \geq t \\ 0, & \text{else} \end{cases} \tag{205}$$

for some choice of real a and where

$$\hat{P} = \frac{P}{\int_t^\infty p(g) g^a dg}. \tag{206}$$

In particular, choosing $a = 0, +1, -1$, we obtain policies that we call truncated constant power (TCP), truncated matched filtering (TMF), and truncated channel inversion (TCI), respectively; see ([5], page 487), [95]. For such policies, we compute

$$\tilde{P}(s_R) = \hat{P} \left(\int_t^\infty p(g|s_R) \sqrt{g^{1+a}} dg \right)^2 \tag{207}$$

$$E[GP(H) | S_R = s_R] = \hat{P} \int_t^\infty p(g|s_R) g^{1+a} dg. \tag{208}$$

These policies all have the form $P(h) = P \cdot f(h)$ for some function $f(\cdot)$ that is independent of P . The minimum SNR in (37) with $C(P)$ replaced with the GMI is thus

$$\frac{E_b}{N_0} \Big|_{\min} = \frac{(\int_t^\infty p(g) g^a dg) \log 2}{E \left[\left(\int_t^\infty p(g|S_R) \sqrt{g^{1+a}} dg \right)^2 \right]}. \tag{209}$$

For instance, consider the threshold $t = 0$ (no truncation). The TCP ($a = 0$) and TMF ($a = 1$) policies have $\hat{P} = P$ while TCI ($a = -1$) has $P = \hat{P}/E[G^{-1}]$. For TCP, TMF, and TCI, we compute the respective

$$\frac{E_b}{N_0} \Big|_{\min} = \frac{\log 2}{E \left[E \left[\sqrt{G} \mid S_R \right]^2 \right]} \tag{210}$$

$$\frac{E_b}{N_0} \Big|_{\min} = \frac{\log 2}{E \left[E \left[G \mid S_R \right]^2 \right]} \tag{211}$$

$$\frac{E_b}{N_0} \Big|_{\min} = E \left[G^{-1} \right] \log 2. \tag{212}$$

Applying Jensen’s inequality to the square root, square, and inverse functions in (210)–(212), we find that for $t = 0$:

- the minimum E_b/N_0 of TCP and TCI is larger (worse) than -1.59 dB unless there is no fading;
- the minimum E_b/N_0 of TMF is smaller (better) than -1.59 dB unless $E[G|S_R] = E[G] = 1$.

However, we emphasize that these claims apply to the GMI and not necessarily the mutual information; see Section 8.4.

GMI-Optimal Policy: The fourth policy is optimal for the GMI (202) and has the form of an MMSE precoder. This policy motivates a truncated MMSE (TMMSE) policy that generalizes and improves TMF and TCI.

Taking the derivative of the Lagrangian

$$I_1(A; Y|S_R) + \lambda(P - E[P(H)]) \tag{213}$$

with respect to $P(h)$ we have the following result.

Theorem 2. *The optimal power control policy for the GMI $I_1(A; Y|S_R)$ for the fading channels (159) with $S_T = H$ is*

$$\sqrt{P(h)} = \frac{\alpha(h)|h|}{\lambda + \beta(h)|h|^2} \tag{214}$$

where $\lambda > 0$ is chosen so that $E[P(H)] = P$ and

$$\alpha(h) = \int_{\mathbb{C}} p(s_R|h) \frac{\sqrt{\tilde{P}(s_R)}}{E[|Y|^2|S_R = s_R] - \tilde{P}(s_R)} ds_R \tag{215}$$

$$\beta(h) = \int_{\mathbb{C}} p(s_R|h) \frac{\tilde{P}(s_R)}{[E[|Y|^2|S_R = s_R] - \tilde{P}(s_R)]E[|Y|^2|S_R = s_R]} ds_R. \tag{216}$$

Proof. Apply (146) with (203) and (204) to obtain

$$\tilde{P}(s_R)' = 2|h| \sqrt{\tilde{P}(s_R)} p(h|s_R) \tag{217}$$

$$E \left[|Y|^2 \mid S_R = s_R \right]' = 2|h|^2 \sqrt{P(h)} p(h|s_R). \tag{218}$$

Inserting into (146) and rearranging terms we obtain (214) with (215) and (216). \square

Remark 62. The expressions (215) and (216) are self-referencing, as $\tilde{P}(s_R)$ itself depends on $\alpha(h)$ and $\beta(h)$. However, one simplification occurs if s_R is a function of H : $\alpha(h)$ and $\beta(h)$ are functions of s_R only since the $p(s_R|h)$ in (215) and (216) is a Dirac generalized function.

Remark 63. Consider the expression (214). We effectively have a matched filter for small $|h|$; for large $|h|$, we effectively have a channel inversion. Recall that LMMSE filtering has similar behavior for low and high SNR, respectively.

Remark 64. A heuristic based on the optimal policy is a TMMSE policy where the transmitter sets $P(h) = 0$ if $G < t$, and otherwise uses (214) but where $\alpha(h)$, $\beta(h)$ are independent of h . There are thus four parameters to optimize: λ , α , β , and t . This TMMSE policy will outperform TMF and TCI in general, as these are special cases where $\beta = 0$ and $\lambda = 0$, respectively.

$s_R = 0$: For this CSIR, the GMI (202) simplifies to $I_1(A; Y)$ and the heuristic policy (TCP, TMF, TCI) rates are

$$I_1(A; Y) = \log \left(1 + \frac{\hat{P} \mathbb{E} \left[\sqrt{G^{1+a}} \cdot \mathbf{1}(G \geq t) \right]^2}{1 + \hat{P} \text{Var} \left[\sqrt{G^{1+a}} \cdot \mathbf{1}(G \geq t) \right]} \right). \tag{219}$$

Moreover, the expression (209) gives

$$\frac{E_b}{N_0} \Big|_{\min} = \frac{\mathbb{E}[G^a \cdot \mathbf{1}(G \geq t)]}{\mathbb{E} \left[\sqrt{G^{1+a}} \cdot \mathbf{1}(G \geq t) \right]^2} \log 2. \tag{220}$$

For TCP, TMF, and TCI, we compute the respective

$$\frac{E_b}{N_0} \Big|_{\min} = \frac{\log 2}{\Pr[G \geq t] \mathbb{E} \left[\sqrt{G} \mid G \geq t \right]^2} \tag{221}$$

$$\frac{E_b}{N_0} \Big|_{\min} = \frac{\log 2}{\int_t^\infty p(g) g dg} \tag{222}$$

$$\frac{E_b}{N_0} \Big|_{\min} = \frac{\mathbb{E}[G^{-1} \mid G \geq t]}{\Pr[G \geq t]} \log 2. \tag{223}$$

Again applying Jensen’s inequality to the various functions in (221)–(223), we find that:

- the minimum E_b/N_0 of TMF is smaller (better) than that of TCP and TCI unless there is no fading, or if the minimal E_b/N_0 is $-\infty$;
- the best threshold for TMF is $t = 0$ and the minimal E_b/N_0 is -1.59 dB.

For the optimal policy, the parameters $\alpha(h)$ and $\beta(h)$ in (215) and (216) are constants independent of h , see Remark 62, and the TMMSE policy with $t = 0$ is the GMI-optimal policy.

Remark 65. The TCI channel densities are

$$p(y|a) = \Pr[G < t] \frac{e^{-|y|^2}}{\pi} + \Pr[G \geq t] \frac{e^{-|y - \sqrt{\hat{P}}u|^2}}{\pi}$$

$$p(y) = \Pr[G < t] \frac{e^{-|y|^2}}{\pi} + \Pr[G \geq t] \frac{e^{-|y|^2/(1+\hat{P})}}{\pi(1+\hat{P})}.$$

Remark 66. At high SNR, one might expect that the receiver can estimate $P(s_T)$ precisely even if $s_R = 0$. We show that this is indeed the case for on-off fading by using the $K = 2$ partition (154) of Remark 46. Moreover, the results prove that at high SNR one can approach $I(A; Y)$; see Section 7.3.

Remark 67. For Rayleigh fading, the GMI with $K = 2$ in (154) is helpful for both high and low SNR. For instance, for $S_R = 0$ and TCI, the $K = 2$ GMI approaches the mutual information for $S_R = 1(G \geq t)$ as the SNR increases; see Remark 74 in Section 8.4. We further show that for $S_R = 0$, the TCI policy can achieve a minimal E_b/N_0 of $-\infty$ dB, see (301) in Section 8.4.

$S_R = 1(G \geq t)$: The heuristic policy rates are now (cf. (219) and note the $\Pr[G \geq t]$ term and conditioning)

$$I_1(A; Y|S_R) = \Pr[G \geq t] \log \left(1 + \frac{\hat{P} \mathbb{E}[\sqrt{G^{1+a}}|G \geq t]^2}{1 + \hat{P} \text{Var}[\sqrt{G^{1+a}}|G \geq t]} \right). \tag{224}$$

Moreover, the expression (209) is

$$\frac{E_b}{N_0} \Big|_{\min} = \frac{\mathbb{E}[G^a|G \geq t]}{\mathbb{E}[\sqrt{G^{1+a}}|G \geq t]^2} \log 2. \tag{225}$$

For TCP, TMF, and TCI we compute the respective

$$\frac{E_b}{N_0} \Big|_{\min} = \frac{\log 2}{\mathbb{E}[\sqrt{G}|G \geq t]^2} \tag{226}$$

$$\frac{E_b}{N_0} \Big|_{\min} = \frac{\log 2}{\mathbb{E}[G|G \geq t]} \tag{227}$$

$$\frac{E_b}{N_0} \Big|_{\min} = \mathbb{E}[G^{-1}|G \geq t] \log 2. \tag{228}$$

Again applying Jensen’s inequality to the various functions in (226)–(228), we find that:

- the minimum E_b/N_0 of all policies can be better than -1.59 dB by choosing $t > 0$;
- the minimum E_b/N_0 of TMF is smaller (better) than that of TCP and TCI unless there is no fading or the minimal E_b/N_0 is $-\infty$.

For the optimal policy, Remark 62 points out that $\alpha(h)$ and $\beta(h)$ depend on s_R only. We compute

$$\sqrt{P(h)} = \begin{cases} \frac{\alpha_0 |h|}{\lambda + \beta_0 |h|^2}, & g < t \\ \frac{\alpha_1 |h|}{\lambda + \beta_1 |h|^2}, & g \geq t \end{cases} \tag{229}$$

where for $s_R \in \{0, 1\}$ we have

$$\alpha_{s_R} = \frac{\sqrt{\tilde{P}(s_R)}}{\mathbb{E}[|Y|^2|S_R = s_R] - \tilde{P}(s_R)}$$

$$\beta_{s_R} = \frac{\sqrt{\tilde{P}(s_R)}}{[\mathbb{E}[|Y|^2|S_R = s_R] - \tilde{P}(s_R)] \mathbb{E}[|Y|^2|S_R = s_R]}.$$

Remark 68. The GMI (224) for TCI ($a = -1$) is the mutual information $I(A; Y|S_R)$. To see this, observe that the model $q(y|a, s_R)$ has

$$q(y|a, 0) = \frac{e^{-|y|^2}}{\pi}, \quad q(y|a, 1) = \frac{e^{-|y - \sqrt{\tilde{P}} u|^2}}{\pi}$$

and thus we have $q(y|a, s_R) = p(y|a, s_R)$ for all y, a, s_R .

6.6. Partial CSIR, CSIT@R

Suppose next that S_R is a noisy function of H (see for instance (172)) and $S_T = f_T(S_R)$. The capacity is given by (147) and we compute

$$I(X; Y|S_R) = E \left[\log \frac{p(Y|X, S_R)}{p(Y|S_R)} \right] \tag{230}$$

where writing $s_T = f_T(s_R)$ we have

$$p(y|s_R, x) = \int_{\mathbb{C}} p(h|s_R) \frac{e^{-|y-hx(s_T)|^2}}{\pi} dh \tag{231}$$

$$p(y|s_R) = \int_{\mathbb{C}^2} p(h|s_R) p(x(s_T)) \frac{e^{-|y-hx(s_T)|^2}}{\pi} dx(s_T) dh. \tag{232}$$

For example, if $X(s_T)$ is CSCG with variance $P(s_T)$ then

$$p(y|s_R) = \int_{\mathbb{C}} p(h|s_R) \frac{\exp\left(-\frac{|y|^2}{1+|h|^2P(s_T)}\right)}{\pi(1+|h|^2P(s_T))} dh. \tag{233}$$

One can compute $I(X; Y|S_R)$ numerically using (231) and (232). However, optimizing over $X(s_T)$ is usually difficult.

For the reverse model GMI (139), the averaging density in (164) and (165) is now (cf. (194) and (201))

$$p(h, s_T|y, s_R) = \delta(s_T - f_T(s_R)) \frac{p(h|s_R) p(y|h, s_R)}{p(y|s_R)}. \tag{234}$$

We use numerical integration to compute the rates.

The forward model GMI again gives more insight. Define the channel gain and variance as the respective

$$\tilde{g}(s_R) = |E[H|S_R = s_R]|^2 \tag{235}$$

$$\tilde{\sigma}^2(s_R) = \text{Var}[H|S_R = s_R]. \tag{236}$$

Theorem 3. An achievable rate for AWGN fading channels (159) with power constraint $E[|X|^2] \leq P$ and with partial CSIR S_R and $S_T = f_T(S_R)$ is

$$I_1(X; Y|S_R) = E \left[\log \left(1 + \frac{\tilde{g}(S_R)P(S_T)}{1 + \tilde{\sigma}^2(S_R)P(S_T)} \right) \right] \tag{237}$$

where $E[P(S_T)] = P$. The optimal power levels $P(s_T)$ are obtained by solving

$$\lambda = \int_{\mathbb{R}} p(s_R|s_T) \frac{\tilde{g}(s_R)}{[1 + (\tilde{g}(s_R) + \tilde{\sigma}^2(s_R))P(s_T)][1 + \tilde{\sigma}^2(s_R)P(s_T)]} ds_R. \tag{238}$$

In particular, if S_T determines S_R (CSIR@T) then we have the quadratic waterfilling expression

$$f\left(P(s_T), \tilde{g}(s_R), \tilde{\sigma}^2(s_R)\right) = \left(\frac{1}{\lambda} - \frac{1}{\tilde{g}(s_R)} \right)^+ \tag{239}$$

where

$$f(Q, g, \sigma^2) = \left(1 + 2\frac{\sigma^2}{g} \right) Q + \left(1 + \frac{\sigma^2}{g} \right) \sigma^2 Q^2 \tag{240}$$

and where λ is chosen so that $E[P(H_R)] = P$.

Proof. Apply Theorem 1 with

$$\tilde{P}(s_R) = \tilde{g}(s_R)P(s_T) \tag{241}$$

$$E\left[|Y|^2|S_R = s_R\right] = 1 + \left(\tilde{g}(s_R) + \tilde{\sigma}^2(s_R)\right)P(s_T) \tag{242}$$

to obtain (237). To optimize the power levels $P(s_T)$ with (146), consider the derivatives

$$\tilde{P}(s_R)' = 2\tilde{g}(s_R)\sqrt{P(s_T)}1_{(s_T = f_T(s_R))} \tag{243}$$

$$E\left[|Y|^2|S_R = s_R\right]' = 2\left(\tilde{g}(s_R) + \tilde{\sigma}^2(s_R)\right)\sqrt{P(s_T)}1_{(s_T = f_T(s_R))}. \tag{244}$$

The expression (146) thus becomes (238). If S_T determines S_R then the expression simplifies to

$$\lambda = \frac{\tilde{g}(s_R)}{[1 + (\tilde{g}(s_R) + \tilde{\sigma}^2(s_R))P(s_T)][1 + \tilde{\sigma}^2(s_R)P(s_T)]}$$

from which we obtain (239). □

Remark 69. The optimal power control policy with CSIT@R and CSIR@T can be written explicitly by solving the quadratic in (239). The result is

$$P(s_T) = \frac{\tilde{g} + 2\tilde{\sigma}^2}{2\tilde{\sigma}^2(\tilde{g} + \tilde{\sigma}^2)} \left[\sqrt{1 + 4\tilde{\sigma}^2 \left(\frac{1}{\lambda} - \frac{1}{\tilde{g}}\right) + \frac{\tilde{g}(\tilde{g} + \tilde{\sigma}^2)}{(\tilde{g} + 2\tilde{\sigma}^2)^2}} - 1 \right] \tag{245}$$

where we have discarded the dependence on s_R for convenience. The alternative form (239) relates to the usual waterfilling where the left-hand side of (239) is $P(s_T)$. Observe that $\tilde{\sigma}^2 = 0$ gives conventional waterfilling.

Remark 70. As in Section 3.3, we show that at high SNR the $K = 2$ GMI of Remark 42 approaches the upper bound of Proposition 2 in some cases; see Section 7.4. The channel parameters depend on s_R , and we choose $h_1(s_R) = 0$ and $\sigma_1^2(s_R) = \sigma_2^2(s_R) = 1$ for all s_R .

7. On-Off Fading

Consider again on-off fading with $P_G(0) = P_G(2) = 1/2$. We study the scenarios listed in Table 1. The case of no CIR and no CSIT was studied in Section 3.3.

7.1. Full CSIR, CSIT@R

Consider $S_R = H$. The capacity with $B = 0$ (no CSIT) is given by (175) (cf. (73)):

$$C(P) = \frac{1}{2} \log(1 + 2P) \tag{246}$$

and the wideband values are given by (177) (cf. (74)); the minimal E_b/N_0 is $\log 2$ and the slope is $S = 1$.

The capacity with $B = \infty$ (or $S_T = G$) increases to

$$C(P) = \frac{1}{2} \log(1 + 4P) \tag{247}$$

where $P(0) = 0$ and $P(2) = 2P$. This capacity is also achieved with $B = 1$ since there are only two values for G . We compute $C'(0) = 2$ and $C''(0) = -8$, and therefore

$$\frac{E_b}{N_0} \Big|_{\min} = \frac{\log 2}{2}, \quad S = 1. \tag{248}$$

The power gain due to CSIT compared to no fading is thus 3.01 dB, but the capacity slope is the same. The rate curves are compared in Figure 4.

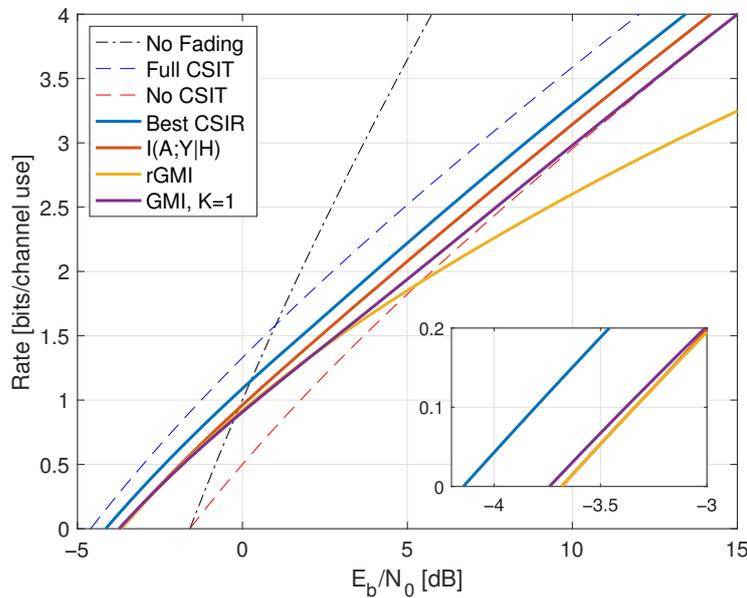


Figure 4. Rates for on-off fading with full CSIR and partial CSIT with noise parameter $\epsilon = 0.1$. The curve “Best CSIR” shows the capacity with $S_R = H\sqrt{P(S_T)}$. The curves for $I(A;Y|H)$, the reverse model GMI (rGMI), and the forward model GMI (GMI, $K = 1$) are for $S_R = H$ with CSCG inputs $X(s_T)$. The $I(A;Y|H)$ and rGMI curves are indistinguishable in the inset.

7.2. Full CSIR, Partial CSIT

Consider next noisy CSIT with $0 \leq \epsilon \leq \frac{1}{2}$ and

$$\Pr[S_T = G] = \bar{\epsilon}, \quad \Pr[S_T \neq G] = \epsilon.$$

$S_R = H\sqrt{P(S_T)}$: The capacity of Proposition 2 is

$$C(P) = \max_{P(0)+P(2)=2P} \frac{\epsilon}{2} \log(1 + 2P(0)) + \frac{\bar{\epsilon}}{2} \log(1 + 2P(2)). \tag{249}$$

Optimizing the power levels, we have

$$P(0) = \left(2\epsilon P - \frac{\bar{\epsilon} - \epsilon}{2}\right)^+, \quad P(2) = 2P - P(0). \tag{250}$$

Figure 4 shows $C(P)$ for $\epsilon = 0.1$ as the curve labeled “Best CSIR”. For $P \geq (\bar{\epsilon} - \epsilon)/(4\epsilon)$, we compute

$$C(P) = \frac{1}{2} \log(1 + 2P) + \frac{1}{2} [1 - H_2(\epsilon)] \log 2 \tag{251}$$

where $H_2(\epsilon) = -\epsilon \log_2 \epsilon - \bar{\epsilon} \log_2 \bar{\epsilon}$ is the binary entropy function. For example, if $\epsilon = 0.1$ then for $P \geq 2$ one gains $\Delta C = [1 - H_2(0.1)]/2 \approx 0.27$ bits over the capacity without CSIT. This translates to an SNR gain of $2\Delta C \cdot 10 \log_{10}(2) \approx 1.60$ dB. On the other hand, for $P \leq (\bar{\epsilon} - \epsilon)/(4\epsilon)$ we have $P(0) = 0, P(2) = 2P$, and the capacity is

$$C(P) = \frac{\bar{\epsilon}}{2} \log(1 + 4P). \tag{252}$$

We have $C'(0) = 2\bar{\epsilon}$ and lose a fraction of $\bar{\epsilon}$ of the power as compared to having full CSIT ($\epsilon = 0$). For example, if $\epsilon = 0.1$, the minimal E_b/N_0 is approximately -4.14 dB.

$S_R = H$: To compute $I(A; Y|H)$ in (190), we write (191) and (193) for CSCG $X(s_T)$ as

$$\begin{aligned}
 p_{Y|A,H}(y|a, 0) &= p_{Y|H}(y|0) = \frac{e^{-|y|^2}}{\pi} \\
 p_{Y|A,H}(y|a, \sqrt{2}) &= \epsilon \frac{e^{-|y-\sqrt{2}x(0)|^2}}{\pi} + \bar{\epsilon} \frac{e^{-|y-\sqrt{2}x(\sqrt{2})|^2}}{\pi} \\
 p_{Y|H}(y|\sqrt{2}) &= \epsilon \frac{\exp\left(-\frac{|y|^2}{1+2P(0)}\right)}{\pi(1+2P(0))} + \bar{\epsilon} \frac{\exp\left(-\frac{|y|^2}{1+2P(2)}\right)}{\pi(1+2P(2))}.
 \end{aligned}$$

Figure 4 shows the rates as the curve labeled “ $I(A; Y|H)$ ”. This curve was computed by Monte Carlo integration with $P(0) = 0.1 \cdot P$ and $P(2) = 1.9 \cdot P$, which is near-optimal for the range of SNRs depicted.

The reverse model GMI (139) requires $\text{Var}[U|Y, H]$. We show how to compute this variance in Appendix C.2 by applying (164) and (165). Figure 4 shows the GMIs as the curve labeled “rGMI”, where we used the same power levels as for the $I(A; Y|H)$ curve. The two curves are indistinguishable for small P , but the “rGMI” rates are poor at large P . This example shows that the forward model GMI with optimized powers can be substantially better than the reverse model GMI with a reasonable but suboptimal power policy.

The forward model GMI (195) is

$$I_1(A; Y|H) = \frac{1}{2} \log\left(1 + \text{SNR}(\sqrt{2})\right) \tag{253}$$

where $\text{SNR}(\sqrt{2})$ is given by (196) with

$$\begin{aligned}
 \tilde{P}_T(\sqrt{2}) &= \left(\epsilon\sqrt{P(0)} + \bar{\epsilon}\sqrt{P(2)}\right)^2 \\
 \text{Var}\left[\sqrt{P(s_T)} \middle| H = h\right] &= 1 + 2\epsilon\bar{\epsilon}\left(\sqrt{P(2)} - \sqrt{P(0)}\right)^2.
 \end{aligned}$$

Applying Remark 61, the optimal power control policy is

$$\sqrt{P(s_T)} = \frac{p_{H|s_T}(\sqrt{2}|s_T)}{\gamma + \beta p_{H|s_T}(\sqrt{2}|s_T)} = \begin{cases} \frac{\epsilon}{\gamma + \beta\epsilon}, & s_T = 0 \\ \frac{\bar{\epsilon}}{\gamma + \beta\bar{\epsilon}}, & s_T = 2 \end{cases} \tag{254}$$

where

$$\beta = \frac{2\sqrt{\tilde{P}_T(\sqrt{2})}}{\mathbb{E}[|Y|^2|H = \sqrt{2}]} \tag{255}$$

and $\gamma \geq 0$ is chosen so that $P(0) + P(2) = 2P$. Figure 4 shows the resulting GMI as the curve labeled “GMI, $K = 1$ ”. At low SNR, we achieve the rate $\tilde{P}_T(\sqrt{2})$ and the optimal power control has $\beta \rightarrow 0$ so that

$$P(0) = \frac{2P\epsilon^2}{\epsilon^2 + \bar{\epsilon}^2}, \quad P(2) = \frac{2P\bar{\epsilon}^2}{\epsilon^2 + \bar{\epsilon}^2} \tag{256}$$

and therefore

$$\tilde{P}_T(\sqrt{2}) = 2(\epsilon^2 + \bar{\epsilon}^2)P. \tag{257}$$

We have $C'(0) = 2(\epsilon^2 + \bar{\epsilon}^2)$ and lose a fraction of $(\epsilon^2 + \bar{\epsilon}^2)$ of the power as compared to having full CSIT ($\epsilon = 0$). For example, if $\epsilon = 0.1$, the minimal E_b/N_0 is approximately -3.74 dB.

We remark that the $I(A; Y|H)$ and reverse model GMI curves lie above the forward model curve if we choose the same power policy as for the forward channel.

7.3. Partial CSIR, Full CSIT

This section studies $S_T = H$. The capacity with partial CSIR is given by (138) for which we need to compute $p(y|a, s_R)$ and $p(y|s_R)$. We consider two cases.

$S_R = 1 (G \geq t)$: Here we recover the case with full CSIR by choosing t to satisfy $0 < t \leq 2$.

$S_R = 0$: The best power policy clearly has $P(0) = 0$ and $P(\sqrt{2}) = 2P$. The mutual information is thus $I(A; Y) = I(X(\sqrt{2}); Y)$ and the channel densities are (cf. (75) and (76))

$$p(y|a) = \frac{e^{-|y|^2}}{2\pi} + \frac{e^{-|y-2\sqrt{P}u(\sqrt{2})|^2}}{2\pi}$$

$$p(y) = \frac{e^{-|y|^2}}{2\pi} + \frac{e^{-|y|^2/(1+4P)}}{2\pi(1+4P)}.$$

The rates $I(A; Y)$ are shown in Figure 5. Observe that the low-SNR rates are larger than without fading; this is a consequence of the slightly bursty nature of transmission.

The reverse model GMI (139) requires $\text{Var}[U|Y]$. We compute this variance in Appendix C.3 by using (164) and (165) with (201) and $\phi(s_T) = 0$. Figure 5 shows the GMIs as the curve labeled “rGMI”.

Next, the TCP, TMF, TCI, and TMMSE policies are the same for $0 < t \leq 2$, since they use $P(0) = 0$ and $P(\sqrt{2}) = 2P$. The resulting rate is given by (202)–(204) with $\tilde{P}(0) = 0$, $\tilde{P}(1) = P$, and $\text{Var}[\sqrt{GP(S_T)} | S_R = 1] = P$ and

$$I_1(A; Y) = \log\left(1 + \frac{P}{1+P}\right). \tag{258}$$

The rates are plotted in Figure 5 as the curve labeled “GMI, $K = 1$ ”. This example again shows that choosing $K = 1$ is a poor choice at high SNR.

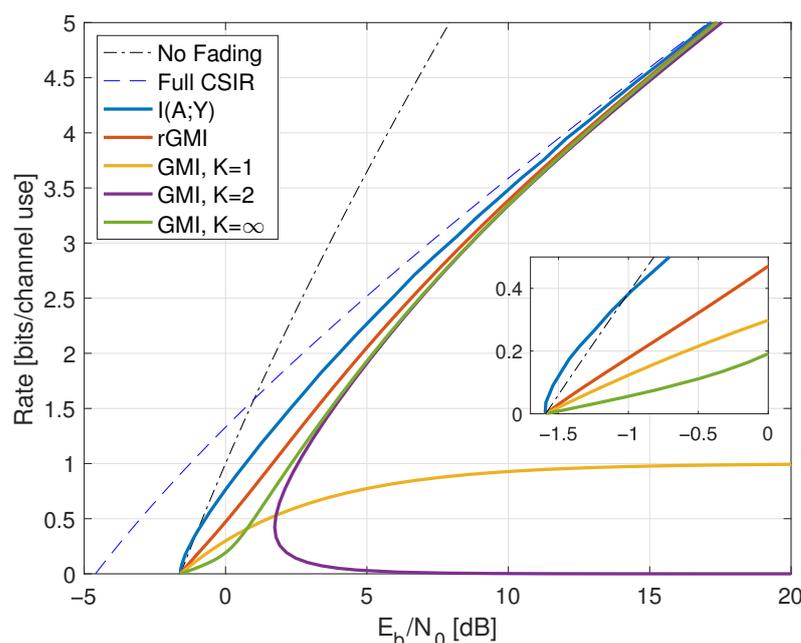


Figure 5. Rates for on-off fading with $S_T = H$ and $S_R = 0$. The GMI for the $K = 2$ partition uses the threshold $t_R = \sqrt{P} + 3$.

To improve the auxiliary model at high SNR, consider the GMI (154) with $K = 2$ and the subsets (65). We further choose the parameters $h_1 = 0, \sigma_1^2 = 0, h_2 = 2, \sigma_2^2 = 1$, and adaptive coding with $X(0) = 0, X(\sqrt{2}) = \sqrt{2P}U, \bar{X} = \sqrt{P}U$, where $U \sim \mathcal{CN}(0, 1)$. The GMI (154) is

$$I_1(A; Y) = \Pr[\mathcal{E}_2] \left[\log(1 + 4P) + \frac{\mathbb{E}[|Y|^2 | \mathcal{E}_2]}{1 + 4P} - \mathbb{E} \left[\left| Y - \sqrt{4P}U \right|^2 \middle| \mathcal{E}_2 \right] \right]. \quad (259)$$

In Appendix B.2, we show that choosing $t_R = P^{\lambda_R} + b$ where $0 < \lambda_R < 1$ and b is a real constant makes all terms behave as desired as P increases:

$$\Pr[\mathcal{E}_2] \rightarrow 1/2, \quad \frac{\mathbb{E}[|Y|^2 | \mathcal{E}_2]}{1 + 4P} \rightarrow 1, \quad \mathbb{E} \left[\left| Y - \sqrt{4P}U \right|^2 \middle| \mathcal{E}_2 \right] \rightarrow 1. \quad (260)$$

We thus have

$$\lim_{P \rightarrow \infty} \left[\frac{1}{2} \log(1 + 4P) - I_1(X; Y) \right] = 0. \quad (261)$$

Figure 5 shows the behavior of $I_1(A; Y)$ for $\lambda_R = 1/2$ and $b = 3$ as the curve labeled ‘‘GMI, $K = 2$ ’’. As for the case without CSIT, the receiver can estimate H accurately at large SNR, and one approaches the capacity with full CSIR.

Finally, the large- K forward model rates are computed using (70) but where \bar{X} replaces X . One may again use the results of Appendix C.3 and the relations

$$\begin{aligned} \mathbb{E}[\bar{X} | Y = y] &= \sqrt{P} \mathbb{E}[U | Y = y] \\ \mathbb{E}[|\bar{X}|^2 | Y = y] &= P \mathbb{E}[|U|^2 | Y = y] \\ \text{Var}[\bar{X} | Y = y] &= P \text{Var}[U | Y = y]. \end{aligned}$$

The rates are shown as the curve labeled ‘‘GMI, $K = \infty$ ’’ in Figure 5. So again, the large- K forward model is good at high SNR but worse than the best $K = 1$ model at low SNR.

7.4. Partial CSIR, CSIT@R

Consider partial CSIR with $S_T = S_R$ and

$$\Pr[S_R = H] = \bar{\epsilon}, \quad \Pr[S_R \neq H] = \epsilon \quad (262)$$

where $0 \leq \epsilon \leq \frac{1}{2}$. We thus have both CSIT@R and CSIR@T. To compute $I(X; Y | S_R)$ in (230), we write (231) and (232) as

$$\begin{aligned} p_{Y|S_R, X}(y|0, x) &= \bar{\epsilon} \frac{e^{-|y|^2}}{\pi} + \epsilon \frac{e^{-|y - \sqrt{2}x(0)|^2}}{\pi} \\ p_{Y|S_R, X}(y|\sqrt{2}, x) &= \bar{\epsilon} \frac{e^{-|y - \sqrt{2}x(\sqrt{2})|^2}}{\pi} + \epsilon \frac{e^{-|y|^2}}{\pi} \\ p_{Y|S_R}(y|0) &= \bar{\epsilon} \frac{e^{-|y|^2}}{\pi} + \epsilon \frac{e^{-|y|^2/[1+2P(0)]}}{\pi[1+2P(0)]} \\ p_{Y|S_R}(y|\sqrt{2}) &= \bar{\epsilon} \frac{e^{-|y|^2/[1+2P(\sqrt{2})]}}{\pi[1+2P(\sqrt{2})]} + \epsilon \frac{e^{-|y|^2}}{\pi} \end{aligned}$$

where $X(s_T)$ is CSCG. We choose the transmit powers $P(0)$ and $P(\sqrt{2})$ as in (250) to compare with the best CSIR. Figure 6 shows the resulting rates for $\epsilon = 0.1$ as the curve labeled ‘‘Partial CSIR, $I(X; Y | S_R)$ ’’. Observe that at high SNR, the curve seems to approach

the best CSIR curve from Figure 4 with $S_R = H\sqrt{P(S_T)}$. We prove this by studying a forward model GMI with $K = 2$.

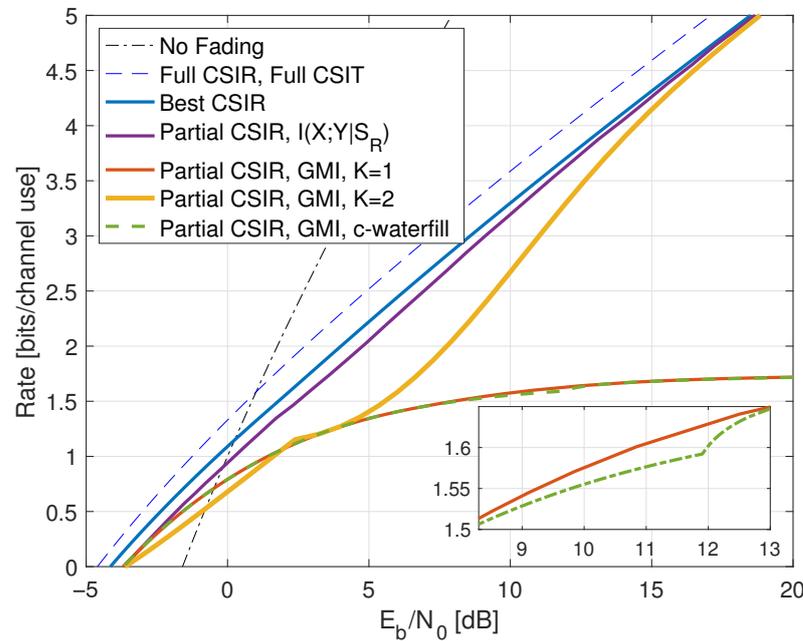


Figure 6. Rates for on-off fading with partial CSIR and CSIT@R. The curve “Best CSIR” shows the capacity with $S_R = H\sqrt{P(S_T)}$. The mutual information $I(X;Y|S_R)$ and the GMI are for $\Pr[S_R \neq H] = 0.1$ and with CSCG inputs $X(s_T)$. The GMI for the $K = 2$ partition uses $t_R = P^{0.4}$. The curve labeled ‘c-waterfill’ shows the conventional waterfilling rates.

The reverse model GMI requires $\text{Var}[U|Y, S_R]$, which can be computed by simulation; see Appendix C.4. However, optimizing the powers seems difficult. We instead focus on the forward model GMI of Theorem 3 for which we compute

$$\tilde{g}(0) = 2\epsilon^2, \quad \tilde{g}(\sqrt{2}) = 2\epsilon\bar{\epsilon}, \quad \tilde{\sigma}^2(0) = \tilde{\sigma}^2(\sqrt{2}) = 2\epsilon\bar{\epsilon}$$

and therefore (237) is

$$I_1(X;Y|S_R) = \frac{1}{2} \log\left(1 + \frac{2\epsilon^2 P(0)}{1 + 2\epsilon\bar{\epsilon} P(0)}\right) + \frac{1}{2} \log\left(1 + \frac{2\bar{\epsilon}^2 P(\sqrt{2})}{1 + 2\epsilon\bar{\epsilon} P(\sqrt{2})}\right). \quad (263)$$

For CSIR@T, the optimal power control policy is given by the quadratic waterfilling specified by (239) or (245):

$$P(0) = \frac{1 + \bar{\epsilon}}{4\epsilon\bar{\epsilon}} \left[\sqrt{1 + 8\epsilon\bar{\epsilon} \left(\frac{1}{\lambda} - \frac{1}{2\epsilon^2}\right)^+ \frac{\epsilon}{(1 + \bar{\epsilon})^2} - 1} \right]$$

$$P(\sqrt{2}) = \frac{1 + \epsilon}{4\epsilon\bar{\epsilon}} \left[\sqrt{1 + 8\epsilon\bar{\epsilon} \left(\frac{1}{\lambda} - \frac{1}{2\bar{\epsilon}^2}\right)^+ \frac{\bar{\epsilon}}{(1 + \epsilon)^2} - 1} \right].$$

The rates are shown in Figure 6 as the curve labeled “Partial CSIR, GMI, $K = 1$ ”. Observe that at high SNR the GMI (263) saturates at

$$\frac{1}{2} \log\left(1 + \frac{\epsilon}{\bar{\epsilon}}\right) + \frac{1}{2} \log\left(1 + \frac{\bar{\epsilon}}{\epsilon}\right). \quad (264)$$

For example, for $\epsilon = 0.1$, we approach 1.74 bits at high SNR. On the other hand, at low SNR, the rate is maximized with $P(0) = 0$ and $P(\sqrt{2}) = 2P$ so that $I_1(X; Y|S_R) \approx 2\bar{\epsilon}^2 P$. We thus achieve a fraction of $\bar{\epsilon}^2$ of the power compared to full CSIT. For example, if $\epsilon = 0.1$, the minimal E_b/N_0 is approximately -3.69 dB.

Figure 6 also shows the conventional waterfilling rates as the curve labeled “Partial CSIR, GMI, c-waterfill”. These rates are almost the same as the quadratic waterfilling rates except for the range of E_b/N_0 between 9 to 13 dB shown in the inset.

To improve the auxiliary model at high SNR, we use a $K = 2$ GMI with (see Remark 70)

$$h_1(s_R) = 0, \quad h_2(s_R) = \sqrt{2}, \quad \sigma_1^2(s_R) = \sigma_2^2(s_R) = 1$$

for $s_R = 0, \sqrt{2}$. The receiver chooses $\bar{X}(s_R) = \sqrt{P(s_R)} U$ (see Remark 41) and we have (see Remark 42)

$$\begin{aligned}
 I_1(X; Y|S_R) = & \frac{1}{2} \Pr[\mathcal{E}_2|S_R = 0] \left\{ \log(1 + 2P(0)) + \frac{\mathbb{E}[|Y|^2|\mathcal{E}_2, S_R = 0]}{1 + 2P(0)} \right. \\
 & \left. - \mathbb{E}[|Y - \sqrt{2}X(0)|^2|\mathcal{E}_2, S_R = 0] \right\} \\
 & + \frac{1}{2} \Pr[\mathcal{E}_2|S_R = \sqrt{2}] \left\{ \log(1 + 2P(\sqrt{2})) + \frac{\mathbb{E}[|Y|^2|\mathcal{E}_2, S_R = \sqrt{2}]}{1 + 2P(\sqrt{2})} \right. \\
 & \left. - \mathbb{E}[|Y - \sqrt{2}X(\sqrt{2})|^2|\mathcal{E}_2, S_R = \sqrt{2}] \right\} \quad (265)
 \end{aligned}$$

where the $X(s_T), s_T \in S_T$, are given by (122). We consider $P(0)$ and $P(\sqrt{2})$ that scale in proportion to P . In this case, Appendix B.3 shows that choosing $t_R = P^{\lambda_R}$ where $0 < \lambda_R < 1$ gives the (best) full-CSIR capacity for large P , which is the rate specified in (249):

$$\lim_{P \rightarrow \infty} \left[\frac{\epsilon}{2} \log(1 + 2P(0)) + \frac{\bar{\epsilon}}{2} \log(1 + 2P(\sqrt{2})) - I_1(X; Y|S_R) \right] = 0. \quad (266)$$

In other words, by optimizing $P(0)$ and $P(\sqrt{2})$, at high SNR the $K = 2$ GMI can approach the capacity of Proposition 2. This is expected since the receiver can estimate $H\sqrt{P(S_T)}$ reliably at high SNR.

Figure 6 shows the behavior of this GMI and $t_R = P^{0.4}$, and where we have chosen $P(0)$ and $P(\sqrt{2})$ according to (250). The abrupt change in slope at approximately 2.5 dB is because $P(0)$ becomes positive beyond this E_b/N_0 . Keeping $P(0) = 0$ for E_b/N_0 up to about 12 dB gives better rates, but for high SNR one should choose the powers according to (250).

8. Rayleigh Fading

Rayleigh fading has $H \sim \mathcal{CN}(0, 1)$. The random variable $G = |H|^2$ thus has the density $p(g) = e^{-g} \cdot 1(g \geq 0)$. Sections 8.1 and 8.2 review known results.

8.1. No CSIR, No CSIT

Suppose $S_R = S_T = 0$ and $X \sim \mathcal{CN}(0, P)$. The densities to compute $I(X; Y)$ for CSCG X are

$$p(y|x) = \frac{e^{-|y|^2/(|x|^2+1)}}{\pi(|x|^2 + 1)} \quad (267)$$

$$p(y) = \int_0^\infty \frac{e^{-g/P}}{P} \frac{e^{-|y|^2/(g+1)}}{\pi(g + 1)} dg. \quad (268)$$

The minimum E_b/N_0 is approximately 9.2 dB, and the forward model GMI (174) is zero. The capacity is achieved by discrete and finite X [96], and at large SNR, the capacity behaves as $\log \log P$ [97]. Further results are derived in [98–102].

8.2. Full CSIR, CSIT@R

The capacity (175) for $B = 0$ (no CSIT) is

$$C(P) = \int_0^\infty e^{-g} \log(1 + gP) dg = e^{1/P} E_1(1/P) \log(e) \tag{269}$$

where the exponential integral $E_1(\cdot)$ is given by (A4) below. The wideband values are given by (177):

$$\left. \frac{E_b}{N_0} \right|_{\min} = \log 2, \quad S = 1.$$

The minimal E_b/N_0 is -1.59 dB, but the fading reduces the capacity slope. At high SNR, we have

$$C(P) \approx \log(P) - \gamma$$

where $\gamma \approx 0.57721$ is Euler’s constant. The capacity thus behaves as for the case without fading but with an SNR loss of approximately 2.5 dB.

The capacity (182) with $B = \infty$ (or $S_T = G$) is (see ([95], Equation (7)))

$$C(P) = \int_\lambda^\infty e^{-g} \log(g/\lambda) dg = E_1(\lambda). \tag{270}$$

where $P(g)$ is given by (181) and λ is chosen so that

$$P = \int_\lambda^\infty e^{-g} P(g) dg = \frac{e^{-\lambda}}{\lambda} - E_1(\lambda).$$

At low SNR we have large λ and using the approximation (A7) below we compute

$$C(P) \approx e^{-\lambda}/\lambda \text{ and } P \approx e^{-\lambda}/\lambda^2. \tag{271}$$

We thus have $E_b/N_0 \approx \log(2)/\lambda$ and the minimal E_b/N_0 is $-\infty$.

Consider now $B = 1$ for which $P_{S_T}(3\Delta/2) = e^{-\Delta}$ and

$$E[G|G \geq \Delta] = 1 + \Delta \tag{272}$$

$$E[G^2|G \geq \Delta] = 2 + 2\Delta + \Delta^2. \tag{273}$$

We thus have the wideband quantities in (186) and (187):

$$\left. \frac{E_b}{N_0} \right|_{\min} = \frac{\log 2}{1 + \Delta} \tag{274}$$

$$S = \frac{2e^{-\Delta}(1 + \Delta)^2}{2 + 2\Delta + \Delta^2}. \tag{275}$$

Figure 7 shows the capacities for $B = 1$ and $\Delta = 1, 2, 1/2$. The minimum E_b/N_0 value is

$$-1.59 \text{ dB} - 10 \log_{10}(1 + \Delta) \tag{276}$$

and for $\Delta = 1, 2, 1/2$ we gain 3 dB, 4.8 dB, 1.8 dB, respectively, over no CSIT at low power. Note that one bit of feedback allows one to approach the full CSIT rates closely.

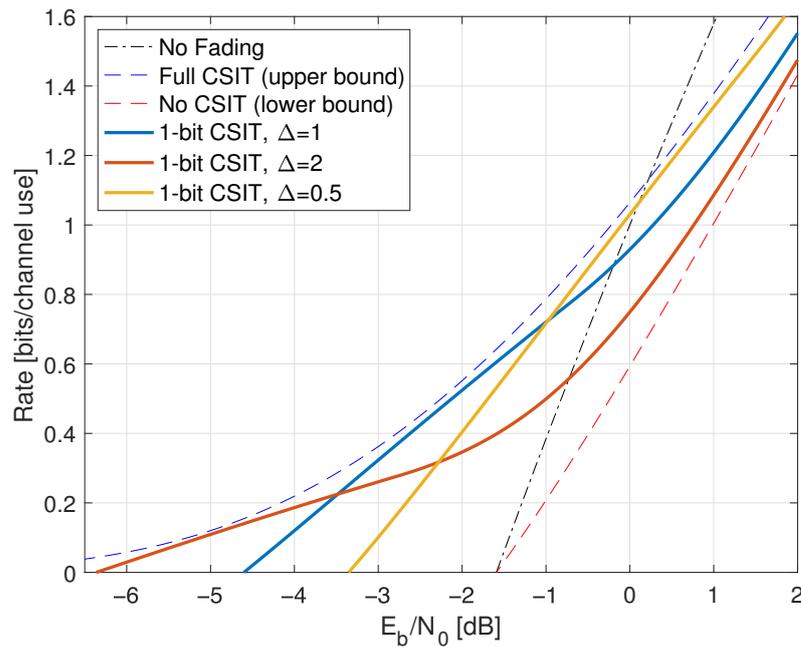


Figure 7. Capacities for Rayleigh fading with full CSIR, a one-bit quantizer with threshold Δ , and CSIT@R.

Remark 71. For the scalar channel (159), knowing H at both the transmitter and receiver provides significant gains at low SNR [73] but small gains at high SNR ([95], Figure 4) as compared to knowing H at the receiver only. Furthermore, the reliability can be improved ([78], Figures 5–7). Significant gains are also possible for MIMO channels.

Remark 72. An alternative way to derive (272)–(275) is as follows. Define $\hat{P} = Pe^\Delta$ so for small P the capacity is

$$\begin{aligned} C(P) &= \int_{\Delta}^{\infty} e^{-g} \log(1 + g \hat{P}) dg \\ &= e^{1/\hat{P}} E_1\left(\frac{1}{\hat{P}} + \Delta\right) + e^{-\Delta} \log(1 + \hat{P}\Delta) \\ &\approx P(1 + \Delta) - \frac{1}{2}P^2e^\Delta(2 + 2\Delta + \Delta^2). \end{aligned}$$

8.3. Full CSIR, Partial CSIT

Consider noisy CSIT with

$$\Pr[S_T = 1(G \geq \Delta)] = \bar{\epsilon}, \quad \Pr[S_T \neq 1(G \geq \Delta)] = \epsilon.$$

We begin with the most informative CSIR.

$S_R = \sqrt{P(S_T)}H$: Proposition 2 gives the capacity

$$\begin{aligned} C(P) &= \int_0^{\infty} e^{-g} \sum_{s_T} P(s_T|g) \log(1 + g P(s_T)) dg \\ &= \int_0^{\Delta} e^{-g} [\bar{\epsilon} \log(1 + g P(0)) + \epsilon \log(1 + g P(1))] dg \\ &\quad + \int_{\Delta}^{\infty} e^{-g} [\bar{\epsilon} \log(1 + g P(1)) + \epsilon \log(1 + g P(0))] dg. \end{aligned} \tag{277}$$

It remains to optimize $P(0)$, $P(1)$ and Δ . The two equations for the Lagrange multiplier λ are

$$\lambda \cdot P_{S_T}(0) = \int_0^\Delta e^{-g} \cdot \frac{\bar{\epsilon} g}{1 + gP(0)} dg + \int_\Delta^\infty e^{-g} \cdot \frac{\epsilon g}{1 + gP(0)} dg \tag{278}$$

$$\lambda \cdot P_{S_T}(1) = \int_0^\Delta e^{-g} \cdot \frac{\epsilon g}{1 + gP(1)} dg + \int_\Delta^\infty e^{-g} \cdot \frac{\bar{\epsilon} g}{1 + gP(1)} dg \tag{279}$$

where $P_{S_T}(0) = \bar{\epsilon} - (\bar{\epsilon} - \epsilon)e^{-\Delta}$ and $P_{S_T}(1) = \epsilon + (\bar{\epsilon} - \epsilon)e^{-\Delta}$. The rates are shown in Figure 8.

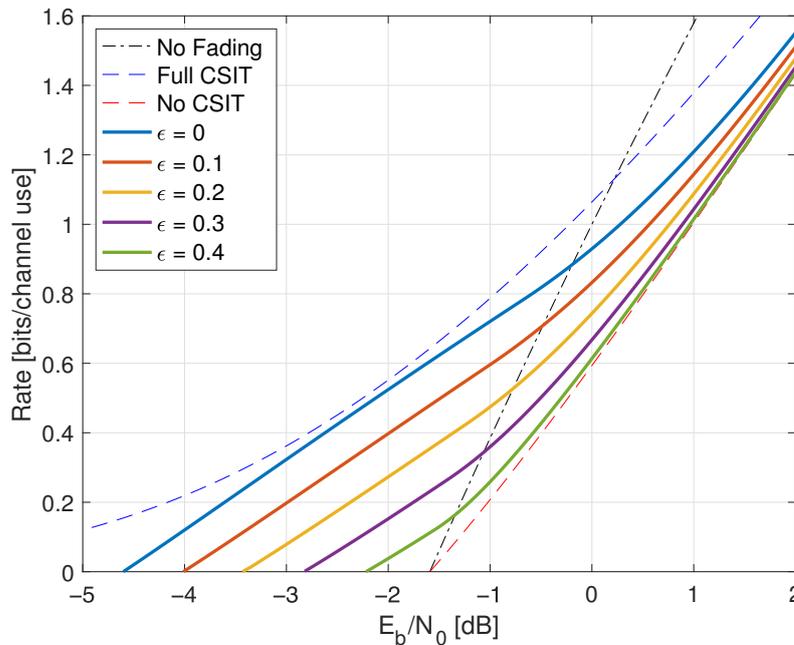


Figure 8. Capacities for Rayleigh fading, $S_R = \sqrt{P(S_T)}H$, and a one-bit quantizer with threshold $\Delta = 1$, and various CSIT error probabilities ϵ .

For fixed Δ and large P , we have $1/\lambda \approx P(0) \approx P(1) \approx P$ and approach the capacity (269) without CSIT. In contrast, for small P we may use similar steps as for (183) and (184). Observe the following for (278) and (279):

- both $P(0)$ and $P(1)$ decrease as λ increases;
- the maximal λ in (278) is obtained with $P(0) = 0$; this value is

$$E[G|S_T = 0] = \frac{\bar{\epsilon} - (\bar{\epsilon} - \epsilon)(1 + \Delta) e^{-\Delta}}{P_{S_T}(0)} \tag{280}$$

- the maximal λ in (279) is obtained with $P(1) = 0$; this value is

$$E[G|S_T = 1] = \frac{\epsilon + (\bar{\epsilon} - \epsilon)(1 + \Delta) e^{-\Delta}}{P_{S_T}(1)}. \tag{281}$$

Thus, if $E[G|S_T = 0] < E[G|S_T = 1]$ and $0 \leq \epsilon < 1/2$, then for P below some threshold we have $P(0) = 0, P(1) = P/P_{S_T}(1)$ and the capacity is

$$C(P) = \int_0^\Delta e^{-g} \epsilon \log\left(1 + \frac{gP}{P_{S_T}(1)}\right) dg + \int_\Delta^\infty e^{-g} \bar{\epsilon} \log\left(1 + \frac{gP}{P_{S_T}(1)}\right) dg. \tag{282}$$

We compute $C'(0) = E[G|S_T = 1]$ which is given by (281) so that $1 \leq C'(0) \leq 1 + \Delta$, as expected from (274). For example, for $\epsilon = 0.1$ and $\Delta = 1$ we have $C'(0) \approx 1.75$ and therefore the minimal E_b/N_0 is approximately -4.01 dB.

The best Δ is the unique solution $\hat{\Delta}$ of the equation

$$e^{-\Delta} = \frac{\epsilon}{\bar{\epsilon} - \epsilon}(\Delta - 1) \tag{283}$$

and the result is $C'(0) = \hat{\Delta} \geq 1$. We have the simple bounds

$$1 + \frac{1}{2} \log\left(\frac{1}{\epsilon} - 2\right) \leq C'(0) \leq 1 + \frac{1}{e} \left(\frac{1}{\epsilon} - 2\right) \tag{284}$$

where the left inequality follows by taking logarithms and using $\log(\Delta - 1) \leq \Delta - 2$, and the right inequality follows by using $e^{-\Delta} \leq e^{-1}$ in (283). For example, for $\epsilon \rightarrow 0$ we have $C'(0) \rightarrow \infty$, and for $\epsilon \rightarrow 1/2$ we have $C'(0) \rightarrow 1$.

$S_R = H$: For the less informative CSIR, one may use (191) and (193) to compute $I(A; Y|H)$. The reverse model GMI requires $\text{Var}[U|Y, S_R]$, which can be computed by simulation; see Appendix C.2. Again, however, optimizing the powers seems difficult. We instead focus on the forward model GMI of Corollary 1, which is

$$I_1(A; Y|H) = \int_0^\infty e^{-g} \log(1 + \text{SNR}(g)) dg \tag{285}$$

where

$$\text{SNR}(g) = \frac{g\tilde{P}_T(g)}{1 + g\epsilon\bar{\epsilon}(\sqrt{P(0)} - \sqrt{P(1)})^2} \tag{286}$$

and

$$\tilde{P}_T(g) = \begin{cases} (\bar{\epsilon}\sqrt{P(0)} + \epsilon\sqrt{P(1)})^2, & g < \Delta \\ (\epsilon\sqrt{P(0)} + \bar{\epsilon}\sqrt{P(1)})^2, & g \geq \Delta. \end{cases} \tag{287}$$

It remains to optimize $P(0)$, $P(1)$ and Δ . Computing the derivatives seems complicated, so we use numerical optimization for fixed $\Delta = 1$ as in Figure 8. The results are shown in Figure 9. For fixed Δ and large P , it is best to choose $P(0) \approx P(1)$ so that $\text{SNR}(g) \approx gP$ and we approach the rate of no CSIT. For small P , however, the best $P(0)$ is no longer zero and $C'(0)$ is smaller than (281).

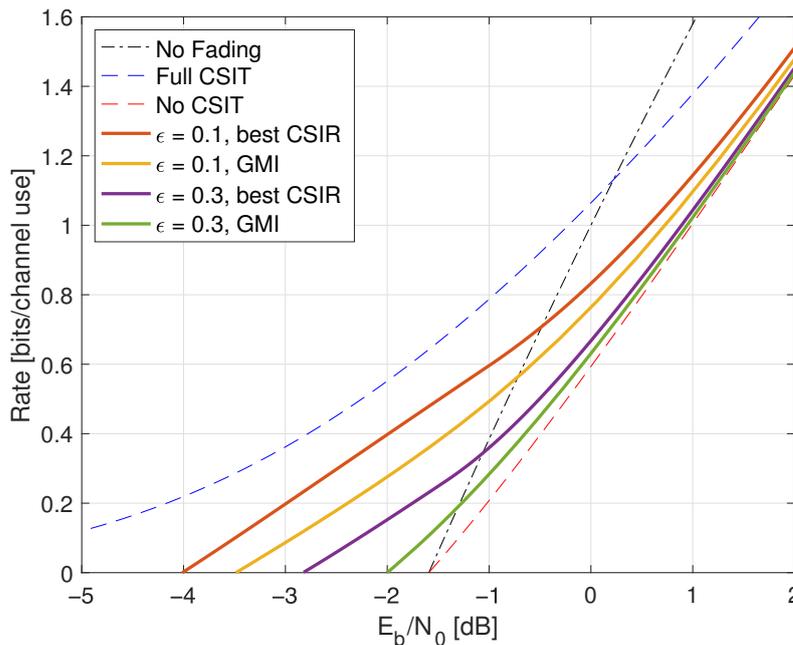


Figure 9. Rates for Rayleigh fading, $S_R = H$ and $S_R = H\sqrt{P(S_T)}$, a one-bit quantizer with threshold $\Delta = 1$, and various ϵ . The curves labeled “best CSIR” show the capacities with $S_R = H\sqrt{P(S_T)}$. The curves labeled “GMI” show the rates (285) for the optimal powers $P(0)$ and $P(1)$.

8.4. Partial CSIR, Full CSIT

Consider $S_T = H$ and suppose we choose the $X(h)$ to be jointly CSCG with variances $E[|X(h)|^2] = P(h)$ and correlation coefficients

$$\rho(h, h') = \frac{E[X(h)X(h')^*]}{\sqrt{P(h)P(h')}}$$

and where $E[P(H)] \leq P$. We then have

$$p(y|s_R) = \int_{\mathbb{C}} p(h|s_R) \frac{e^{-|y|^2/(|h|^2P(h)+1)}}{\pi(|h|^2P(h)+1)} dh.$$

As in (98), $p(y|s_R)$ and $h(Y|S_R)$ depend only on the marginals of A and not on the $\rho(h, h')$. We thus have the problem of finding the $\rho(h, h')$ that minimize

$$h(Y|A, S_R) = \int_{\mathcal{A}} p(a) h(Y|S_R, A = a) da.$$

We will use fully-correlated $X(h)$ as discussed in Section 6.5. We again consider $S_R = 0$ and $S_R = 1(G \geq t)$.

$S_R = 0$: For the heuristic policies, the power (206) is

$$\hat{P} = \frac{P}{\Gamma(1+a, t)} \tag{288}$$

and the rate (219) is

$$I_1(A; Y) = \log \left(1 + \frac{P \Gamma(\frac{3+a}{2}, t)^2}{\Gamma(1+a, t) + P [\Gamma(2+a, t) - \Gamma(\frac{3+a}{2}, t)^2]} \right) \tag{289}$$

where $\Gamma(s, x)$ is the upper incomplete gamma function; see Appendix A.3. Moreover, the expression (220) is

$$\frac{E_b}{N_0} \Big|_{\min} = \frac{\Gamma(1+a, t)}{\Gamma(\frac{3+a}{2}, t)^2} \cdot \log 2. \tag{290}$$

We remark that $\Gamma(s, 0) = \Gamma(s)$ where $\Gamma(x)$ is the gamma function. We further have

$$\begin{aligned} \Gamma(0, t) &= E_1(t), & \Gamma(1, t) &= e^{-t}, \\ \Gamma(2, t) &= e^{-t}(t+1), & \Gamma(3, t) &= e^{-t}(t^2 + 2t + 2). \end{aligned}$$

For example, the TCP policy ($a = 0$) has $\hat{P} = P e^t$. At low SNR, it turns out that the best choice is $t = 0.283$ for which we have $\Gamma(1, t)/\Gamma(3/2, t)^2 \approx 1.174$. The minimum E_b/N_0 in (222) is thus -0.90 dB. At high SNR, the best choice is $t = 0$ so that (289) with $\Gamma(3/2, 0) = \Gamma(3/2) = \sqrt{\pi}/2$ gives

$$I_1(A; Y) = \log \left(1 + \frac{P \pi/4}{1 + P(1 - \pi/4)} \right). \tag{291}$$

The TCP rate thus saturates at 2.22 bits per channel use; see the curve labeled ‘‘TCP, GMI, $K = 1$ ’’ in Figure 10.

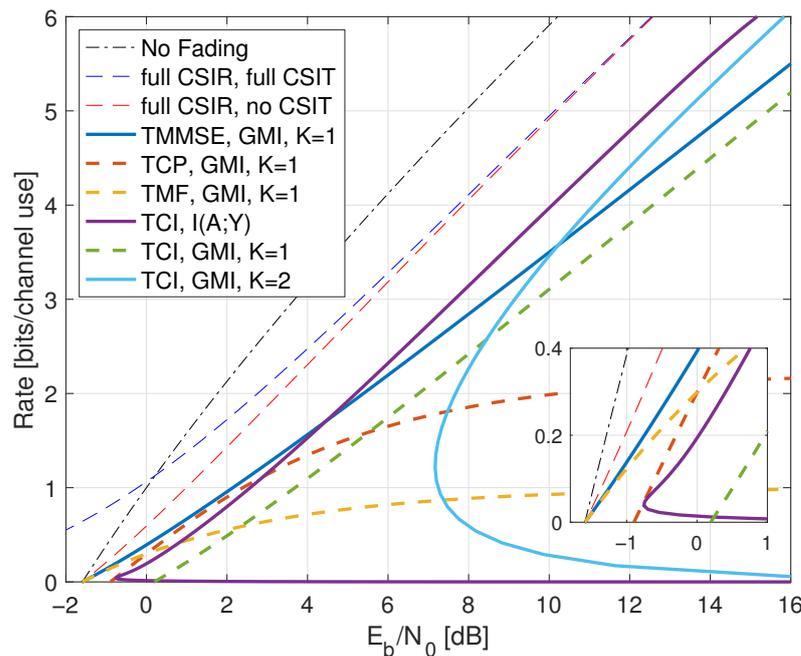


Figure 10. Rates for Rayleigh fading with $S_T = H$ and $S_R = 0$. The threshold t was optimized for the $K = 1$ curves, while $t = P^{-0.4}$ for the $I(A;Y)$ and $K = 2$ curves. The $K = 2$ GMI uses $t_R = P^{0.4}$.

The TMF policy ($a = 1$) has $\hat{P} = P e^t / (t + 1)$. The best choice is $t = 0$ for which we have $\Gamma(2) = 1$ and $\Gamma(3) = 2$ and therefore (289) is

$$I_1(A;Y) = \log\left(1 + \frac{P}{1 + P}\right). \tag{292}$$

The minimum E_b/N_0 in (222) is -1.59 dB, and at high SNR, the TMF rate saturates at 1 bit per channel use. The rates are shown as the curve labeled “TMF, GMI, $K = 1$ ” in Figure 10.

The TCI policy ($a = -1$) has $\hat{P} = P/E_1(t)$ and using $\Gamma(0, t) = E_1(t)$ and $\Gamma(1, t) = e^{-t}$ gives

$$I_1(A;Y) = \log\left(1 + \frac{P}{e^{2t} E_1(t) + P(e^t - 1)}\right). \tag{293}$$

The minimum E_b/N_0 in (290) is

$$\left.\frac{E_b}{N_0}\right|_{\min} = E_1(t) e^{2t} \cdot \log 2. \tag{294}$$

Optimizing over t by taking derivatives (see (A5) below), the best t satisfies the equation $2te^t E_1(t) = 1$ which gives $t \approx 0.61$ and the minimal E_b/N_0 is approximately 0.194 dB. On the other hand, for large SNR, we may choose $t = 1/P$ and using $E_1(t) \approx \log(1/t)$ for small t gives

$$I_1(A;Y) \approx \log\left(1 + \frac{P}{1 + \log P}\right).$$

Since the pre-log is at most 1, the capacity grows with pre-log 1 for large P . We see that TMF is best at small P while TCI is best at large P . The rates are shown as the curve labeled “TCI, GMI, $K = 1$ ” in Figure 10.

The simple channel output of TCI permits further analysis. Using Remark 65, we compute the mutual information $I(A;Y)$ by numerical integration; see the curve labeled “TCI, $I(A;Y)$ ” in Figure 10. We see that at high SNR, the TCI mutual information is larger

than the GMI for TCP, TMF, and (of course) TCI. Moreover, as we show, the TCI mutual information can work well at low SNR.

Motivated by Section 7.3 and Figure 5, we again use the GMI (154) with $K = 2$ and (65). We further choose $h_1 = 0, \sigma_1^2 = \sigma_2^2 = 1$, and

$$\bar{X} = \frac{\sqrt{\hat{P}}}{h_2} U, \quad U \sim \mathcal{CN}(0, 1).$$

The expression (154) simplifies to

$$I_1(A; Y) = \Pr[\mathcal{E}_2] \left[\log(1 + \hat{P}) + \frac{\mathbb{E}[|Y|^2 | \mathcal{E}_2]}{1 + \hat{P}} - \mathbb{E} \left[\left| Y - \sqrt{\hat{P}} U \right|^2 \middle| \mathcal{E}_2 \right] \right]. \quad (295)$$

The GMI (295) exhibits interesting high and low SNR scaling by choosing the following thresholds t, t_R .

- For high SNR, we choose

$$t = P^{-\lambda} \text{ and } t_R = \hat{P}^{\lambda_R} \quad (296)$$

where $0 < \lambda < 1$ and $0 < \lambda_R < 1$. As P increases, t decreases and Appendix B.4 shows that

$$\Pr[\mathcal{E}_2] \rightarrow 1, \quad \frac{\mathbb{E}[|Y|^2 | \mathcal{E}_2]}{1 + \hat{P}} \rightarrow 1, \quad \mathbb{E} \left[\left| Y - \sqrt{\hat{P}} U \right|^2 \middle| \mathcal{E}_2 \right] \rightarrow 1. \quad (297)$$

Inserting $\hat{P} = P/E_1(t)$, we thus have

$$\lim_{P \rightarrow \infty} \left[I_1(A; Y) - \log \left(1 + \frac{P}{E_1(t)} \right) \right] = 0. \quad (298)$$

We further have $E_1(t) \approx \lambda \log P$ by using (A6) in Appendix A.2, and the high-SNR slope of the GMI matches the slope of $\log P$ but the additive gap to $\log P$ increases. The high SNR rates are shown as the curve labeled “TCI, GMI, $K = 2$ ” in Figure 10 for $\lambda = \lambda_R = 0.4$.

- For low SNR, we choose

$$t = -\log(P/c) \text{ and } t_R = \hat{P} \quad (299)$$

for a constant $c > 0$. As P decreases, both t and $\hat{P} = P/E_1(t)$ increase and Appendix B.4 shows that

$$\Pr[\mathcal{E}_2] \approx e^{-t-1}, \quad \frac{\mathbb{E}[|Y|^2 | \mathcal{E}_2]}{1 + 2\hat{P}} \rightarrow 1, \quad \mathbb{E} \left[\left| Y - \sqrt{\hat{P}} U \right|^2 \middle| \mathcal{E}_2 \right] \rightarrow 1. \quad (300)$$

Using (A7), we have $I_1(A; Y) \approx e^{-t-1} \log t$ which vanishes as t grows. But we also have

$$\frac{E_b}{N_0} = \frac{P}{R} \log 2 \approx \frac{c e^{-t} \log 2}{e^{-t-1} \log t} \approx \frac{c e \log 2}{\log(-\log P)} \quad (301)$$

which decreases (very slowly) as P decreases. The minimal E_b/N_0 is therefore $-\infty$. The low SNR rates are shown as the curve labeled “TCI, GMI, $K = 2$ ” in Figure 11 for $c = 1.4$.

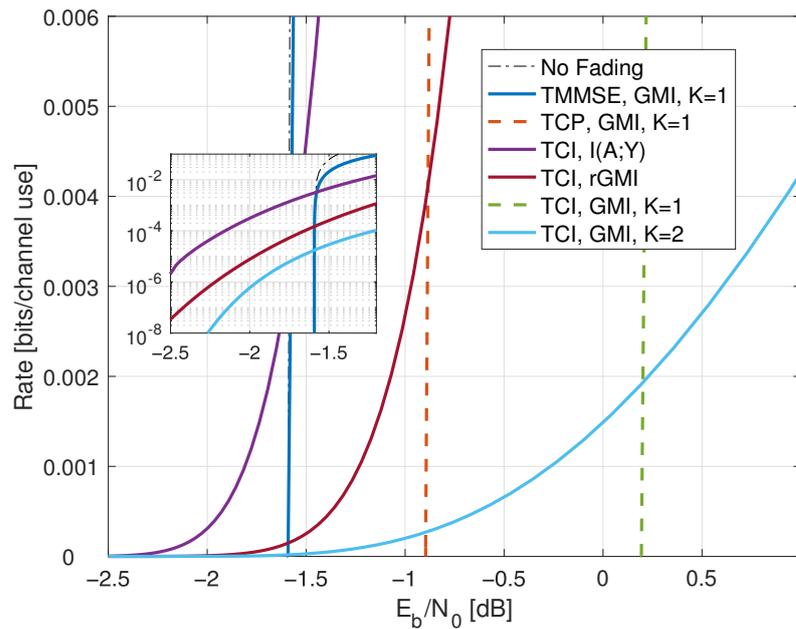


Figure 11. Low-SNR rates for Rayleigh fading with $S_T = H$ and $S_R = 0$. The threshold t was optimized for the $K = 1$ curves, while $t = -\log(P/1.4)$ for the $I(A; Y)$, rGMI, and $K = 2$ curves. The $K = 2$ GMI uses $t_R = \hat{P}$. The TMF and TMMSE GMIs are indistinguishable for this range of rates.

Figure 11 shows that the TCI mutual information achieves a minimal E_b/N_0 below -1.59 dB. At $E_b/N_0 = -2$ dB, we computed $I_1(A; Y) \approx 6 \times 10^{-7}$ and $I(A; Y) \approx 3 \times 10^{-4}$. The $K = 2$ partition is thus useful to prove that TCI can achieve E_b/N_0 arbitrarily close to zero. Figure 11 also shows the reverse model GMI as the curve labeled “TCI, rGMI” which has the rate $I_1(A; Y) \approx 8 \times 10^{-6}$ at $E_b/N_0 = -2$ dB.

We compare the full CSIR and full CSIT rates. At high SNR, the GMI for $S_R = 0$ achieves the same capacity pre-log as $S_R = H$. At low SNR, recall from (271) that with full CSIR/CSIT we have $E_b/N_0 \approx \log(2)/\lambda$. To compare the rates for similar E_b/N_0 , we set $\lambda = \log t$, where t is as in (299) and $c \approx 1$. The TCI $K = 2$ GMI without CSIR is approximately $e^{-t} \log t$ while the full CSIR rate (271) is approximately $e^{-\lambda}/\lambda \approx 1/(t \log(t))$. Thus, the $K = 2$ GMI with no CSIR is a fraction $te^{-t} \log(t)^2$ of the full CSIR capacity.

$S_R = 1(G \geq t)$: The power in (206) is again (288) and the rate (224) is

$$I_1(A; Y|S_R) = e^{-t} \cdot \log \left(1 + \frac{P e^{2t} \Gamma(\frac{3+a}{2}, t)^2}{\Gamma(1+a, t) + P [e^t \Gamma(2+a, t) - e^{2t} \Gamma(\frac{3+a}{2}, t)^2]} \right). \tag{302}$$

Moreover, the expression (225) is

$$\frac{E_b}{N_0} \Big|_{\min} = \frac{\Gamma(1+a, t)}{e^t \cdot \Gamma(\frac{3+a}{2}, t)^2} \cdot \log 2 \tag{303}$$

which is the same as (290) except for the factor e^t in the denominator. This implies that the minimal E_b/N_0 can be improved for $t > 0$.

The TCP, TMF, and TCI rates (302) are the respective

$$I_1(A; Y|S_R) = e^{-t} \log \left(1 + \frac{P e^{2t} \Gamma(\frac{3}{2}, t)^2}{e^{-t} + P [t + 1 - e^{2t} \Gamma(\frac{3}{2}, t)^2]} \right) \tag{304}$$

$$I_1(A; Y|S_R) = e^{-t} \log \left(1 + \frac{P (t + 1)^2}{e^{-t}(t + 1) + P} \right) \tag{305}$$

$$I_1(A; Y|S_R) = e^{-t} \log \left(1 + \frac{P}{E_1(t)} \right). \tag{306}$$

Remark 73. As pointed out in Remark 68, the TCI GMI (306) is $I(A; Y|S_R)$. One can also understand this by observing that the receiver knows $\sqrt{GP(G)}$ for all G . The mutual information is thus related to the rate (189) of Proposition 2.

The minimal E_b/N_0 in (303) are the respective

$$\frac{E_b}{N_0} \Big|_{\min} = \frac{1}{e^{2t} \cdot \Gamma(\frac{3}{2}, t)^2} \cdot \log 2 \tag{307}$$

$$\frac{E_b}{N_0} \Big|_{\min} = \frac{1}{t + 1} \cdot \log 2 \tag{308}$$

$$\frac{E_b}{N_0} \Big|_{\min} = e^t E_1(t) \cdot \log 2. \tag{309}$$

The above expressions mean that, for all three policies, we can make the minimal E_b/N_0 as small as desired by increasing t . For example, for TCI, we can bound (see (A9) below)

$$\frac{1}{t + 1} < e^t E_1(t) < \frac{1}{t}. \tag{310}$$

TCI thus has a slightly larger (slightly worse) minimal E_b/N_0 than TMF for the same t , as discussed after (212).

For large P , the TCP rate (304) is optimized by $t \approx 0.163$ and the rate saturates at ≈ 2.35 bits per channel use. The TMF rate (305) is optimized with $t = 0$, and the rate saturates at 1 bit per channel use. For the TCI rate (306), we again choose $t = 1/P$ and use $E_1(t) \approx \log(1/t)$ for small t to show that the capacity grows with pre-log 1:

$$I_1(A; Y|S_R) \approx \log \left(1 + \frac{P}{\log P} \right).$$

Again, TMF is best at small P while TCI is best at large P .

Remark 74. Comparing (298) and (306), the $S_R = 0$, $K = 2$, TCI GMI in (295) approaches the $S_R = 1(G \geq t)$ mutual information $I(A; Y|S_R)$ in (306) at high SNR.

Optimal Policy: Consider now the optimal power control policy. Suppose first that $S_R = 0$ for which Theorem 2 gives the TMMSE policy with $t = 0$:

$$\sqrt{P(h)} = \frac{\alpha|h|}{\beta + |h|^2}. \tag{311}$$

For Rayleigh fading, we thus have (see (A13) below)

$$P = \int_0^\infty e^{-g} \frac{\alpha^2 g}{(\beta + g)^2} dg = \alpha^2 [(\beta + 1)e^\beta E_1(\beta) - 1] \tag{312}$$

with the two expressions (see (A12) and (A14) below)

$$\tilde{P} = \int_0^\infty e^{-g} \frac{\alpha^2 g}{\beta + g} dg = \alpha^2 \left[1 - \beta e^\beta E_1(\beta) \right]^2 \tag{313}$$

$$E[GP(H)] = \int_0^\infty e^{-g} \frac{\alpha^2 g^2}{(\beta + g)^2} dg = \alpha^2 \left[1 + \beta - \beta(\beta + 2)e^\beta E_1(\beta) \right]. \tag{314}$$

Given P and β , we may compute α^2 from (312). We then search for the optimal β for fixed P . The rates are shown as the curve labeled “TMMSE, GMI, $K = 1$ ” in Figures 10 and 11 and we see that the TMMSE strategy has the best $K = 1$ rates.

Consider next $S_R = 1(G \geq t)$ and the TMMSE policy. We compute (see (A13) below)

$$P = \int_t^\infty e^{-g} \frac{\alpha^2 g}{(\beta + g)^2} dg = \alpha^2 \left[(\beta + 1)e^\beta E_1(t + \beta) - e^{-t} \frac{\beta}{t + \beta} \right] \tag{315}$$

and (see (A12) and (A14) below)

$$\begin{aligned} \sqrt{\tilde{P}(1)} &= \int_t^\infty \frac{e^{-g}}{e^{-t}} \frac{\alpha g}{\beta + g} dg = \alpha \left[1 - \beta e^{t+\beta} E_1(t + \beta) \right] \tag{316} \\ E[|Y|^2 | S_R = 1] &= \int_t^\infty \frac{e^{-g}}{e^{-t}} \left(1 + \frac{\alpha^2 g^2}{(\beta + g)^2} \right) dg \\ &= 1 + \alpha^2 \left[1 + \frac{\beta^2}{t + \beta} - \beta(\beta + 2)e^{t+\beta} E_1(t + \beta) \right]. \tag{317} \end{aligned}$$

We optimize as for the $S_R = 0$ case: given P, β, t , we compute α^2 from (315). We then search for the optimal β for fixed P and t . The optimal t is approximately a factor of 1.1 smaller than for the TCI policy. The rates are shown in Figure 12 as the curve labeled “TMMSE, GMI”.

8.5. Partial CSIR, CSIT@R

Suppose S_R is defined by (see (172))

$$H = \sqrt{\epsilon} S_R + \sqrt{1-\epsilon} Z_R$$

where $0 \leq \epsilon \leq 1$ and S_R, Z_R are independent with distribution $\mathcal{CN}(0, 1)$. We further consider the CSIT $S_T = |S_R|^2$.

The reverse model GMI again requires $\text{Var}[U|Y, S_R]$, which can be computed by simulation; see Appendix C.4. However, as in Sections 7.4 and 8.3, optimizing the powers seems difficult, and we instead focus on forward models. The expressions (235) and (236) are

$$\tilde{g}(s_R) = \bar{\epsilon} s_T, \quad \tilde{\sigma}^2(s_R) = \epsilon. \tag{318}$$

The GMI (237) of Theorem 3 is

$$I_1(X; Y | S_R) = \int_{\lambda/\bar{\epsilon}}^\infty e^{-s_T} \log \left(1 + \frac{\bar{\epsilon} s_T P(s_T)}{1 + \epsilon P(s_T)} \right) ds_T \tag{319}$$

where the power control policy $P(s_T)$ is given by (245). The parameter λ is chosen so that $E[P(s_T)] = P$. For example, for $\epsilon \rightarrow 0$ we recover the waterfilling solution (181). Figure 13 shows the quadratic and conventional waterfilling rates, which lie almost on top of each other. For example, the inset shows the rates for $\epsilon = 0.2$ and a small range of E_b/N_0 .

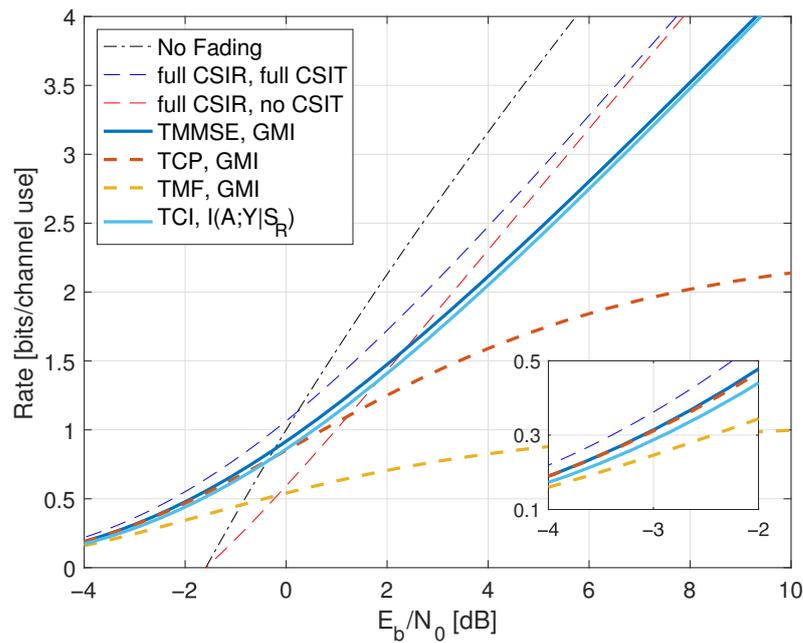


Figure 12. Rates for Rayleigh fading with full CSIT and $S_R = 1 (G \geq t)$.

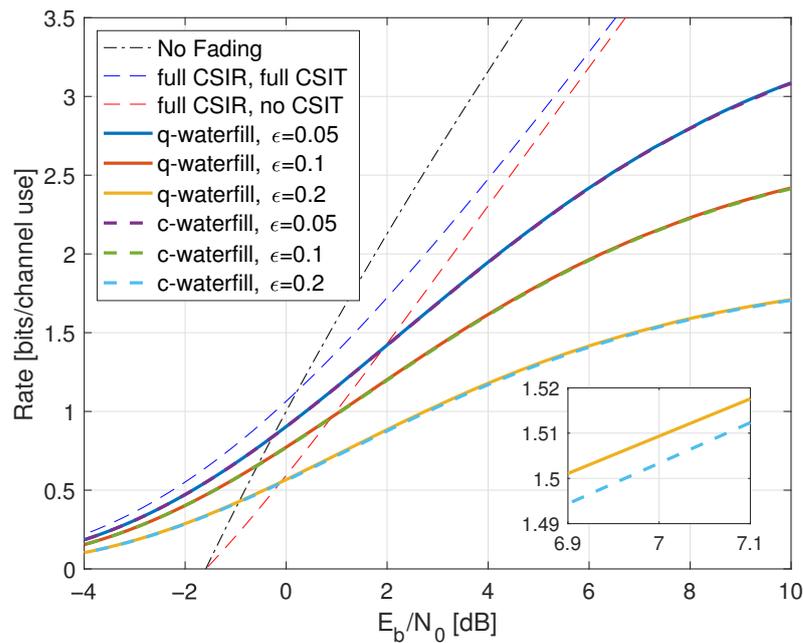


Figure 13. Rates for Rayleigh fading with partial CSIR and CSIT@R. The curves labeled ‘q-waterfill’ and ‘c-waterfill’ are the quadratic and conventional waterfilling rates, respectively.

9. Channels with In-Block Feedback

This section generalizes Shannon’s model described in Section 4.1 to include block fading with in-block feedback. For example, the model lets one include delay in the CSIT and permits many other generalizations for network models [22].

9.1. Model and Capacity

The problem is specified by the FDG in Figure 14. The model has a message M , and the channel input and output strings

$$X_i^L = (X_{i1}, \dots, X_{iL}), \quad Y_i^L = (Y_{i1}, \dots, Y_{iL})$$

for blocks $i = 1, \dots, n$. The channel is specified by a string $S_H^n = (S_{H1}, \dots, S_{Hn})$ of i.i.d. hidden channel states. The CSIR $S_{Ri\ell}$ is a (possibly noisy) function of S_{Hi} for all i and ℓ . The receiver sees the channel outputs (see (159))

$$(Y_{i\ell}, S_{Ri\ell}) = \left(f_\ell(X_i^\ell, S_{Hi}, Z_i^L), S_{Ri\ell} \right) \tag{320}$$

for some functions $f_\ell(\cdot)$, $\ell = 1, \dots, L$. Observe that the X_i^ℓ influence the $Y_{i\ell}$ in a causal fashion. The random variables $M, S_{H1}, \dots, S_{Hn}, Z_1^L, \dots, Z_n^L$ are mutually independent.

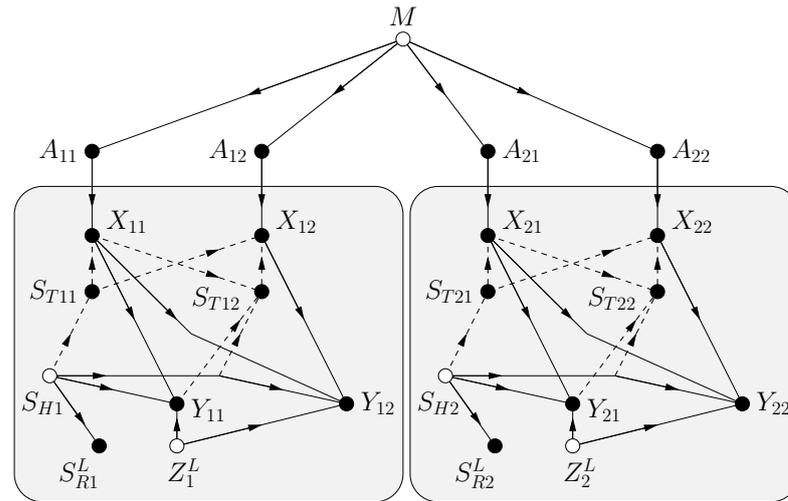


Figure 14. FDG for a block fading model with $n = 2$ blocks of length $L = 2$ and in-block feedback. Cross-block dependence via past $S_{Ti\ell}$ is not shown.

We now permit past channel symbols to influence the CSIT; see Section 1.2. Suppose the CSIT has the form

$$S_{Ti\ell} = f_{T\ell}(S_{Hi}, X_i^{\ell-1}, Y_i^{\ell-1}) \tag{321}$$

for some function $f_{T\ell}(\cdot)$ and for all i and ℓ . The motivation for (321) is that useful CSIR may not be available until the end of a block or even much later. In the meantime, the receiver can, e.g., quantize the $Y_i^{\ell-1}$ and transmit the quantization bits via feedback. This lets one study fast power control and beamforming without precise knowledge of the channel coefficients.

Define the string of past and current states as

$$s_T^{i\ell} = \left(s_{T1}^L, \dots, s_{T(i-1)}^L, s_{Ti}^L \right). \tag{322}$$

The channel input at time $i\ell$ is $X(s_T^{i\ell})$ and the adaptive codeword A^{nL} is defined by the ordered lists

$$A_{i\ell} = \left[X(s_T^{i\ell}), \forall s_T^{i\ell} \right] \tag{323}$$

for $1 \leq i \leq n$ and $1 \leq \ell \leq L$. The adaptive codeword A^{nL} is a function of M and is thus independent of S_H^n and S_R^{nL} .

The model under consideration is a special case of the channels introduced in ([22], Section V). However, the model in [22] has transmission and reception begin at time $\ell = 2$ rather than $\ell = 1$. To compare the theory, one must thus shift the time indexes by 1 unit and increase L to $L + 1$. The capacity for our model is given by ([22], Theorem 2) which we write as

$$C \stackrel{(a)}{=} \max_{A^L} \frac{1}{L} I(A^L; Y^L, S_R^L) \stackrel{(b)}{=} \max_{A^L} \frac{1}{L} I(A^L; Y^L | S_R^L). \tag{324}$$

where (a) follows by normalizing by L rather than $L + 1$, and step (b) follows by the independence of A^L and S_R^L .

9.2. GMI for Scalar Channels

We will study scalar block fading channels; extensions to vector channels follow as described in Section 4.4. Let $\underline{Y} = [Y_1, \dots, Y_L]^T$ be the vector form of Y^L and similarly for other strings with L symbols. The GMI with parameter s is

$$I_s(A^L; Y^L | S_R^L) = \mathbb{E} \left[\log \frac{q(\underline{Y} | \underline{A}, \underline{S}_R)^s}{q(\underline{Y} | \underline{S}_R)} \right] \tag{325}$$

Reverse Model: For the reverse model, let \underline{A} be a column vector that stacks the $X_\ell(s_T^\ell)$ for all s_T^ℓ and ℓ . Consider a reverse density as in (105):

$$q(a^L | y^L) = \frac{\exp \left(-\underline{z}(\underline{y}, \underline{s}_R)^\dagger \mathbf{Q}_{\underline{A} | \underline{Y}=\underline{y}, \underline{S}_R=\underline{s}_R}^{-1} \underline{z}(\underline{y}, \underline{s}_R) \right)}{\pi^N \det \mathbf{Q}_{\underline{A} | \underline{Y}=\underline{y}, \underline{S}_R=\underline{s}_R}}$$

where

$$\underline{z}(\underline{y}, \underline{s}_R) = \underline{a} - \mathbb{E} \left[\underline{A} | \underline{Y} = \underline{y}, \underline{S}_R = \underline{s}_R \right].$$

Using the forward model $q(y^L | a^L) = q(a^L | y^L) / p(a^L)$, the GMI with $s = 1$ becomes

$$I_1(A^L; Y^L, S_R^L) = \mathbb{E} \left[\log \frac{\det \mathbf{Q}_{\underline{A}}}{\det \mathbf{Q}_{\underline{A} | \underline{Y}, \underline{S}_R}} \right]. \tag{326}$$

To simplify, consider adaptive symbols as in (89) (cf. (107)):

$$X_\ell(s_T^\ell) = \sqrt{P_\ell(s_T^\ell)} e^{j\phi_\ell(s_T^\ell)} U_\ell \tag{327}$$

where $\underline{U} \sim \mathcal{CN}(0, \mathbf{I})$. In other words, consider a conventional codebook represented by the U_ℓ and adapt the power and phase based on the available CSIT. The mutual information becomes $I(A^L; Y^L, S_R^L) = I(U^L; Y^L, S_R^L)$ (cf. (96)) and the GMI with $s = 1$ is (cf. (108))

$$I_1(A^L; Y^L | S_R^L) = \mathbb{E} \left[-\log \det \mathbf{Q}_{\underline{U} | \underline{Y}, \underline{S}_R} \right]. \tag{328}$$

In fact, one may also consider choosing $U_\ell = U$ for all ℓ in which case we compute (cf. (139))

$$I_1(A^L; Y^L | S_R^L) = \mathbb{E} \left[-\log \text{Var} [U | \underline{Y}, \underline{S}_R] \right]. \tag{329}$$

Forward Model: Consider the following forward model (cf. (111) and (141)):

$$q(\underline{y} | \underline{a}, \underline{s}_R) = \frac{\exp \left(-\underline{z}(\underline{s}_R)^\dagger \mathbf{Q}_{\underline{Z}(\underline{s}_R)}^{-1} \underline{z}(\underline{s}_R) \right)}{\pi^L \det \mathbf{Q}_{\underline{Z}(\underline{s}_R)}}. \tag{330}$$

with

$$\underline{z}(\underline{s}_R) = \underline{y} - \mathbf{H}(\underline{s}_R) \underline{\tilde{x}}(\underline{s}_R)$$

and where similar to (142) we define

$$\underline{X}(\underline{s}_R) = \sum_{\underline{s}_T} \mathbf{W}(\underline{s}_T, \underline{s}_R) \underline{X}(\underline{s}_T) \tag{331}$$

where the $\mathbf{W}(\underline{s}_T, \underline{s}_R)$ are $L \times L$ complex matrices. Note that

$$\underline{X}(\underline{s}_T) = [X_1(s_{T1}), X_2(s_{T2}^2), \dots, X_L(s_{TL}^L)]^T \tag{332}$$

so X_ℓ is a function of A^L and $S_T^\ell, \ell = 1, \dots, L$.

We have the following generalization of Lemma 4 (see also Theorem 1) where the novelty is that S_T is replaced with \underline{S}_T . Define $\underline{U}(\underline{s}_T) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ and $\underline{X}(\underline{s}_T) = \mathbf{Q}_{\underline{X}(\underline{s}_T)}^{1/2} \underline{U}(\underline{s}_T)$ for all \underline{s}_T .

Theorem 4. A GMI (325) for the scalar block fading channel $p(y^L|a^L, s_R^L)$, an adaptive codeword A^L with jointly CSCG entries, the auxiliary model (330), and with fixed $\mathbf{Q}_{X(\underline{s}_T)}$ is

$$I_1(A^L; Y^L | S_R^L) = \mathbb{E} \left[\log \left(\frac{\det \mathbf{Q}_Y(\underline{s}_R)}{\det(\mathbf{Q}_Y(\underline{s}_R) - \tilde{\mathbf{D}}(\underline{s}_R) \tilde{\mathbf{D}}(\underline{s}_R)^\dagger)} \right) \right]. \tag{333}$$

where

$$\mathbf{Q}_Y(\underline{s}_R) = \mathbb{E} \left[\underline{Y} \underline{Y}^\dagger \mid \underline{s}_R = \underline{s}_R \right] \tag{334}$$

and for $M \times M$ unitary $\mathbf{V}_R(\underline{s}_T, \underline{s}_R)$ the matrix $\tilde{\mathbf{D}}(\underline{s}_R)$ is

$$\mathbb{E} \left[\mathbf{U}_T(\underline{s}_T, \underline{s}_R) \mathbf{\Sigma}(\underline{s}_T, \underline{s}_R) \mathbf{V}_R(\underline{s}_T, \underline{s}_R)^\dagger \mid \underline{s}_R = \underline{s}_R \right] \tag{335}$$

and $\mathbf{U}_T(\underline{s}_T, \underline{s}_R)$ and $\mathbf{\Sigma}(\underline{s}_T, \underline{s}_R)$ are $N \times N$ unitary and $N \times M$ rectangular diagonal matrices, respectively, of the SVD

$$\mathbb{E} \left[\underline{Y} \underline{U}(\underline{s}_T)^\dagger \mid \underline{s}_T = \underline{s}_T, \underline{s}_R = \underline{s}_R \right] = \mathbf{U}_T(\underline{s}_T, \underline{s}_R) \mathbf{\Sigma}(\underline{s}_T, \underline{s}_R) \mathbf{V}_T(\underline{s}_T, \underline{s}_R)^\dagger \tag{336}$$

for all $\underline{s}_T, \underline{s}_R$ and the $\mathbf{V}_T(\underline{s}_T, \underline{s}_R)$ are $M \times M$ unitary matrices. One may maximize (333) over the unitary $\mathbf{V}_R(\underline{s}_T, \underline{s}_R)$.

Suppose next that the actual channel is $\underline{Y} = H \underline{X} + \underline{Z}$ where $\underline{Z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. The extension of (136) and (168) to block fading channels with CSIR is

$$I_1(A^L; Y^L | S_R^L) = \sum_{\ell=1}^L \mathbb{E} \left[\log \left(1 + \frac{\tilde{P}_\ell(\underline{s}_R)}{1 + \mathbb{E}[GP_\ell(S_T^\ell) | \underline{s}_R] - \tilde{P}_\ell(\underline{s}_R)} \right) \right] \tag{337}$$

where (cf. (166) and (167))

$$\begin{aligned} \tilde{P}_\ell(\underline{s}_R) &= \mathbb{E} \left[\left| \mathbb{E} \left[H \sqrt{P_\ell(S_T^\ell)} \mid \underline{s}_T, \underline{s}_R = \underline{s}_R \right] \right|^2 \right] \\ \mathbb{E} \left[|Y_\ell|^2 \mid \underline{s}_R = \underline{s}_R \right] &= 1 + \mathbb{E} \left[GP_\ell(S_T^\ell) \mid \underline{s}_R = \underline{s}_R \right]. \end{aligned}$$

9.3. CSIT@ R

Continuing as in Section 5.2, suppose the CSIT in (321) can be written by replacing S_{Hi} with S_{Ri}^ℓ for all i and ℓ :

$$S_{Til} = f_{Til}(S_{Ri}^\ell, X_i^{\ell-1}, Y_i^{\ell-1}). \tag{338}$$

The capacity (324) then simplifies to a directed information. To see this, expand the mutual information in (324) as

$$\begin{aligned}
 I(A^L; Y^L | S_R^L) &\stackrel{(a)}{=} \sum_{\ell=1}^L I(A^L, X^\ell; Y_\ell | S_R^L, Y^{\ell-1}) \\
 &\stackrel{(b)}{=} \sum_{\ell=1}^L I(X^\ell; Y_\ell | S_R^L, Y^{\ell-1})
 \end{aligned}
 \tag{339}$$

where step (a) follows because X^ℓ is a function of A^L and S_T^ℓ in (338), and step (b) follows by the Markov chains

$$A^L - [S_R^L, X^\ell, Y^{\ell-1}] - Y_\ell.
 \tag{340}$$

The capacity is therefore (see the definition (27))

$$C = \max_{X_\ell(S_T^\ell), \ell=1, \dots, L} \frac{1}{L} I(X^L \rightarrow Y^L | S_R^L).
 \tag{341}$$

The maximization in (341) under a cost constraint becomes a constrained maximization for which $E[c(X^L, Y^L)] \leq LP$ for some cost function $c(\cdot)$.

Remark 75. As outlined at the end of Section 9.1, the capacity (341) is a special case of the theory in ([22], Equation (48)). To see this, define the extended and time-shifted strings

$$\hat{A}^{L+1} = (0, A^L), \quad \hat{X}^{L+1} = (0, X^L), \quad \hat{Y}^{L+1} = (0, Y^L).$$

Since A^L and S_R^L are independent, one may expand (339) as

$$\begin{aligned}
 I(A^L; Y^L | S_R^L) &= I(A^L; (S_{R2}, \dots, S_{RL}, 0), Y^L | S_{R1}) \\
 &\stackrel{(a)}{=} \sum_{\ell=1}^L I(A^L, X^\ell; S_{R(\ell+1)}, Y_\ell | S_R^\ell, Y^{\ell-1}) \\
 &\stackrel{(b)}{=} \sum_{\ell=1}^L I(X^\ell; S_{R(\ell+1)}, Y_\ell | S_R^\ell, Y^{\ell-1}) \\
 &= \sum_{\ell=2}^{L+1} I(\hat{X}^\ell; S_{R\ell}, \hat{Y}_\ell | S_R^{\ell-1}, \hat{Y}^{\ell-1})
 \end{aligned}
 \tag{342}$$

where step (a) follows because X^ℓ is a function of A^L and S_T^ℓ in (338), and where $S_{R(L+1)} = 0$, and step (b) follows by the Markov chains

$$A^L - [X^\ell, Y^{\ell-1}, S_R^\ell] - [Y_\ell, S_{R(\ell+1)}].
 \tag{343}$$

The expression (342) is the desired directed information

$$I(A^L; Y^L, S_R^L) = I(\hat{X}^{L+1} \rightarrow \hat{Y}^{L+1}, S_R^{L+1}).
 \tag{344}$$

Remark 76. Consider the basic CSIT model

$$S_{Til} = f_T(S_{Ril})
 \tag{345}$$

for some function $f_T(\cdot)$ and for $\ell = 1, \dots, L$ and $i = 1, \dots, n$. This model was studied in ([103], Section III.C) and its capacity is given as (see ([103], Equation (35) with Equation (13)))

$$C = \max_{X_\ell(S_T^\ell), \ell=1, \dots, L} \frac{1}{L} I(X^L; Y^L | S_R^L, S_T^L).
 \tag{346}$$

To see that (346) is a special case of (341), observe that

$$\begin{aligned}
 I(X^L \rightarrow Y^L | S_R^L) &\stackrel{(a)}{=} \sum_{\ell=1}^L I(X^\ell; Y_\ell | S_R^L, S_T^L, Y^{\ell-1}) \\
 &\stackrel{(b)}{=} \sum_{\ell=1}^L I(X^L; Y_\ell | S_R^L, S_T^L, Y^{\ell-1})
 \end{aligned}
 \tag{347}$$

where step (a) follows by (339), and step (b) follows by the Markov chains

$$[X_{\ell+1}, \dots, X_L] - [S_R^L, S_T^L, Y^{\ell-1}, X^\ell] - Y_\ell.
 \tag{348}$$

The expression (347) gives (346). Related results are available in ([10], Section III) and [104,105].

Remark 77. The capacity (341) has only S_R^L in the conditioning while (346) has both S_R^L and S_T^L in the conditioning. This subtle difference is due to permitting $X^{\ell-1}$ to influence the $S_{T\ell}$ in (338), and it complicates the analysis. On the other hand, if we remove only $X^{\ell-1}$ from (338) then the receiver knows $S_{T\ell}$ at time ℓ and the capacity (341) can be written as (see the definition (28))

$$C = \max_{X_\ell(S_T^\ell), \ell=1, \dots, L} \frac{1}{L} I(X^L \rightarrow Y^L | S_T^L | S_R^L).
 \tag{349}$$

We treat such a model in Section 9.7 below.

9.4. Fading Channels with AWGN

The expression (341) is valid for general statistics. We next specialize to the block-fading AWGN model

$$Y_\ell = H X_\ell + Z_\ell
 \tag{350}$$

where $\ell = 1, \dots, L$, $Z^L \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, and (H, S_R^L) , A^L, Z^L are mutually independent. Consider the power constraint

$$\sum_{\ell=1}^L \mathbb{E} [P_\ell(S_T^\ell)] \leq LP
 \tag{351}$$

where $P_\ell(s_T^\ell) = \mathbb{E} [|X_\ell(s_T^\ell)|^2]$. The optimization of (341) under the constraint (351) is usually intractable, and we again desire expressions with $\log(1 + \text{SNR})$ terms to obtain insight.

Capacity Upper Bound: Using similar steps as in (162), we have

$$\begin{aligned}
 I(A^L; Y^L | S_R^L) &\leq I(A^L; Y^L, H | S_R^L) \\
 &= \sum_{\ell=1}^L I(A^L; Y_\ell | S_R^L, H, Y^{\ell-1}) \\
 &\leq \sum_{\ell=1}^L [h(Y_\ell | S_R^L, H, Y^{\ell-1}) - h(Z_\ell)] \\
 &\stackrel{(a)}{\leq} \sum_{\ell=1}^L \mathbb{E} \left[\log \left(1 + \mathbb{E} [G P_\ell(S_T^\ell) | S_R^L, H, Y^{\ell-1}] \right) \right]
 \end{aligned}
 \tag{352}$$

where $G = |H|^2$ and step (a) follows by (163). However, CSCG inputs do not necessarily maximize the RHS of (352) because the inputs affect the CSIT.

Remark 78. The expectation inside the logarithm in (352) becomes $G P_\ell(S_T^\ell)$ if S_T^ℓ is a function of $S_R^L, H, Y^{\ell-1}$; see (161), Remark 77, and Proposition 3 below.

Achievable Rates: Deriving achievable rates is more subtle than in Section 6. Consider the CSIT model (338) where for each block, we have

$$S_{T\ell} = f_{T\ell}(H, X^{\ell-1}, Y^{\ell-1})$$

for all ℓ . The capacity (341) is

$$C(P) = \max_{X_\ell(S_T^\ell), \ell=1, \dots, L} \frac{1}{L} I(X^L \rightarrow Y^L | H) \tag{353}$$

$$= \max_{X_\ell(S_T^\ell), \ell=1, \dots, L} \left[\frac{1}{L} h(Y^L | H) \right] - \log(\pi e). \tag{354}$$

However, CSCG inputs are not necessarily optimal since the inputs affect the CSIT.

Instead of trying to optimize the input, consider X_ℓ that are CSCG. We may write

$$I(X^L \rightarrow Y^L | H) = \sum_{\ell=1}^L \mathbb{E} \left[\log \left(1 + GP_\ell(S_T^\ell) \right) \right] \tag{355}$$

and the Lagrangians to maximize (355) are

$$\sum_{\ell=1}^L \mathbb{E} \left[\log \left(1 + GP_\ell(S_T^\ell) \right) \right] + \lambda \left(LP - \sum_{\ell=1}^L \mathbb{E} \left[P_\ell(S_T^\ell) \right] \right). \tag{356}$$

Suppose the $S_{T\ell}$ are discrete random variables. Taking the derivative with respect to $P_\ell(s_T^\ell)$, we obtain

$$\begin{aligned} \lambda &= \int_0^\infty p(g|s_T^\ell) \frac{g}{1 + gP_\ell(s_T^\ell)} dg \\ &+ \sum_{k=\ell+1}^L \sum_{s_T^k} \int_0^\infty p(g) \frac{dP_{S_T^k|G}(s_T^k|g) \log \left(1 + gP_k(s_T^k) \right)}{dP_\ell(s_T^\ell) P_{S_T^k}(s_T^k)} dg \end{aligned} \tag{357}$$

as long as $P_\ell(s_T^\ell) > 0$. This expression is complicated because the choice of transmit powers $P_\ell(s_T^\ell)$ influences the statistics of the future CSIT $S_{T(\ell+1)}, \dots, S_{TL}$. If (357) cannot be satisfied, choose $P_\ell(s_T^\ell) = 0$. Finally, set λ so that $\sum_{\ell=1}^L \mathbb{E} \left[P_\ell(S_T^\ell) \right] = LP$.

Instead of the above, consider the simpler CSIT model with $S_{T\ell} = f_{T\ell}(H)$ for all ℓ , cf. (345). The capacity (346) is now given by (355) with CSCG inputs and (357) simplifies because the derivatives with respect to $P_\ell(s_T^\ell)$ are zero, i.e., the double sum in (357) disappears and for all ℓ and s_T^ℓ we have

$$\lambda = \int_0^\infty p(g|s_T^\ell) \frac{g}{1 + gP_\ell(s_T^\ell)} dg. \tag{358}$$

We use (358) for (362)–(364) in Section 9.7 below.

9.5. Full CSIR, Partial CSIT

We next generalize Proposition 2 in Section 6.4 to the block-fading AWGN model (350) with the CSIR

$$S_{R\ell} = H \sqrt{P(S_T^\ell)}, \quad \ell = 1, \dots, L \tag{359}$$

and where $S_{T\ell} = f_{T\ell}(S_H)$, i.e., we have discarded $X_i^{\ell-1}$ and $Y_i^{\ell-1}$ in (321). We then have the following capacity result that implies this CSIR is the best possible since one achieves a capacity upper bound similar to (161).

Proposition 3. The capacity of the channel (350) with the CSIR (359) and $S_{T\ell} = f_{T\ell}(S_H)$ for $\ell = 1, \dots, L$ is

$$C(P) = \max \frac{1}{L} \sum_{\ell=1}^L \mathbb{E} \left[\log \left(1 + GP_{\ell}(S_T^{\ell}) \right) \right] \tag{360}$$

where the maximization is over the power control policies $P_{\ell}(S_T^{\ell})$ such that $\sum_{\ell=1}^L \mathbb{E} \left[P_{\ell}(S_T^{\ell}) \right] \leq LP$. One may use (358) to compute the $P_{\ell}(S_T^{\ell})$.

Proof. For achievability, apply (337) with

$$\tilde{P}_{\ell}(\underline{S}_R) = GP_{\ell}(S_T^{\ell}) \text{ and } \mathbb{E} \left[|Y_{\ell}|^2 | \underline{S}_R \right] = 1 + \tilde{P}_{\ell}(\underline{S}_R).$$

The converse follows by applying similar steps as in (162):

$$\begin{aligned} I(A^L; Y^L | S_R^L) &\leq I(A^L; Y^L, S_T^L, H | S_R^L) \\ &= \sum_{\ell=1}^L I \left(A^{\ell}; Y_{\ell} | S_R^{\ell}, S_T^{\ell}, H, Y^{\ell-1} \right) \\ &\leq \sum_{\ell=1}^L \left[h(Y_{\ell} | S_R^{\ell}, S_T^{\ell}, H, Y^{\ell-1}) - h(Z_{\ell}) \right] \\ &\stackrel{(a)}{\leq} \sum_{\ell=1}^L \mathbb{E} \left[\log \text{Var} \left[Y_{\ell} | S_R^{\ell}, S_T^{\ell}, H, Y^{\ell-1} \right] \right]. \end{aligned} \tag{361}$$

Finally, insert $\text{Var} \left[Y_{\ell} | S_R^{\ell}, S_T^{\ell}, H, Y^{\ell-1} \right] = 1 + GP_{\ell}(S_T^{\ell})$. \square

The RHS of (361) is at most the RHS of (352) and hence (361) gives a better bound. However, the bound (361) is valid only for particular CSIT, as in Remark 78.

9.6. On-Off Fading with Delayed CSIT

Consider on-off fading where the CSIT is delayed by D symbols, i.e., we have $S_{T\ell} = 0$ for $\ell = 1, \dots, D$ and $S_{T(D+1)} = H$. Define the transmit powers as $P_{\ell}(s_T^{\ell}) = \mathbb{E} \left[|X(s_T^{\ell})|^2 \right]$ for $\ell = 1, \dots, L$. The capacity is

$$C(P) = \frac{D}{2L} \log(1 + 2P_1) + \frac{L-D}{2L} \log(1 + 2P_{D+1})$$

where we write $P_{D+1} = P_{D+1}(s_T^{D+1})$. Optimizing the powers, we obtain

$$\begin{cases} \left\{ \begin{array}{l} P_1 = P - \frac{L-D}{4L} \\ P_{D+1} = 2P + \frac{D}{2L} \end{array} \right\} & \text{if } P \geq \frac{L-D}{4L} \\ \left\{ \begin{array}{l} P_1 = 0 \\ P_{D+1} = \frac{2LP}{L-D} \end{array} \right\} & \text{else.} \end{cases}$$

For large P , we thus have $C(P) \approx \frac{1}{2} \log(P)$ for all $0 \leq D \leq L$. For small P , we have

$$\begin{aligned} C(P) &= \begin{cases} \frac{L-D}{2L} \log \left(1 + \frac{4LP}{L-D} \right), & \text{if } 0 \leq D < L \\ \log(1 + 2P)/2, & \text{if } D = L \end{cases} \\ &\approx \begin{cases} \left(2P - \frac{4L}{L-D} P^2 \right) \log(e), & \text{if } 0 \leq D < L \\ (P - P^2) \log(e), & \text{if } D = L. \end{cases} \end{aligned}$$

The CSIT thus gives a 3 dB power gain at low SNR since $C(P) \approx 2P \log(e)$ for $0 \leq D < L$ and $C(P) \approx P \log(e)$ for $D = L$. Furthermore, using (37), the slope of the capacity versus E_b/N_0 in bits/s/Hz/(3 dB) is

$$\begin{aligned} &1 - D/L && \text{if } 0 \leq D < L \\ &1 && \text{if } D = L. \end{aligned}$$

In other words, the delay reduces the low-SNR rate by a factor of $1 - D/L$ for $0 \leq D < L$.

9.7. Rayleigh Fading and One-Bit Feedback

Let $q_u(\cdot)$ be the one-bit ($B = 1$) quantizer in Section 2.9. We study Rayleigh fading for two scenarios with $S_R^L = H$, i.e., the receiver knows H after the L transmissions of each block.

- For the CSIT (345), we study delayed feedback where $S_{T\ell} = 0$ for $\ell = 1, \dots, L - 1$ and $S_{TL} = q_u(G)$. The delay is thus $D = L - 1$ in the sense of Section 9.6.
- For the CSIT (338), we study the case $S_{T1} = 0$, $S_{T2} = q_u(|Y_1|)$, and $S_{T\ell} = 0$ for $\ell = 3, \dots, L$. The delay is thus $D = 1$ in the sense of Section 9.6.

Delayed Quantized CSIR Feedback: Consider $S_{T\ell} = 0$ for $\ell = 1, \dots, L - 1$ and $S_{TL} = q_u(G)$. CSCG inputs are optimal, and (347) has the same form as (360). The Lagrangians are given by (356), and we again obtain (358). For the case at hand, we have $L + 1$ equations for λ , namely

$$\lambda = \int_0^\infty e^{-g} \frac{g}{1 + gP_\ell} dg, \quad \ell = 1, \dots, L - 1 \tag{362}$$

$$\lambda = \int_0^\Delta \frac{e^{-g}}{1 - e^{-\Delta}} \frac{g}{1 + gP_L(\Delta/2)} dg \tag{363}$$

$$\lambda = \int_\Delta^\infty \frac{e^{-g}}{e^{-\Delta}} \frac{g}{1 + gP_L(3\Delta/2)} dg \tag{364}$$

where we used (40) and (41) and abused notation by writing $P_L(s_{TL})$ for $P_L(s_T^L)$. We thus have $P_1 = \dots = P_{L-1}$ and obtain three equations. We now search for λ such that

$$(L - 1)P_1 + \sum_s P_{S_{TL}}(s)P_L(s) = LP$$

and the capacity (353) is

$$C(P) = \frac{L - 1}{L} e^{1/P_1} E_1(1/P_1) + \frac{1}{L} \sum_s \int_{\mathcal{I}(s)} e^{-g} \log(1 + gP_L(s)) dg \tag{365}$$

where the sums are over $s = \Delta/2, 3\Delta/2$ and

$$\mathcal{I}(\Delta/2) = [0, \Delta), \quad \mathcal{I}(3\Delta/2) = [\Delta, \infty).$$

We remark that, if $P_1 = 0$, then we set $e^{1/P_1} E_1(1/P_1) = 0$ since $\lim_{x \rightarrow \infty} e^x E_1(x) = 0$.

Figure 15 shows these capacities for $L = 1, 2, 3$ and $\Delta = 1$. At low SNR (e.g., for $L = 3$ below -2.97 dB) we have $P_1 = 0$ and $P_L(\Delta/2) = 0$, i.e., the transmitter is silent unless $S_{TL} = 3\Delta/2$ and it uses power at time $\ell = L$ only. Observe that, as in Section 9.6, a delay of L steps reduces the low-SNR slope, and therefore the low-SNR rates, by a factor of L . Delay can thus be costly at low SNR.

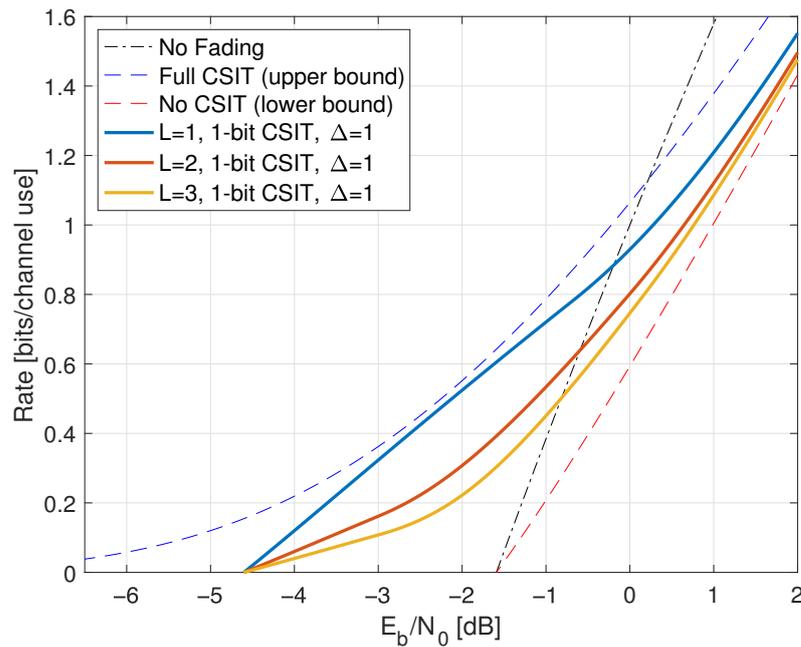


Figure 15. Capacities for Rayleigh block fading with $L = 1, 2, 3$ and a CSIT delay of $D = L - 1$. The CSIT at symbol L is $S_{TL} = q_u(G)$.

Quantized Channel Output Feedback: Consider $S_{T1} = 0$, $S_{T2} = q_u(|Y_1|)$, and $S_{T\ell} = 0$ for $\ell = 3, \dots, L$. As discussed in Remark 77, the capacity is given by the directed information expression (349). However, optimizing the input statistics seems difficult, i.e., CSCG inputs are not necessarily optimal. Instead, we compute achievable rates for a strategy where one symbol partially acts as a pilot.

Suppose the transmitter sends $X_1 = \sqrt{P_1}e^{j\Phi}$ as the first symbol of each block, where Φ is uniformly distributed in $[0, 2\pi)$. The idea is that $|X_1| = \sqrt{P_1}$ is known at the receiver, and thus X_1 acts as a pilot to test the channel amplitude. Next, we choose a variation of flash signaling. Define the event $\mathcal{E} = \{|Y_1| \geq \Delta\} = \{S_{T2} = 3\Delta/2\}$. If this event does not occur, the transmitter sends $X_\ell = 0$ for $\ell = 2, \dots, L$. Otherwise, the transmitter sends independent CSCG X_ℓ with variance $P_2/\Pr[\mathcal{E}]$ for $\ell = 2, \dots, L$. Define $P_\ell(s_T^\ell) = \mathbb{E}[|X(s_T^\ell)|^2]$. We have $P_\ell = P_2$ for $\ell \geq 2$ and the power constraint is $P_1 + (L - 1)P_2 \leq LP$.

We use (347) to write

$$C(P) \geq \frac{1}{L}I(X_1; Y_1|H) + \frac{L-1}{L}I(X^2; Y_2|H, Y_1). \tag{366}$$

The first mutual information in (366) is

$$I(X_1; Y_1|H) = h(Y_1|H) - \log(\pi e)$$

and we compute (see ([52], Appendix A))

$$p(y_1|h) = \frac{1}{\pi}e^{-(|y_1|^2 + P_1|h^2)}I_0\left(2|y_1||h|\sqrt{P_1}\right)$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind of order zero. The Jacobian of the mapping from Cartesian coordinates $[\Re(y_1), \Im(y_1)]$ to polar coordinates $[|y_1|, \arg y_1]$ is $|y_1|$, so we have

$$h(Y_1|H = h) = \int_0^\infty -p(y_1|h) \log(p(y_1|h)) 2\pi|y_1| d|y_1|.$$

We further compute

$$I(X^2; Y_2 | H, Y_1) = \int_0^\infty e^{-g} \Pr[\mathcal{E} | G = g] \log\left(1 + \frac{gP_2}{\Pr[\mathcal{E}]}\right) dg. \tag{367}$$

The conditional probability of a high-energy Y_1 is

$$\Pr[\mathcal{E} | G = g] = Q_1\left(\sqrt{2gP_1}, \sqrt{2}\Delta\right)$$

where $Q_1(\cdot)$ is the Marcum Q-function of order 1; see (A3) in Appendix A.1. For Rayleigh fading, we compute

$$\Pr[\mathcal{E}] = \Pr\left[\left|H\sqrt{P_1}e^{j\Phi} + Z_1\right|^2 \geq \Delta^2\right] = e^{-\Delta^2/(P_1+1)}.$$

The resulting rates are shown in Figure 16 for the block lengths $L = 10, 20, 100$. Observe that each curve turns back on itself, which reflects the non-concavity of the directed information rates in P ; see ([74], Section III). All rates below the curves are achievable by “time-wasting”, i.e., by transmitting for some fraction of the time only. This suggests that flash signaling [73] will improve the rates since one sends information by choosing whether to transmit energy.

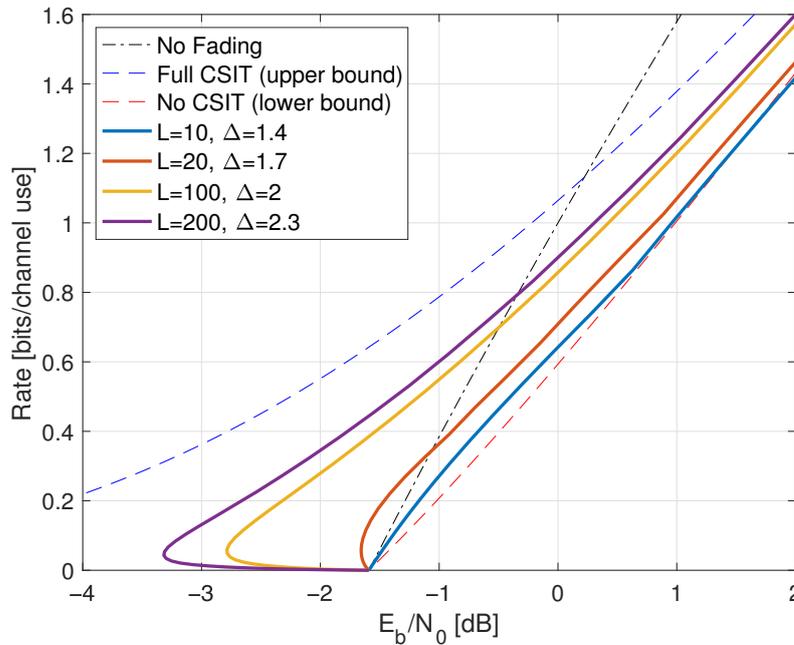


Figure 16. Rates for Rayleigh block fading with block lengths $L = 10, 20, 100$. The CSIT at symbol 2 is $S_{T2} = q_u(|Y_1|)$.

10. Conclusions

This paper reviewed and derived achievable rates for channels with CSIR, CSIT, block fading, and in-block feedback. GMI expressions were developed for adaptive codewords and two classes of auxiliary channel models with AWGN and CSCG inputs: reverse and forward channel models. The forward model inputs were chosen as linear functions of the adaptive codeword’s symbols. We showed that, for scalar channels, an input distribution that maximizes the GMI generates a conventional codebook, where the codeword symbols are multiplied by a complex number that depends on the CSIT. The GMI increases by partitioning the channel output alphabet and modifying the auxiliary model parameters for each partition subset. The partitioning helps to determine the capacity scaling at high and low SNR. Power control policies were developed for full CSIT, including TMMSE policies. The theory was applied to channels with on-off fading and Rayleigh fading. The capacities

with in-block feedback simplify to directed information expressions if the CSIT is a function of the CSIR and past channel inputs and outputs.

There are many possible applications and extensions of this work. For example, adaptive coding and modulation are important for all practical communication systems, including wireless, copper, and fiber-optic networks. Shannon's adaptive codewords can improve current systems since the CSIT is usually a noisy version of the CSIR; see Remark 25. Moreover, the information theory for in-block feedback [22] applies to beamforming [106] and intelligent reflecting surfaces [107,108]. One may also apply GMI to multi-user channels with in-block feedback, such as multi-access and broadcast channels. Finally, it is important to develop improved capacity upper bounds. The standard approach here is the duality framework described in [97,109]; see also ([110], page 128).

Funding: This work was supported by the 6G Future Lab Bavaria funded by the Bavarian State Ministry of Science and the Arts, the project 6G-life funded by the Germany Federal Ministry for Education and Research (BMBF), and by the German Research Foundation (DFG) through projects 390777439 and 509917421.

Acknowledgments: The author wishes to thank the reviewers for their helpful comments and W. Zhang for sending his recent paper [50].

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Special Functions

This appendix reviews three classes of functions that we use to analyze information rates: the non-central chi-squared distribution, the exponential integral, and gamma functions.

Appendix A.1. Non-Central Chi-Squared Distribution

The non-central chi-squared distribution with two degrees of freedom is the probability distribution of $Y = |x + Z|^2$ where $x \in \mathbb{C}$ and $Z \sim \mathcal{CN}(0, 2)$. The density is

$$p(y) = \frac{1}{2} e^{-(y+|x|^2)/2} I_0(|x|\sqrt{y}) \cdot 1(y \geq 0) \quad (\text{A1})$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind of order zero. The cumulative distribution function is

$$\Pr[Y \leq t] = 1 - Q_1(|x|, \sqrt{t}) \quad (\text{A2})$$

where $Q_1(\cdot)$ is the Marcum Q-function of order 1. Observe that if we change Z to $Z \sim \mathcal{CN}(0, \sigma^2)$ then for $Y = |x + Z|^2$ we instead have

$$\Pr[Y \leq t] = 1 - Q_1\left(\sqrt{2|x|^2/\sigma^2}, \sqrt{2t/\sigma^2}\right). \quad (\text{A3})$$

Appendix A.2. Exponential Integral

The exponential integral is defined for $x > 0$ as

$$E_1(x) = \int_x^\infty \frac{e^{-t}}{t} dt. \quad (\text{A4})$$

The derivative of $E_1(x)$ is

$$\frac{dE_1(x)}{dx} = \frac{-e^{-x}}{x}. \quad (\text{A5})$$

For small x one may apply ([111], Equation (3))

$$E_1(x) \approx -\gamma - \log x + x \tag{A6}$$

where $\gamma \approx 0.57721$ is Euler’s constant. For large x we have

$$E_1(x) \approx \frac{e^{-x}}{x} \left(1 - \frac{1}{x} + \frac{2}{x^2} - \frac{6}{x^3} \right). \tag{A7}$$

We have the bounds [112]

$$\frac{1}{2} \log \left(1 + \frac{2}{x} \right) < e^x E_1(x) < \log \left(1 + \frac{1}{x} \right) \tag{A8}$$

$$\frac{1}{x+1} < e^x E_1(x) < \frac{x+1}{x(x+2)}. \tag{A9}$$

Using integration by parts, for $x > 0$, we have

$$\int_x^\infty e^{-t} \log t \, dt = E_1(x) + e^{-x} \log(x) \tag{A10}$$

$$\int_x^\infty e^{-t} \frac{1}{t^2} \, dt = \frac{e^{-x}}{x} - E_1(x). \tag{A11}$$

Using the translation $\tilde{t} = t + y$ we also have

$$\int_x^\infty e^{-t} \frac{t}{t+y} \, dt = e^{-x} - y e^y E_1(x+y) \tag{A12}$$

$$\int_x^\infty e^{-t} \frac{t}{(t+y)^2} \, dt = -e^{-x} \frac{y}{x+y} + (y+1) e^y E_1(x+y) \tag{A13}$$

$$\int_x^\infty e^{-t} \frac{t^2}{(t+y)^2} \, dt = e^{-x} \left(1 + \frac{y^2}{x+y} \right) - y(y+2) e^y E_1(x+y). \tag{A14}$$

Appendix A.3. Gamma Functions

The upper and lower incomplete gamma functions are the respective

$$\Gamma(s, t) = \int_t^\infty e^{-g} g^{s-1} \, dg \tag{A15}$$

$$\gamma(s, t) = \int_0^t e^{-g} g^{s-1} \, dg. \tag{A16}$$

For instance, we have $\Gamma(1, t) = e^{-t}$ and $\gamma(1, t) = 1 - e^{-t}$. We further have $\Gamma(0, t) = E_1(t)$ where $E_1(x)$ is the exponential integral defined in Appendix A.2.

The Gamma function is $\Gamma(s) = \Gamma(s, 0) = \gamma(s, \infty)$ and for positive integers n we have

$$\Gamma(n) = (n-1)!, \quad \Gamma\left(n - \frac{1}{2}\right) = \frac{(2n-2)!}{4^{n-1} (n-1)!} \sqrt{\pi}.$$

For example, the following cases are used in Section 8.4:

$$\Gamma(1) = \Gamma(2) = 1, \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}, \quad \Gamma\left(\frac{3}{2}\right) = \frac{\sqrt{\pi}}{2}, \quad \Gamma\left(\frac{5}{2}\right) = \frac{3}{4} \sqrt{\pi}.$$

The value $\Gamma(0)$ is undefined but we have $\lim_{x \rightarrow 0^+} \Gamma(x) = \infty$.

Appendix B. Forward Model GMIs with $K = 2$

This appendix studies $K = 2$ GMIs to develop high and low SNR capacity scaling results. Consider the independent random variables $Z \sim \mathcal{CN}(0,1)$ and $X \sim \mathcal{CN}(0,P)$. We need the following expression for the event $\mathcal{E} = \{|X + Z|^2 \geq t_R\}$:

$$\begin{aligned} \mathbb{E}\left[|Z|^2 \mid \mathcal{E}\right] &= \int_{\mathbb{C}} p_{Z|\mathcal{E}}(z) |z|^2 dz \\ &= \frac{1}{\Pr[\mathcal{E}]} \int_{\mathbb{C}} \frac{e^{-|z|^2}}{\pi} |z|^2 \Pr\left[|X + z|^2 \geq t_R\right] dz \\ &= e^{t_R/(1+P)} \int_0^\infty e^{-g} g Q_1\left(\sqrt{\frac{2g}{P}}, \sqrt{\frac{2t_R}{P}}\right) dg. \end{aligned} \tag{A17}$$

The integral can be computed using ([113], Equation (12)) with $k = 2, m = 1, p = 1$, the Gamma functions above, and the following identities for Kummer’s confluent hypergeometric function:

$${}_1F_1(1;2;z) = (e^z - 1)/z, \quad {}_1F_1(2;2;z) = e^z.$$

The result is

$$\mathbb{E}\left[|Z|^2 \mid |X + Z|^2 \geq t_R\right] = 1 + \frac{t_R}{(1+P)^2}. \tag{A18}$$

Appendix B.1. On-Off Fading

Consider on-off fading as in Section 3.3 and the $K = 2$ partition in Remark 14 with $h_2 = \sqrt{2}$. We compute

$$\Pr[\mathcal{E}_2] = \sum_{h=0,\sqrt{2}} \Pr[H = h] \Pr[\mathcal{E}_2 | H = h] = \frac{1}{2} e^{-t_R} + \frac{1}{2} e^{-t_R/(1+2P)}. \tag{A19}$$

If $t_R = P^{\lambda_R} + b$ where $0 < \lambda_R < 1$ and b is a real constant then $\Pr[\mathcal{E}_2] \rightarrow 1/2$ as $P \rightarrow \infty$, as desired. We further have

$$\Pr[H = 0 | \mathcal{E}_2] = \frac{e^{-t_R}}{2\Pr[\mathcal{E}_2]} \tag{A20}$$

$$\Pr[H = \sqrt{2} | \mathcal{E}_2] = \frac{e^{-t_R/(1+2P)}}{2\Pr[\mathcal{E}_2]}. \tag{A21}$$

The choice $t_R = P^{\lambda_R} + b$ gives $\Pr[H = \sqrt{2} | \mathcal{E}_2] \rightarrow 1$ as $P \rightarrow \infty$. In other words, the receiver can reliably determine H by choosing t_R to grow with P , but not too fast.

We next compute

$$\begin{aligned} \mathbb{E}\left[|Y|^2 | \mathcal{E}_2\right] &= \sum_{h=0,\sqrt{2}} \Pr[H = h | \mathcal{E}_2] \mathbb{E}\left[|Y|^2 | \mathcal{E}_2, H = h\right] \\ &= \frac{e^{-t_R}(t_R + 1) + e^{-t_R/(1+2P)}(t_R + 1 + 2P)}{2\Pr[\mathcal{E}_2]}. \end{aligned} \tag{A22}$$

The choice $t_R = P^{\lambda_R} + b$ makes $\mathbb{E}\left[|Y|^2 | \mathcal{E}_2\right] / (1 + 2P) \rightarrow 1$ as $P \rightarrow \infty$. Finally, we compute

$$\begin{aligned} \mathbb{E}\left[|Y - \sqrt{2}X|^2 | \mathcal{E}_2\right] &= \sum_{h=0,\sqrt{2}} \Pr[H = h | \mathcal{E}_2] \mathbb{E}\left[|Y - \sqrt{2}X|^2 | \mathcal{E}_2, H = h\right] \\ &= \frac{1}{2\Pr[\mathcal{E}_2]} \left\{ e^{-t_R}(t_R + 1 + 2P) + e^{-t_R/(1+2P)} \left(1 + \frac{t_R}{(1+2P)^2}\right) \right\} \end{aligned} \tag{A23}$$

where the last step uses (A18). The choice $t_R = P^{\lambda_R} + b$ makes $\mathbb{E}\left[|Y - \sqrt{2}X|^2 | \mathcal{E}_2\right] \rightarrow 1$ as $P \rightarrow \infty$.

Appendix B.2. On-Off Fading, Partial CSIR, and Full CSIT

The analysis for Section 7.3 is similar to that of Appendix B.1. Consider the GMI (259) and observe that we can replace $2P$ with $4P$ in (A19)–(A22). We also have

$$\begin{aligned} E\left[|Y - \sqrt{4P}U|^2 \mid \mathcal{E}_2\right] &= \sum_{h=0, \sqrt{2}} \Pr[H = h \mid \mathcal{E}_2] E\left[|Y - \sqrt{4P}U|^2 \mid \mathcal{E}_2, H = h\right] \\ &= \frac{1}{2\Pr[\mathcal{E}_2]} \left\{ e^{-t_R}(t_R + 1 + 4P) + e^{-t_R/(1+4P)} \left(1 + \frac{t_R}{(1+4P)^2}\right) \right\}. \end{aligned} \tag{A24}$$

The choice $t_R = P^{\lambda_R} + b$ as in Appendix B.1 gives (260).

Appendix B.3. On-Off Fading, Partial CSIR, and CSIT@R

The analysis for Section 7.4 is similar to that of Appendices B.1 and B.2. We compute

$$\Pr[\mathcal{E}_2 \mid S_R = 0] = \bar{\epsilon} e^{-t_R} + \epsilon e^{-t_R/[1+2P(0)]} \tag{A25}$$

$$\Pr[\mathcal{E}_2 \mid S_R = \sqrt{2}] = \epsilon e^{-t_R} + \bar{\epsilon} e^{-t_R/[1+2P(\sqrt{2})]}. \tag{A26}$$

Suppose $P(0)$ and $P(\sqrt{2})$ both scale in proportion to P . If we choose $t_R = P^{\lambda_R} + b$ as in Appendix B.1 then $\Pr[\mathcal{E}_2 \mid S_R = 0] \rightarrow \epsilon$ and $\Pr[\mathcal{E}_2 \mid S_R = \sqrt{2}] \rightarrow \bar{\epsilon}$ as $P \rightarrow \infty$. We also have

$$\Pr[H = 0 \mid \mathcal{E}_2, S_R = 0] = \frac{\bar{\epsilon} e^{-t_R}}{\Pr[\mathcal{E}_2 \mid S_R = 0]} \tag{A27}$$

$$\Pr[H = \sqrt{2} \mid \mathcal{E}_2, S_R = 0] = \frac{\epsilon e^{-t_R/[1+2P(0)]}}{\Pr[\mathcal{E}_2 \mid S_R = 0]} \tag{A28}$$

and similarly for the probabilities $\Pr[H = 0 \mid \mathcal{E}_2, S_R = \sqrt{2}]$ and $\Pr[H = \sqrt{2} \mid \mathcal{E}_2, S_R = \sqrt{2}]$. Choosing $t_R = P^{\lambda_R} + b$ gives the desired behavior $\Pr[H = \sqrt{2} \mid \mathcal{E}_2, S_R = 0] \rightarrow 1$ and $\Pr[H = \sqrt{2} \mid \mathcal{E}_2, S_R = \sqrt{2}] \rightarrow 1$ as $P \rightarrow \infty$. Again, the receiver can reliably determine H by choosing t_R to grow with P , but not too fast.

We next have

$$E\left[|Y|^2 \mid \mathcal{E}_2, S_R = 0\right] = \frac{\bar{\epsilon} e^{-t_R}(t_R + 1) + \epsilon e^{-t_R/[1+2P(0)]}(t_R + 1 + 2P(0))}{\Pr[\mathcal{E}_2 \mid S_R = 0]}. \tag{A29}$$

The expression for $E\left[|Y|^2 \mid \mathcal{E}_2, S_R = \sqrt{2}\right]$ is similar but ϵ and $\bar{\epsilon}$ are swapped and $P(0)$ is replaced with $P(\sqrt{2})$. We also have

$$\begin{aligned} E\left[|Y - \sqrt{2}X(0)|^2 \mid \mathcal{E}_2, S_R = 0\right] &= \frac{1}{\Pr[\mathcal{E}_2 \mid S_R = 0]} \left\{ \bar{\epsilon} e^{-t_R}(t_R + 1 + 2P(0)) + \right. \\ &\quad \left. \epsilon e^{-t_R/[1+2P(0)]} \left(1 + \frac{t_R}{(1+2P(0))^2}\right) \right\}. \end{aligned} \tag{A30}$$

The expression for $E\left[|Y - \sqrt{2}X(0)|^2 \mid \mathcal{E}_2, S_R = \sqrt{2}\right]$ is similar: swap ϵ and $\bar{\epsilon}$ and replace $P(0)$ with $P(\sqrt{2})$. The choice $t = P^{\lambda_R} + b$ makes all terms in (265) behave as desired. We thus obtain (266).

Appendix B.4. Rayleigh Fading, No CSIR, full CSIT, and TCI

The analysis for Section 8.4 is similar to that of Appendices B.1–B.3, but we now have a continuous H . Recall that $\mathcal{E}_2 = \{|Y|^2 \geq t_R\}$ and $Y = \sqrt{P(h)}U + Z$ where $P(h) = 0$ for $g < t$ and $P(h) = \hat{P}$ otherwise. We compute

$$\begin{aligned} \Pr[\mathcal{E}_2] &= \Pr[G < t]\Pr[\mathcal{E}_2|G < t] + \Pr[G \geq t]\Pr[\mathcal{E}_2|G \geq t] \\ &= (1 - e^{-t})e^{-tR} + e^{-t}e^{-tR/(1+\hat{P})} \end{aligned} \tag{A31}$$

where we used $\Pr[\mathcal{E}_2|G < t] = \Pr[|Z|^2 \geq t_R]$ and similarly for $\Pr[\mathcal{E}_2|G \geq t]$. For example, for the t and t_R in (296) we find that $\Pr[\mathcal{E}_2] \rightarrow 1$ as P grows. Similarly, for the t and t_R in (299) we find that $\Pr[\mathcal{E}_2] \approx e^{-t-1}$ as P decreases.

We write

$$\begin{aligned} \mathbb{E}[|Y|^2|\mathcal{E}_2] &= \Pr[G < t|\mathcal{E}_2]\mathbb{E}[|Z|^2|\mathcal{E}_2, G < t] \\ &\quad + \Pr[G \geq t|\mathcal{E}_2]\mathbb{E}[|\sqrt{\hat{P}}U + Z|^2|\mathcal{E}_2, G \geq t] \\ &= \frac{(1 - e^{-t})e^{-tR}(t_R + 1) + e^{-t}e^{-tR/(1+\hat{P})}(t_R + 1 + \hat{P})}{\Pr[\mathcal{E}_2]} \end{aligned} \tag{A32}$$

For the t and t_R in (296) we have $\mathbb{E}[|Y|^2|\mathcal{E}_2]/(1 + \hat{P}) \rightarrow 1$ as P grows. Similarly, for the t and t_R in (299) we find that $\mathbb{E}[|Y|^2|\mathcal{E}_2]/(1 + 2\hat{P}) \rightarrow 1$ as P decreases. Next, we write

$$\begin{aligned} &\mathbb{E}[|Y - \sqrt{\hat{P}}U|^2|\mathcal{E}_2] \\ &= \Pr[G < t|\mathcal{E}_2]\mathbb{E}[|Z - \sqrt{\hat{P}}U|^2|Z|^2 \geq t_R] + \Pr[G \geq t|\mathcal{E}_2]\mathbb{E}[|Z|^2|\sqrt{\hat{P}}U + Z|^2 \geq t_R] \\ &= \frac{1}{\Pr[\mathcal{E}_2]} \left\{ (1 - e^{-t})e^{-tR}(t_R + 1 + \hat{P}) + e^{-t}e^{-tR/(1+\hat{P})} \left(1 + \frac{t_R}{(1 + \hat{P})^2} \right) \right\} \end{aligned} \tag{A33}$$

For the t and t_R in (296) the expression (A33) approaches 1 as P grows. Similarly, for the t and t_R in (299) we find that (A33) approaches 1 as P decreases.

Appendix C. Conditional Second-Order Statistics

This appendix shows how to compute conditional second-order statistics for the reverse model GMIs and the forward model GMIs with $K = \infty$. Suppose that U, Y are jointly CSCG given $H = h$. Using (25) and (26), we have

$$\mathbb{E}[U|Y = y, H = h] = \frac{\mathbb{E}[UY^*|H = h]}{\mathbb{E}[|Y|^2|H = h]} \cdot y \tag{A34}$$

$$\text{Var}[U|Y = y, H = h] = \mathbb{E}[|U|^2|H = h] - \frac{|\mathbb{E}[UY^*|H = h]|^2}{\mathbb{E}[|Y|^2|H = h]} \tag{A35}$$

Now consider the channel $Y = HX + Z$ where $X = \sqrt{P(s_T)}e^{j\phi(s_T)}U$ with $U \sim \mathcal{CN}(0, 1)$. We may write

$$\mathbb{E}[U|Y = y, S_R = s_R] = \int_{\mathbb{C} \times \mathcal{S}_T} p(h, s_T|y, s_R) \frac{h^* \sqrt{P(s_T)}e^{j\phi(s_T)}y}{1 + |h|^2P(s_T)} ds_T dh \tag{A36}$$

and

$$\begin{aligned} \mathbb{E}[|U|^2|Y = y, S_R = s_R] &= \int_{\mathbb{C} \times \mathcal{S}_T} p(h, s_T|y, s_R) \\ &\quad \left(\frac{1}{1 + |h|^2P(s_T)} + \frac{|h|^2P(s_T)|y|^2}{(1 + |h|^2P(s_T))^2} \right) ds_T dh \end{aligned} \tag{A37}$$

Appendix C.1. No CSIR, No CSIT

Consider $S_R = S_T = 0$. The expectations in (A36) and (A37) are computed via

$$p(h|y) = \frac{p(h)p(y|h)}{p(y)}. \tag{A38}$$

The expression (A36) with $\phi(0) = 0$ gives

$$E[U|Y = y] = \int_{\mathbb{C}} p(h|y) \frac{h^* \sqrt{P} y}{1 + |h|^2 P} dh. \tag{A39}$$

Similarly, the expression (A37) gives

$$\begin{aligned} E[|U|^2|Y = y] &= \int_{\mathbb{C}} p(h|y) E[|X|^2|Y = y, H = h] dh \\ &= \int_{\mathbb{C}} p(h|y) \left(\frac{1}{1 + |h|^2 P} + \frac{|h|^2 P |y|^2}{(1 + |h|^2 P)^2} \right) dh \end{aligned} \tag{A40}$$

We may now compute $\text{Var}[U|Y = y]$ using (A39) and (A40). For the expressions (69) and (70), one may use

$$E[X|Y = y] = \sqrt{P} E[U|Y = y], \quad E[|X|^2|Y = y] = P E[|U|^2|Y = y].$$

For example, for on-off fading as in Section 3.3 we compute

$$E[X|Y = y] = P_{H|Y}(\sqrt{2}|y) \frac{\sqrt{2}P}{1 + 2P} \cdot y \tag{A41}$$

$$E[|X|^2|Y = y] = P_{H|Y}(0|y) P + P_{H|Y}(\sqrt{2}|y) \left(\frac{P}{1 + 2P} + \frac{2P^2|y|^2}{(1 + 2P)^2} \right) \tag{A42}$$

and therefore

$$\text{Var}[X|Y = y] = P_{H|Y}(0|y) P + P_{H|Y}(\sqrt{2}|y) \left(\frac{P}{1 + 2P} + \frac{2P^2|y|^2}{(1 + 2P)^2} P_{H|Y}(0|y) \right) \tag{A43}$$

where $P_{H|Y}(\sqrt{2}|y) = 1 - P_{H|Y}(0|y)$ and

$$P_{H|Y}(0|y) = \frac{e^{-|y|^2}}{e^{-|y|^2} + \frac{1}{1+2P} e^{-|y|^2/(1+2P)}}. \tag{A44}$$

For Rayleigh fading as in Section 8.1, the density (A38) is

$$p(h|y) = \frac{e^{-g} e^{-|y|^2/(1+gP)}}{\pi^2(1 + gP)} \cdot \frac{1}{p(y)}$$

where $g = |h|^2$. Moreover, $p(y)$ in (268) depends on g only. We thus have $E[U|Y = y] = 0$ and the integrand in (A40) depends on g and $|y|^2$ only.

Appendix C.2. Full CSIR, Partial CSIT

Consider $S_R = H$ and partial S_T . The expectations in (A36) and (A37) are computed via (194) that we repeat here:

$$p(h, s_T|y, s_R) = \delta(h - s_R) \frac{p(s_T|h) p(y|h, s_T)}{p(y|h)}.$$

For on-off fading as in Section 7.2, the expression (A36) with $\phi(0) = 0$ gives the expectations $E[U|Y = y, H = 0] = 0$ and

$$E[U|Y = y, H = \sqrt{2}] = \sum_{s_T=0,2} P_{S_T|Y,H}(s_T|y, \sqrt{2}) \frac{\sqrt{2P(s_T)} y}{1 + 2P(s_T)}$$

and, similarly, (A37) gives $E[|U|^2|Y = y, H = 0] = 1$ and

$$E[|U|^2|Y = y, H = \sqrt{2}] = \sum_{s_T=0,2} P_{S_T|Y,H}(s_T|y, \sqrt{2}) \left(\frac{1}{1 + 2P(s_T)} + \frac{2P(s_T)|y|^2}{(1 + 2P(s_T))^2} \right)$$

where $P_{S_T|Y,H}(2|y, \sqrt{2}) = 1 - P_{S_T|Y,H}(0|y, \sqrt{2})$ and

$$P_{S_T|Y,H}(0|y, \sqrt{2}) = \frac{\frac{\epsilon}{1+2P(0)} e^{-|y|^2/(1+2P(0))}}{\frac{\epsilon}{1+2P(0)} e^{-|y|^2/(1+2P(0))} + \frac{\epsilon}{1+2P(2)} e^{-|y|^2/(1+2P(2))}}.$$

For Rayleigh fading as in Section 8.3, the sums over $s_T = 0, 2$ become sums over $s_T = 0, 1$ and the probabilities $P(s_T|y, h)$ take on similar forms as above.

Appendix C.3. Partial CSIR, Full CSIT

Consider $S_T = H$ and partial S_R . The expectations in (A36) and (A37) are computed via (201) that we repeat here:

$$p(h, s_T|y, s_R) = \delta(s_T - h) \frac{p(h|s_R) p(y|h, s_R)}{p(y|s_R)}.$$

For on-off fading with $S_R = 0$ as in Section 7.3, the expression (A36) with $\phi(0) = 0$ gives

$$E[U|Y = y] = P_{H|Y}(\sqrt{2}|y) \frac{\sqrt{4P} y}{1 + 4P}$$

and (A37) gives

$$E[|U|^2|Y = y] = P_{H|Y}(0|y) + P_{H|Y}(\sqrt{2}|y) \left(\frac{1}{1 + 4P} + \frac{4P|y|^2}{(1 + 4P)^2} \right)$$

where $P_{H|Y}(\sqrt{2}|y) = 1 - P_{H|Y}(0|y)$ and

$$P_{H|Y}(0|y) = \frac{e^{-|y|^2}}{e^{-|y|^2} + \frac{1}{1+4P} e^{-|y|^2/(1+4P)}}.$$

For Rayleigh fading with $S_R = 0$ and TCI as in Section 8.4, the expressions (A36) and (A37) give (cf. (A41) and (A42))

$$E[U|Y = y] = \Pr[G \geq t|Y = y] \frac{\sqrt{\hat{P}} y}{1 + \hat{P}}$$

$$E[|U|^2|Y = y] = \Pr[G < t|Y = y] + \Pr[G \geq t|Y = y] \left(\frac{1}{1 + \hat{P}} + \frac{\hat{P}|y|^2}{(1 + \hat{P})^2} \right)$$

and therefore (cf. (A43))

$$\begin{aligned} \text{Var}[U|Y = y] &= \Pr[G < t|Y = y] + \Pr[G \geq t|Y = y] \\ &\cdot \left(\frac{1}{1 + \hat{P}} + \frac{\hat{P}|y|^2}{(1 + \hat{P})^2} \Pr[G < t|Y = y] \right) \end{aligned}$$

where (cf. (A44))

$$\Pr[G < t|Y = y] = \frac{(1 - e^{-t}) e^{-|y|^2}}{(1 - e^{-t}) e^{-|y|^2} + e^{-t} \frac{1}{1 + \hat{P}} e^{-|y|^2/(1 + \hat{P})}}.$$

Appendix C.4. Partial CSIR, CSIT@R

Consider $S_T = S_R$ and partial S_R . The expectations in (A36) and (A37) are computed via (234) that we repeat here:

$$p(h, s_T|y, s_R) = \delta(s_T - f(s_R)) \frac{p(h|s_R) p(y|h, s_R)}{p(y|s_R)}.$$

For on-off fading as in Section 7.3, the expression (A36) with $\phi(0) = 0$ gives

$$\begin{aligned} \mathbb{E}[U|Y = y, S_R = 0] &= P_{H|Y, S_R}(1|y, 0) \frac{\sqrt{2P(0)} y}{1 + 2P(0)} \\ \mathbb{E}[U|Y = y, S_R = \sqrt{2}] &= P_{H|Y, S_R}(1|y, \sqrt{2}) \frac{\sqrt{2P(\sqrt{2})} y}{1 + 2P(\sqrt{2})} \end{aligned}$$

and (A37) gives

$$\begin{aligned} \mathbb{E}[|U|^2|Y = y, S_R = 0] &= P_{H|Y, S_R}(0|y, 0) \\ &+ P_{H|Y, S_R}(1|y, 0) \left(\frac{1}{1 + 2P(0)} + \frac{2P(0)|y|^2}{(1 + 2P(0))^2} \right) \\ \mathbb{E}[|U|^2|Y = y, S_R = \sqrt{2}] &= P_{H|Y, S_R}(0|y, \sqrt{2}) \\ &+ P_{H|Y, S_R}(1|y, \sqrt{2}) \left(\frac{1}{1 + 2P(\sqrt{2})} + \frac{2P(\sqrt{2})|y|^2}{(1 + 2P(\sqrt{2}))^2} \right) \end{aligned}$$

where $P_{H|Y, S_R}(\sqrt{2}|y, s_R) = 1 - P_{H|Y, S_R}(0|y, s_R)$ and

$$\begin{aligned} P_{H|Y, S_R}(0|y, 0) &= \frac{\bar{\epsilon} e^{-|y|^2}}{\bar{\epsilon} e^{-|y|^2} + \frac{\epsilon}{1 + 2P(0)} e^{-|y|^2/(1 + 2P(0))}} \\ P_{H|Y, S_R}(0|y, \sqrt{2}) &= \frac{\epsilon e^{-|y|^2}}{\epsilon e^{-|y|^2} + \frac{\bar{\epsilon}}{1 + 2P(\sqrt{2})} e^{-|y|^2/(1 + 2P(\sqrt{2}))}}. \end{aligned}$$

For Rayleigh fading as in Section 8.5, the probabilities $P(h|y, s_R)$ take on similar forms as above.

Appendix D. Proof of Lemma 2 and (119)

We prove Lemma 2 by using the same steps as in the proof of Proposition 1. The GMI (102) with a vector \underline{Y} is

$$\begin{aligned}
 I_s(A; \underline{Y}) &= \log \det \left(\mathbf{I} + (\mathbf{Q}_{\underline{Z}}/s)^{-1} \mathbf{H} \mathbf{Q}_{\underline{X}} \mathbf{H}^\dagger \right) \\
 &\quad + \mathbb{E} \left[\underline{Y}^\dagger \left(\mathbf{Q}_{\underline{Z}}/s + \mathbf{H} \mathbf{Q}_{\underline{X}} \mathbf{H}^\dagger \right)^{-1} \underline{Y} \right] \\
 &\quad - \mathbb{E} \left[(\underline{Y} - \mathbf{H} \underline{X})^\dagger (\mathbf{Q}_{\underline{Z}}/s)^{-1} (\underline{Y} - \mathbf{H} \underline{X}) \right].
 \end{aligned} \tag{A45}$$

One can again set $s = 1$. Choosing $\mathbf{H} = \tilde{\mathbf{H}}$ and $\mathbf{Q}_{\underline{Z}} = \tilde{\mathbf{Q}}_{\underline{Z}}$ then gives (112).

Next, consider the channel $\underline{Y}_a = \tilde{\mathbf{H}} \underline{X} + \tilde{\underline{Z}}$ where $\tilde{\underline{Z}}$ is CSCG with covariance matrix $\mathbf{Q}_{\tilde{\underline{Z}}}$ and $\tilde{\underline{Z}}$ is independent of \underline{X} . Generalizing (50) and (51), we compute $\mathbf{Q}_{\underline{Y}_a} = \mathbf{Q}_{\underline{Y}}$ and

$$\mathbb{E} \left[(\underline{Y}_a - \tilde{\mathbf{H}} \underline{X}) (\underline{Y}_a - \tilde{\mathbf{H}} \underline{X})^\dagger \right] = \mathbb{E} \left[(\underline{Y} - \mathbf{H} \underline{X}) (\underline{Y} - \mathbf{H} \underline{X})^\dagger \right]. \tag{A46}$$

In other words, the second-order statistics for the two channels with outputs \underline{Y} (the actual channel output) and \underline{Y}_a are the same. Moreover, the GMI (112) is the mutual information $I(A; \underline{Y}_a)$. Using (104) and (A45), for any s , \mathbf{H} and $\mathbf{Q}_{\underline{Z}}$ we have

$$I(A; \underline{Y}_a) = \log \det \left(\mathbf{I} + \mathbf{Q}_{\tilde{\underline{Z}}}^{-1} \tilde{\mathbf{H}} \mathbf{Q}_{\underline{X}} \tilde{\mathbf{H}}^\dagger \right) \geq I_s(A; \underline{Y}_a) = I_s(A; \underline{Y}) \tag{A47}$$

and equality holds if $\mathbf{H} = \tilde{\mathbf{H}}$ and $\mathbf{Q}_{\underline{Z}}/s = \mathbf{Q}_{\tilde{\underline{Z}}}$.

To prove (119), recall that $\text{tr}(\mathbf{A}\mathbf{B}) = \text{tr}(\mathbf{B}\mathbf{A})$ for matrices \mathbf{A} and \mathbf{B} with appropriate dimensions. Furthermore, for Hermitian matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ with the same dimensions we have

$$\text{tr}(\mathbf{A}\mathbf{B}\mathbf{C}) = \text{tr}((\mathbf{A}\mathbf{B}\mathbf{C})^\dagger) = \text{tr}(\mathbf{C}\mathbf{B}\mathbf{A}) = \text{tr}(\mathbf{A}\mathbf{C}\mathbf{B}). \tag{A48}$$

For notational convenience, consider the covariance matrix (117) with $s = 1$ and use

$$\begin{aligned}
 \mathbf{A} &= \mathbf{Q}_{\underline{Z}}, \quad \mathbf{B} = \left(\mathbf{H} \mathbf{Q}_{\underline{X}} \mathbf{H}^\dagger \right)^{-1/2} \left(\mathbf{Q}_{\underline{Y}} - \mathbf{Q}_{\underline{Z}} \right)^{1/2} \\
 \mathbf{C} &= \mathbf{Q}_{\tilde{\underline{Z}}}^{-1} \left(\mathbf{Q}_{\underline{Y}} - \mathbf{Q}_{\underline{Z}} \right)^{1/2} \left(\mathbf{H} \mathbf{Q}_{\underline{X}} \mathbf{H}^\dagger \right)^{-1/2}
 \end{aligned}$$

to compute (cf. (A45))

$$\mathbb{E} \left[(\underline{Y} - \mathbf{H} \underline{X})^\dagger \mathbf{Q}_{\underline{Z}}^{-1} (\underline{Y} - \mathbf{H} \underline{X}) \right] = \text{tr} \left(\mathbf{Q}_{\underline{Z}} \mathbf{Q}_{\underline{Z}}^{-1} \right) \stackrel{(a)}{=} \text{tr} \left(\left(\mathbf{Q}_{\underline{Y}} - \mathbf{Q}_{\underline{Z}} \right) \left(\mathbf{H} \mathbf{Q}_{\underline{X}} \mathbf{H}^\dagger \right)^{-1} \right) \tag{A49}$$

where step (a) follows by (A48). Next, by using (117) we have

$$\begin{aligned}
 \left(\mathbf{Q}_{\underline{Z}} + \mathbf{H} \mathbf{Q}_{\underline{X}} \mathbf{H}^\dagger \right)^{-1} &= \left(\mathbf{Q}_{\underline{Y}} - \mathbf{Q}_{\underline{Z}} \right)^{1/2} \left(\mathbf{H} \mathbf{Q}_{\underline{X}} \mathbf{H}^\dagger \right)^{-1/2} \mathbf{Q}_{\underline{Y}}^{-1} \\
 &\quad \cdot \left(\mathbf{H} \mathbf{Q}_{\underline{X}} \mathbf{H}^\dagger \right)^{-1/2} \left(\mathbf{Q}_{\underline{Y}} - \mathbf{Q}_{\underline{Z}} \right)^{1/2}
 \end{aligned} \tag{A50}$$

and therefore (cf. (A45))

$$\begin{aligned}
 \mathbb{E} \left[\underline{Y}^\dagger \left(\mathbf{Q}_{\underline{Z}} + \mathbf{H} \mathbf{Q}_{\underline{X}} \mathbf{H}^\dagger \right)^{-1} \underline{Y} \right] &= \text{tr} \left(\mathbf{Q}_{\underline{Y}} \left(\mathbf{Q}_{\underline{Z}} + \mathbf{H} \mathbf{Q}_{\underline{X}} \mathbf{H}^\dagger \right)^{-1} \right) \\
 &\stackrel{(a)}{=} \text{tr} \left(\left(\mathbf{Q}_{\underline{Y}} - \mathbf{Q}_{\underline{Z}} \right) \left(\mathbf{H} \mathbf{Q}_{\underline{X}} \mathbf{H}^\dagger \right)^{-1} \right)
 \end{aligned} \tag{A51}$$

where step (a) again follows by (A48). We are thus left with the logarithm term in (A45). Finally, the determinant in (A45) is

$$\det(\mathbf{I} + \mathbf{Q}_{\underline{Z}}^{-1} \mathbf{H} \mathbf{Q}_{\underline{X}} \mathbf{H}^\dagger) = \det(\mathbf{Q}_{\underline{Z}}^{-1} \mathbf{Q}_{\underline{Y}}) \tag{A52}$$

where we applied (117) and Sylvester’s identity (33).

Appendix E. Proof of Lemma 3

Let $\bar{P} = E[|\bar{X}|^2]$ and write

$$\bar{X} = \sqrt{\bar{P}} \bar{U}, \quad X(s_T) = \sqrt{P(s_T)} U(s_T). \tag{A53}$$

Since the $U(s_T)$ are CSCG we have

$$U(s'_T) = \rho(s'_T, s_T) U(s_T) + Z(s'_T) \tag{A54}$$

where $\rho(s'_T, s_T) = E[U(s'_T) U(s_T)^*]$ and

$$Z(s'_T) \sim \mathcal{CN}(0, 1 - |\rho(s'_T, s_T)|^2) \tag{A55}$$

is independent of $U(s_T)$. As in (109), define

$$\begin{aligned} \bar{X} &= \sum_{s'_T} w(s'_T) X(s'_T) \\ &= \sum_{s_T} w(s'_T) \sqrt{P(s'_T)} [U(s_T) \rho(s'_T, s_T) + Z(s'_T)] \\ &= \sqrt{\bar{P}} \bar{\rho}(s_T) U(s_T) + \sum_{s'_T} w(s'_T) \sqrt{P(s'_T)} Z(s'_T) \end{aligned} \tag{A56}$$

where, assuming that $\bar{P} > 0$, we have

$$\bar{\rho}(s_T) = E[\bar{U} U(s_T)^*] = \sum_{s'_T} w(s'_T) \sqrt{\frac{P(s'_T)}{\bar{P}}} \rho(s'_T, s_T). \tag{A57}$$

Observe that $\sqrt{\bar{P}} \bar{\rho}(s_T) U(s_T)$ is the LMMSE estimate of \bar{X} given $U(s_T)$.

Using Lemma 2, we have the auxiliary variables

$$\tilde{h} = \frac{E[Y \bar{X}^*]}{\bar{P}}, \quad \tilde{\sigma}^2 = E[|Y|^2] - |\tilde{h}|^2 \bar{P} \tag{A58}$$

and the GMI

$$I_1(A; Y) = \log \left(\frac{E[|Y|^2]}{E[|Y|^2] - |\tilde{h}|^2 \bar{P}} \right). \tag{A59}$$

If the $P(s_T)$ are fixed, then so is $E[|Y|^2]$ because $U(s_T)$ is CSCG and independent of Z given $S_T = s_T$. The GMI (A59) is thus maximized by maximizing $|\tilde{h}|^2 \bar{P}$. We compute

$$\begin{aligned} |\tilde{h}|^2 \bar{P} &= \left| \sum_{s_T} P_{S_T}(s_T) \frac{E[Y \bar{X}^* | S_T = s_T]}{\sqrt{\bar{P}}} \right|^2 \\ &\stackrel{(a)}{=} \left| \sum_{s_T} P_{S_T}(s_T) E[Y U(s_T)^* | S_T = s_T] \bar{\rho}(s_T)^* \right|^2 \end{aligned} \tag{A60}$$

$$\leq \left(\sum_{s_T} P_{S_T}(s_T) |E[Y U(s_T)^* | S_T = s_T]| \right)^2 \tag{A61}$$

where step (a) follows because we have the Markov chain $A - [U(S_T), S_T] - Y$ which implies that Y and the $Z(s'_T)$ in (A56) are independent give $S_T = s_T$.

Equality holds in (A61) if the summands in (A60) all have the same phase and $|\bar{\rho}(s_T)| = 1$ for all s_T . But this is possible by choosing $X(s_T)$ as given in (122) so that $U(s_T) = e^{j\phi(s_T)} U$. Moreover, choose the receiver weights as

$$w(\tilde{s}_T) = \sqrt{\frac{\bar{P}}{P(\tilde{s}_T)}} e^{-j\phi(\tilde{s}_T)} \tag{A62}$$

for one $\tilde{s}_T \in \mathcal{S}_T$ with $P(\tilde{s}_T) > 0$, and $w(s_T) = 0$ otherwise. We then have $\bar{X} = \sqrt{\bar{P}} U$ and

$$\rho(s'_T, s_T) = e^{j(\phi(s'_T) - \phi(s_T))}, \quad \bar{\rho}(s_T) = e^{-j\phi(s_T)} \tag{A63}$$

and the resulting maximal $I_1(A; Y)$ is given by (120) and (121).

Remark A1. The full correlation permits many choices for the $w(s_T)$; hence, these weights do not seem central to the design. However, including weights can be useful if the codebook is not designed for the CSIR. For example, suppose A has independent entries $X(s_T)$ for which we compute

$$\bar{\rho}(s_T) = \frac{w(s_T)\sqrt{P(s_T)}}{\sqrt{\sum_{s'_T} |w(s'_T)|^2 P(s'_T)}} \tag{A64}$$

and thus (A60) becomes

$$\frac{|\sum_{s_T} P_{S_T}(s_T) E[YX(s_T)^* | S_T = s_T] w(s_T)^*|^2}{\sum_{s_T} |w(s_T)|^2 P(s_T)}. \tag{A65}$$

Using Bergström’s inequality (or the Cauchy-Schwarz inequality), the expression (A65) is maximized by

$$w(s_T) = P_{S_T}(s_T) \frac{E[YX(s_T)^* | S_T = s_T]}{P(s_T)} \cdot c \tag{A66}$$

for some constant $c \neq 0$. The expression (A60) is therefore

$$\sum_{s_T} P_{S_T}(s_T | h)^2 |E[YU(s_T)^* | S_T = s_T]|^2 \tag{A67}$$

which is generally smaller than $E[|E[YU(s_T)^* | S_T]|]^2$ (apply $\sum_i a_i^2 \leq (\sum_i a_i)^2$ for non-negative a_i).

Remark A2. The following example shows that more general signaling and more general \bar{X} can be useful. Consider the channel with two equally-likely states $\mathcal{S}_T = \{+1, -1\}$ and $Y = |X| \exp(js_T \arg(X)) + Z$. We compute

$$\begin{aligned} E[YU(+1)^* | S_T = +1] &= \sqrt{P(1)} \\ E[YU(-1)^* | S_T = -1] &= 0 \\ \bar{\rho}(+1) &= \frac{w(1)\sqrt{P(1)} + w(-1)\sqrt{P(-1)}\rho(-1, +1)}{\sqrt{\bar{P}}} \end{aligned}$$

and one should choose $P(-1) = 0$ and $P(1) = 2P$ if the power constraint is $E[P(S_T)] \leq P$. We thus have

$$E[|Y|^2] = P + 1, \quad \bar{P} = \frac{P}{2}$$

and therefore (120) gives

$$I_1(A; Y) = \log\left(1 + \frac{P}{2 + P}\right).$$

However, one can achieve the rate $\log(1 + P)$ with other Gaussian \bar{X} , namely linear combinations of both the $X(s_T)$ and the $X(s_T)^*$ in (A56). This idea permits circularly asymmetric \bar{X} , also known as improper \bar{X} [114]. Alternatively, the transmitter can send the complex-conjugate symbols if $S_T = -1$.

Appendix F. Large K for Section 5.3

We complete Remark 49 by proceeding as in Appendix C.1. To generalize (70), we must deal with unit-rank matrices $\underline{y}\underline{y}^\dagger$ that do not have inverses. Consider first finite K . Conditioned on the event \mathcal{E}_k , we may write

$$\underline{Y} = \underline{y}_k + \epsilon^{1/2}\tilde{\underline{Z}}_k \tag{A68}$$

where $\underline{y}_k = E[\underline{Y}|\mathcal{E}_k]$ and $E[\tilde{\underline{Z}}_k|\mathcal{E}_k] = \underline{0}$. We abuse notation and write the conditional covariance matrix of $\tilde{\underline{Z}}_k$ as $\mathbf{Q}_{\tilde{\underline{Z}}_k}$, and we assume that $\mathbf{Q}_{\tilde{\underline{Z}}_k}$ is invertible. Define $\tilde{\underline{y}}_k = \mathbf{Q}_{\tilde{\underline{Z}}_k}^{-1/2}\underline{y}_k$ and compute

$$\mathbf{Q}_{\underline{Y}}^{(k)} = \epsilon \mathbf{Q}_{\tilde{\underline{Z}}_k}^{1/2} \left[\mathbf{I} + \frac{1}{\epsilon} \tilde{\underline{y}}_k \tilde{\underline{y}}_k^\dagger \right] \mathbf{Q}_{\tilde{\underline{Z}}_k}^{1/2} \tag{A69}$$

$$\left(\mathbf{Q}_{\underline{Y}}^{(k)}\right)^{-1} = \frac{1}{\epsilon} \mathbf{Q}_{\tilde{\underline{Z}}_k}^{-1/2} \left[\mathbf{I} - \frac{\tilde{\underline{y}}_k \tilde{\underline{y}}_k^\dagger}{\epsilon + \|\tilde{\underline{y}}\|^2} \right] \mathbf{Q}_{\tilde{\underline{Z}}_k}^{-1/2}. \tag{A70}$$

We further compute approximations for small ϵ :

$$\underline{y}_k^\dagger \left(\mathbf{Q}_{\underline{Y}}^{(k)}\right)^{-1} \underline{y}_k = \frac{\|\tilde{\underline{y}}_k\|^2}{\epsilon + \|\tilde{\underline{y}}_k\|^2} \approx 1 \tag{A71}$$

$$\mathbf{H}_k = \left(\underline{y}_k E[\bar{\underline{X}}^\dagger|\mathcal{E}_k] + \epsilon^{1/2} E[\tilde{\underline{Z}}_k \bar{\underline{X}}^\dagger|\mathcal{E}_k]\right) \left(\mathbf{Q}_{\underline{Y}}^{(k)}\right)^{-1} \approx \underline{y}_k E[\bar{\underline{X}}^\dagger|\mathcal{E}_k] \left(\mathbf{Q}_{\bar{\underline{X}}}^{(k)}\right)^{-1}. \tag{A72}$$

We can now treat the limit of large K for which ϵ approaches zero, i.e., we choose a different auxiliary model for each $\underline{Y} = \underline{y}$. Applying the Woodbury and Sylvester identities (32) and (33) several times, (158) becomes

$$I_1(A; \underline{Y}) = \int_{\mathbb{C}^N} p(\underline{y}) \left[\log \det \left(\mathbf{I} + \left(\mathbf{Q}_{\bar{\underline{X}}}^{(y)} - E_{\underline{y}} E_{\underline{y}}^\dagger \right)^{-1} \mathbf{Q}_{\bar{\underline{X}}} \left(\mathbf{Q}_{\bar{\underline{X}}}^{(y)} \right)^{-1} E_{\underline{y}} E_{\underline{y}}^\dagger \right) - \text{tr} \left(\left(\mathbf{Q}_{\bar{\underline{X}}}^{(y)} \left(\mathbf{D}_{\bar{\underline{X}}}^{(y)} \right)^{-1} \mathbf{Q}_{\bar{\underline{X}}}^{(y)} - E_{\underline{y}} E_{\underline{y}}^\dagger \right)^{-1} E_{\underline{y}} E_{\underline{y}}^\dagger \right) \right] d\underline{y} \tag{A73}$$

where

$$E_{\underline{y}} = E[\bar{\underline{X}}|\underline{Y} = \underline{y}], \quad \mathbf{Q}_{\bar{\underline{X}}}^{(y)} = E[\bar{\underline{X}} \bar{\underline{X}}^\dagger|\underline{Y} = \underline{y}], \quad \mathbf{D}_{\bar{\underline{X}}}^{(y)} = \mathbf{Q}_{\bar{\underline{X}}} - \mathbf{Q}_{\bar{\underline{X}}}^{(y)}.$$

If $\bar{\underline{X}}, \underline{Y}$ are jointly CSCG, then using (25) and (26) we have

$$E_{\underline{y}} = E[\bar{\underline{X}} \underline{Y}^\dagger] \mathbf{Q}_{\underline{Y}}^{-1} \cdot \underline{y} \tag{A74}$$

$$\mathbf{Q}_{\bar{\underline{X}}}^{(y)} - E_{\underline{y}} E_{\underline{y}}^\dagger = \mathbf{Q}_{\bar{\underline{X}}} - E[\bar{\underline{X}} \underline{Y}^\dagger] \mathbf{Q}_{\underline{Y}}^{-1} E[\bar{\underline{X}} \underline{Y}^\dagger]^\dagger. \tag{A75}$$

For example, if $\underline{Y} = \mathbf{H}\underline{X} + \underline{Z}$ where $\mathbf{H}, \underline{X}, \underline{Z}$ are mutually independent and $E[\underline{Z}] = \mathbf{0}$, then we have (cf. (A39))

$$\begin{aligned} \underline{E}_y &= \int_{\mathbb{C}^{N \times M}} p(\mathbf{h}|\underline{y}) E[\underline{X}|\underline{Y} = \underline{y}, \mathbf{H} = \mathbf{h}] d\mathbf{h} \\ &= \int_{\mathbb{C}^{N \times M}} p(\mathbf{h}|\underline{y}) \mathbf{Q}_{\underline{X}} \mathbf{h}^\dagger (\mathbf{I} + \mathbf{h}\mathbf{Q}_{\underline{X}}\mathbf{h}^\dagger)^{-1} \underline{y} d\mathbf{h} \end{aligned} \tag{A76}$$

$$= E\left[\mathbf{Q}_{\underline{X}}\mathbf{H}^\dagger (\mathbf{I} + \mathbf{H}\mathbf{Q}_{\underline{X}}\mathbf{H}^\dagger)^{-1} \Big| \underline{Y} = \underline{y}\right] \cdot \underline{y} \tag{A77}$$

where we have applied (A74) with conditioning on the event $\mathbf{H} = \mathbf{h}$. Similarly, we apply a conditional version of (A75) and the step (A76) to compute (cf. (A40))

$$\begin{aligned} \mathbf{Q}_{\underline{X}}^{(y)} &= \int_{\mathbb{C}^{N \times M}} p(\mathbf{h}|\underline{y}) E[\underline{X}\underline{X}^\dagger | \underline{Y} = \underline{y}, \mathbf{H} = \mathbf{h}] d\mathbf{h} \\ &= \int_{\mathbb{C}^{N \times M}} p(\mathbf{h}|\underline{y}) \left(\mathbf{Q}_{\underline{X}}^{(y, \mathbf{h})} + \underline{E}_{y, \mathbf{h}} \underline{E}_{y, \mathbf{h}}^\dagger\right) d\mathbf{h} \\ &= E\left[\mathbf{Q}_{\underline{X}}^{(y, \mathbf{H})} + \underline{E}_{y, \mathbf{H}} \underline{E}_{y, \mathbf{H}}^\dagger \Big| \underline{Y} = \underline{y}\right] \end{aligned} \tag{A78}$$

where

$$\begin{aligned} \mathbf{Q}_{\underline{X}}^{(y, \mathbf{h})} &= \mathbf{Q}_{\underline{X}} - \mathbf{Q}_{\underline{X}}\mathbf{h}^\dagger (\mathbf{I} + \mathbf{h}\mathbf{Q}_{\underline{X}}\mathbf{h}^\dagger)^{-1} \mathbf{h}\mathbf{Q}_{\underline{X}} \\ \underline{E}_{y, \mathbf{h}} &= E[\underline{X}|\underline{Y} = \underline{y}, \mathbf{H} = \mathbf{h}] = \mathbf{Q}_{\underline{X}} \mathbf{h}^\dagger (\mathbf{I} + \mathbf{h}\mathbf{Q}_{\underline{X}}\mathbf{h}^\dagger)^{-1} \underline{y}. \end{aligned}$$

Appendix G. Proof of Lemma 4

We mimic the steps of Appendix E. Consider the SVDs

$$\mathbf{Q}_{\underline{X}} = \mathbf{V}_{\underline{X}} \mathbf{\Sigma}_{\underline{X}} \mathbf{V}_{\underline{X}}^\dagger, \quad \mathbf{Q}_{\underline{X}(s_T)} = \mathbf{V}_{\underline{X}(s_T)} \mathbf{\Sigma}_{\underline{X}(s_T)} \mathbf{V}_{\underline{X}(s_T)}^\dagger.$$

Let $\underline{U} \sim \mathcal{CN}(0, \mathbf{I})$ and write

$$\underline{X} = \mathbf{Q}_{\underline{X}}^{1/2} \underline{U}.$$

Since the $\underline{U}(s_T)$ are CSCG, we have

$$\underline{U}(s'_T) = \mathbf{R}(s'_T, s_T) \underline{U}(s_T) + \underline{Z}(s'_T) \tag{A79}$$

where $\mathbf{R}(s'_T, s_T) = E[\underline{U}(s'_T) \underline{U}(s_T)^\dagger]$ and

$$\underline{Z}(s'_T) \sim \mathcal{CN}(0, \mathbf{I} - \mathbf{R}(s'_T, s_T)\mathbf{R}(s'_T, s_T)^\dagger) \tag{A80}$$

is independent of $\underline{U}(s_T)$. As in (109), define

$$\begin{aligned} \underline{X} &= \sum_{s'_T} \mathbf{W}(s'_T) \underline{X}(s'_T) \\ &= \sum_{s'_T} \mathbf{W}(s'_T) \mathbf{Q}_{\underline{X}(s'_T)}^{1/2} [\mathbf{R}(s'_T, s_T) \underline{U}(s_T) + \underline{Z}(s'_T)] \\ &= \mathbf{Q}_{\underline{X}}^{1/2} \bar{\mathbf{R}}(s_T) \underline{U}(s_T) + \sum_{s'_T} \mathbf{W}(s'_T) \mathbf{Q}_{\underline{X}(s'_T)}^{1/2} \underline{Z}(s'_T) \end{aligned} \tag{A81}$$

where as in (A57), and assuming $\mathbf{Q}_{\underline{X}} \succ \mathbf{0}$, we write

$$\bar{\mathbf{R}}(s_T) = \mathbb{E} \left[\underline{\bar{U}} \underline{U}(s_T)^\dagger \right] = \sum_{s'_T} \mathbf{Q}_{\underline{X}}^{-1/2} \mathbf{W}(s'_T) \mathbf{Q}_{\underline{X}(s'_T)}^{1/2} \mathbf{R}(s'_T, s_T). \tag{A82}$$

Observe that the vector $\mathbf{Q}_{\underline{X}}^{1/2} \bar{\mathbf{R}}(s_T) \underline{U}(s_T)$ is the LMMSE estimate of \underline{X} given $\underline{U}(s_T)$.

Using Lemma 2, we have (see (A58))

$$\tilde{\mathbf{H}} = \mathbb{E} \left[\underline{Y} \underline{X}^\dagger \right] \mathbf{Q}_{\underline{X}}^{-1}, \quad \mathbf{Q}_{\underline{Z}} = \mathbf{Q}_{\underline{Y}} - \tilde{\mathbf{H}} \mathbf{Q}_{\underline{X}} \tilde{\mathbf{H}}^\dagger \tag{A83}$$

and we have the GMI (124) that we repeat here:

$$I_1(A; \underline{Y}) = \log \left(\frac{\det \mathbf{Q}_{\underline{Y}}}{\det (\mathbf{Q}_{\underline{Y}} - \tilde{\mathbf{H}} \mathbf{Q}_{\underline{X}} \tilde{\mathbf{H}}^\dagger)} \right). \tag{A84}$$

As in Appendix E, if the $\mathbf{Q}_{\underline{X}(s_T)}$ are fixed, then so is $\mathbf{Q}_{\underline{Y}}$ because $\underline{U}(s_T) \sim \mathcal{CN}(\underline{0}, \mathbf{I})$ is independent of \underline{Z} given $S_T = s_T$. We want to maximize the GMI (A84). Similar to (A60), we have the decomposition

$$\tilde{\mathbf{H}} \mathbf{Q}_{\underline{X}} \tilde{\mathbf{H}}^\dagger = \tilde{\mathbf{D}} \tilde{\mathbf{D}}^\dagger \tag{A85}$$

where

$$\tilde{\mathbf{D}} = \sum_{s_T} P_{S_T}(s_T) \mathbb{E} \left[\underline{Y} \underline{U}(s_T)^\dagger \mid S_T = s_T \right] \bar{\mathbf{R}}(s_T)^\dagger. \tag{A86}$$

As in (A60), we have the Markov chain $A - [\underline{U}(S_T), S_T] - \underline{Y}$ which implies that \underline{Y} and the $\underline{Z}(s'_T)$ in (A81) are independent given $S_T = s_T$. It is natural to expect that the matrix $\bar{\mathbf{R}}(s_T)$ of correlation coefficients should be “maximized” somehow. Indeed, the Cauchy-Schwarz inequality gives

$$v_1^\dagger \bar{\mathbf{R}}(s_T) v_2 = \mathbb{E} \left[v_1^\dagger \underline{\bar{U}} \cdot \underline{U}(s_T)^\dagger v_2 \right] \leq \sqrt{\mathbb{E} \left[\left| \underline{\bar{U}}^\dagger v_1 \right|^2 \right]} \cdot \sqrt{\mathbb{E} \left[\left| \underline{U}(s_T)^\dagger v_2 \right|^2 \right]} = \|v_1\| \cdot \|v_2\|$$

for any complex M -dimensional vectors v_1 and v_2 . The singular values of $\bar{\mathbf{R}}(s_T)$ are thus at most 1. We will choose the $\underline{U}(s_T)$ so that the $\bar{\mathbf{R}}(s_T)$ are unitary matrices, and thus all singular values are 1.

Consider the SVD decompositions (126) and a codebook based on scaling and rotating a common $\underline{U} \sim \mathcal{CN}(\underline{0}, \mathbf{I})$ of dimension N (see (122)):

$$\underline{U}(s_T) = \mathbf{V}_T(s_T) \underline{U}. \tag{A87}$$

The receiver chooses $M \times M$ unitary matrices $\mathbf{V}_R(s_T)$ for all s_T and uses the weighting matrix (cf. (A62))

$$\mathbf{W}(\tilde{s}_T) = \mathbf{Q}_{\underline{X}}^{1/2} \mathbf{V}_R(\tilde{s}_T) \mathbf{V}_T(\tilde{s}_T)^\dagger \mathbf{Q}_{\underline{X}(\tilde{s}_T)}^{-1/2} \tag{A88}$$

for one $\tilde{s}_T \in \mathcal{S}_T$ with $\mathbf{Q}_{\underline{X}(\tilde{s}_T)} \succ \mathbf{0}$, and $\mathbf{W}(\tilde{s}_T) = \mathbf{0}$ otherwise. These choices give $\underline{X} = \mathbf{Q}_{\underline{X}}^{1/2} \underline{U}$ and (cf. (A63))

$$\mathbf{R}(s'_T, s_T) = \mathbf{V}_T(s'_T) \mathbf{V}_T(s_T)^\dagger, \quad \bar{\mathbf{R}}(s_T) = \mathbf{V}_R(s_T) \mathbf{V}_T(s_T)^\dagger. \tag{A89}$$

Using (126), (A86), and (A89), we have

$$\tilde{\mathbf{D}} = \sum_{s_T} P_{s_T}(s_T) \mathbf{U}_T(s_T) \boldsymbol{\Sigma}(s_T) \mathbf{V}_R(s_T)^\dagger. \quad (\text{A90})$$

References

- Ozarow, L.; Shamai, S.; Wyner, A.D. Information theoretic consideration for cellular mobile radio. *IEEE Trans. Inf. Theory* **1994**, *43*, 359–378. [\[CrossRef\]](#)
- Biglieri, E.; Proakis, J.; Shamai (Shitz), S. Fading channels: Information-theoretic and communications aspects. *IEEE Trans. Inf. Theory* **1998**, *44*, 2619–2692. [\[CrossRef\]](#)
- Love, D.J.; Heath, R.W., Jr.; Lau, V.K.N.; Gesbert, D.; Rao, B.D.; Andrews, M. An overview of limited feedback in wireless communication systems. *IEEE J. Select. Areas Commun.* **2008**, *26*, 1341–1365. [\[CrossRef\]](#)
- Kim, Y.H.; Kramer, G. Information theory for cellular wireless networks. In *Information Theoretic Perspectives on 5G Systems and Beyond*; Cambridge University Press: Cambridge, UK, 2022; pp. 10–92. [\[CrossRef\]](#)
- Keshet, G.; Steinberg, Y.; Merhav, N. Channel coding in the presence of side information. *Found. Trends Commun. Inf. Theory* **2008**, *4*, 445–586. [\[CrossRef\]](#)
- Shannon, C.E. Channels with side information at the transmitter. *IBM J. Res. Develop.* **1958**, *2*, 289–293; Reprinted in *Claude Elwood Shannon: Collected Papers*; Sloane, N.J.A., Wyner, A.D., Eds.; IEEE Press: Piscataway, NJ, USA, 1993; pp. 273–278. [\[CrossRef\]](#)
- Shannon, C.E. Two-way communication channels. In *Proceedings of the Proc. 4th Berkeley Symp. on Mathematical Statistics and Probability*; Neyman, J., Ed.; Univ. Calif. Press: Berkeley, CA, USA, 1961; Volume 1, pp. 611–644; Reprinted in *Claude Elwood Shannon: Collected Papers*; Sloane, N.J.A., Wyner, A.D., Eds.; IEEE Press: Piscataway, NJ, USA, 1993; pp. 351–384.
- Blahut, R. *Principles and Practice of Information Theory*; Addison-Wesley: Reading, MA, USA, 1987.
- Kramer, G. *Directed Information for Channels with Feedback*; Vol. ETH Series in Information Processing; Hartung-Gorre Verlag: Konstanz, Germany, 1998; Volume 11. [\[CrossRef\]](#)
- Caire, G.; Shamai (Shitz), S. On the capacity of some channels with channel state information. *IEEE Trans. Inf. Theory* **1999**, *45*, 2007–2019. [\[CrossRef\]](#)
- McEliece, R.J.; Stark, W.E. Channels with block interference. *IEEE Trans. Inf. Theory* **1984**, *30*, 44–53. [\[CrossRef\]](#)
- Stark, W.; McEliece, R. On the capacity of channels with block memory. *IEEE Trans. Inf. Theory* **1988**, *34*, 322–324. [\[CrossRef\]](#)
- Wang, H.S.; Moayeri, N. Finite-state Markov channel—a useful model for radio communication channels. *IEEE Trans. Vehic. Technol.* **1995**, *44*, 163–171. [\[CrossRef\]](#)
- Wang, H.S.; Chang, P.C. On verifying the first-order Markovian assumption for a Rayleigh fading channel model. *IEEE Trans. Vehic. Technol.* **1996**, *45*, 353–357. [\[CrossRef\]](#)
- Viswanathan, H. Capacity of Markov channels with receiver CSI and delayed feedback. *IEEE Trans. Inf. Theory* **1999**, *45*, 761–771. [\[CrossRef\]](#)
- Zhang, Q.; Kassam, S. Finite-state Markov model for Rayleigh fading channels. *IEEE Trans. Commun.* **1999**, *47*, 1688–1692. [\[CrossRef\]](#)
- Tan, C.C.; Beaulieu, N.C. On first-order Markov modeling for the Rayleigh fading channel. *IEEE Trans. Commun.* **2000**, *48*, 2032–2040. [\[CrossRef\]](#)
- Médard, M. The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel. *IEEE Trans. Inf. Theory* **2000**, *46*, 933–946. [\[CrossRef\]](#)
- Riediger, M.; Shwedyk, E. Communication receivers based on Markov models of the fading channel. In *Proceedings of the IEEE CCECE2002, Canadian Conference on Electrical and Computer Engineering, Conference Proceedings (Cat. No.02CH37373)*, Winnipeg, MB, Canada, 12–15 May 2002; Volume 3, pp. 1255–1260. [\[CrossRef\]](#)
- Agarwal, M.; Honig, M.L.; Ata, B. Adaptive training for correlated fading channels With feedback. *IEEE Trans. Inf. Theory* **2012**, *58*, 5398–5417. [\[CrossRef\]](#)
- Ezzine, R.; Wiese, M.; Deppe, C.; Boche, H. A rigorous proof of the capacity of MIMO Gauss-Markov Rayleigh fading channels. In *Proceedings of the 2022 IEEE International Symposium on Information Theory (ISIT)*, Espoo, Finland, 26 June–1 July 2022; pp. 2732–2737. [\[CrossRef\]](#)
- Kramer, G. Information networks with in-block memory. *IEEE Trans. Inf. Theory* **2014**, *60*, 2105–2120. [\[CrossRef\]](#)
- Pinsker, M.S. Calculation of the rate of information production by means of stationary random processes and the capacity of stationary channel. *Dokl. Akad. Nauk USSR* **1956**, *111*, 753–756.
- Ihara, S. On the capacity of channels with additive non-Gaussian noise. *Inf. Control* **1978**, *37*, 34–39. [\[CrossRef\]](#)
- Pinsker, M.; Prelov, V.; Verdú, S. Sensitivity of channel capacity. *IEEE Trans. Inf. Theory* **1995**, *41*, 1877–1888. [\[CrossRef\]](#)
- Shamai, S. On the capacity of a twisted-wire pair: Peak-power constraint. *IEEE Trans. Commun.* **1990**, *38*, 368–378. [\[CrossRef\]](#)
- Kalet, I.; Shamai, S. On the capacity of a twisted-wire pair: Gaussian model. *IEEE Trans. Commun.* **1990**, *38*, 379–383. [\[CrossRef\]](#)
- Diggavi, S.; Cover, T. The worst additive noise under a covariance constraint. *IEEE Trans. Inf. Theory* **2001**, *47*, 3072–3081. [\[CrossRef\]](#)

29. Klein, T.; Gallager, R. Power control for the additive white Gaussian noise channel under channel estimation errors. In Proceedings of the 2001 IEEE International Symposium on Information Theory (IEEE Cat. No.01CH37252), Washington, DC, USA, 29–29 June 2001; p. 304. [\[CrossRef\]](#)
30. Bhashyam, S.; Sabharwal, A.; Aazhang, B. Feedback gain in multiple antenna systems. *IEEE Trans. Commun.* **2002**, *50*, 785–798. [\[CrossRef\]](#)
31. Hassibi, B.; Hochwald, B. How much training is needed in multiple-antenna wireless links? *IEEE Trans. Inf. Theory* **2003**, *49*, 951–963. [\[CrossRef\]](#)
32. Yoo, T.; Goldsmith, A. Capacity and power allocation for fading MIMO channels with channel estimation error. *IEEE Trans. Inf. Theory* **2006**, *52*, 2203–2214. [\[CrossRef\]](#)
33. Agarwal, M.; Honig, M.L. Wideband fading channel capacity With training and partial feedback. *IEEE Trans. Inf. Theory* **2010**, *56*, 4865–4873. [\[CrossRef\]](#)
34. Soysal, A.; Ulukus, S. Joint channel estimation and resource allocation for MIMO systems-part I: Single-user analysis. *IEEE Trans. Wireless Commun.* **2010**, *9*, 624–631. [\[CrossRef\]](#)
35. Marzetta, T.L.; Larsson, E.G.; Yang, H.; Ngo, H.Q. *Fundamentals of Massive MIMO*; Cambridge University Press: Cambridge, UK, 2016. [\[CrossRef\]](#)
36. Li, Y.; Tao, C.; Lee Swindlehurst, A.; Mezghani, A.; Liu, L. Downlink achievable rate analysis in massive MIMO systems with one-bit DACs. *IEEE Commun. Lett.* **2017**, *21*, 1669–1672. [\[CrossRef\]](#)
37. Caire, G. On the ergodic rate lower bounds With applications to massive MIMO. *IEEE Trans. Wireless Commun.* **2018**, *17*, 3258–3268. [\[CrossRef\]](#)
38. Noam, Y.; Zaidel, B.M. On the two-user MISO interference channel with single-user decoding: Impact of imperfect CSIT and channel dimension reduction. *IEEE Trans. Signal Proc.* **2019**, *67*, 2608–2623. [\[CrossRef\]](#)
39. Kaplan, G.; Shamai (Shitz), S. Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment. *Arch. für Elektron. Und Übertragungstechnik* **1993**, *47*, 228–239.
40. Merhav, N.; Kaplan, G.; Lapidot, A.; Shamai Shitz, S. On information rates for mismatched decoders. *IEEE Trans. Inf. Theory* **1994**, *40*, 1953–1967. [\[CrossRef\]](#)
41. Scarlett, J.; i Fàbregas, A.G.; Somekh-Baruch, A.; Martinez, A. Information-Theoretic Foundations of Mismatched Decoding. *Found. Trends Commun. Inf. Theory* **2020**, *17*, 149–401. [\[CrossRef\]](#)
42. Lapidot, A. Nearest neighbor decoding for additive non-Gaussian noise channels. *IEEE Trans. Inf. Theory* **1996**, *42*, 1520–1529. [\[CrossRef\]](#)
43. Lapidot, A.; Shamai, S. Fading channels: How perfect need “perfect side information” be? *IEEE Trans. Inf. Theory* **2002**, *48*, 1118–1134. [\[CrossRef\]](#)
44. Weingarten, H.; Steinberg, Y.; Shamai, S. Gaussian codes and weighted nearest neighbor decoding in fading multiple-antenna channels. *IEEE Trans. Inf. Theory* **2004**, *50*, 1665–1686. [\[CrossRef\]](#)
45. Asyhari, A.T.; Fàbregas, A.G.i. MIMO block-fading channels With mismatched CSI. *IEEE Trans. Inf. Theory* **2014**, *60*, 7166–7185. [\[CrossRef\]](#)
46. Östman, J.; Lancho, A.; Durisi, G.; Sanguinetti, L. URLLC With massive MIMO: Analysis and design at finite blocklength. *IEEE Trans. Wireless Commun.* **2021**, *20*, 6387–6401. [\[CrossRef\]](#)
47. Zhang, W. A general framework for transmission with transceiver distortion and some applications. *IEEE Trans. Commun.* **2012**, *60*, 384–399. [\[CrossRef\]](#)
48. Zhang, W.; Wang, Y.; Shen, C.; Liang, N. A regression approach to certain information transmission problems. *IEEE J. Selected Areas Commun.* **2019**, *37*, 2517–2531. [\[CrossRef\]](#)
49. Pang, S.; Zhang, W. Generalized nearest neighbor decoding for MIMO channels with imperfect channel state information. In Proceedings of the IEEE Inf. Theory Workshop, Kanazawa, Japan, 17–21 October 2021; pp. 1–6. [\[CrossRef\]](#)
50. Wang, Y.; Zhang, W. Generalized nearest neighbor decoding. *IEEE Trans. Inf. Theory* **2022**, *68*, 5852–5865. [\[CrossRef\]](#)
51. Nedelcu, A.S.; Steiner, F.; Kramer, G. Low-resolution precoding for multi-antenna downlink channels and OFDM. *Entropy* **2022**, *24*, 504. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Essiambre, R.J.; Kramer, G.; Winzer, P.J.; Foschini, G.J.; Goebel, B. Capacity Limits of Optical Fiber Networks. *IEEE/OSA J. Lightw. Technol.* **2010**, *28*, 662–701. [\[CrossRef\]](#)
53. Dar, R.; Shtaif, M.; Feder, M. New bounds on the capacity of the nonlinear fiber-optic channel. *Opt. Lett.* **2014**, *39*, 398–401. [\[CrossRef\]](#)
54. Secondini, M.; Agrell, E.; Forestieri, E.; Marsella, D.; Camara, M.R. Nonlinearity mitigation in WDM systems: Models, strategies, and achievable rates. *IEEE/OSA J. Lightw. Technol.* **2019**, *37*, 2270–2283. [\[CrossRef\]](#)
55. García-Gómez, F.J.; Kramer, G. Mismatched models to lower bound the capacity of optical fiber channels. *IEEE/OSA J. Lightw. Technol.* **2020**, *38*, 6779–6787. [\[CrossRef\]](#)
56. García-Gómez, F.J.; Kramer, G. Mismatched models to lower bound the capacity of dual-polarization optical fiber channels. *IEEE/OSA J. Lightw. Technol.* **2021**, *39*, 3390–3399. [\[CrossRef\]](#)
57. García-Gómez, F.J.; Kramer, G. Rate and power scaling of space-division multiplexing via nonlinear perturbation. *J. Lightw. Technol.* **2022**, *40*, 5077–5082. [\[CrossRef\]](#)

58. Secondini, M.; Civelli, S.; Forestieri, E.; Khan, L.Z. New lower bounds on the capacity of optical fiber channels via optimized shaping and detection. *J. Lightw. Technol.* **2022**, *40*, 3197–3209. [[CrossRef](#)]
59. Shtaif, M.; Antonelli, C.; Mecozzi, A.; Chen, X. Challenges in estimating the information capacity of the fiber-optic channel. *Proc. IEEE* **2022**, *110*, 1655–1678. [[CrossRef](#)]
60. Mecozzi, A.; Shtaif, M. Information Capacity of Direct Detection Optical Transmission Systems. *IEEE/OSA Trans. Lightw. Technol.* **2018**, *36*, 689–694. [[CrossRef](#)]
61. Plabst, D.; Prinz, T.; Wiegart, T.; Rahman, T.; Stojanović, N.; Calabrò, S.; Hanik, N.; Kramer, G. Achievable rates for short-reach fiber-optic channels with direct detection. *IEEE/OSA J. Lightw. Technol.* **2022**, *40*, 3602–3613. [[CrossRef](#)]
62. Gallager, R.G. *Information Theory and Reliable Communication*; Wiley: New York, NY, USA, 1968.
63. Divsalar, D. Performance of Mismatched Receivers on Bandlimited Channels. Ph.D. Thesis, Univ. California, Los Angeles, CA, USA, 1978.
64. Ozarow, L.; Wyner, A. On the capacity of the Gaussian channel with a finite number of input levels. *IEEE Trans. Inf. Theory* **1990**, *36*, 1426–1428. [[CrossRef](#)]
65. Chagnon, M. Optical Communications for Short Reach. *IEEE/OSA Trans. Lightw. Technol.* **2019**, *37*, 1779–1797. [[CrossRef](#)]
66. Arnold, D.; Loeliger, H.A.; Vontobel, P.; Kavcic, A.; Zeng, W. Simulation-based computation of information rates for channels with memory. *IEEE Trans. Inf. Theory* **2006**, *52*, 3498–3508. [[CrossRef](#)]
67. Aboy-Faycal, I.; Lapidoth, A. On the capacity of reduced-complexity receivers for intersymbol interference channels. In Proceedings of the Convention of the Electrical and Electronic Engineers in Israel, Tel-Aviv, Israel, 11–12 April 2000; pp. 263–266. [[CrossRef](#)]
68. Rusek, F.; Prlja, A. Optimal channel shortening for MIMO and ISI channels. *IEEE Trans. Wireless Commun.* **2012**, *11*, 810–818. [[CrossRef](#)]
69. Hu, S.; Rusek, F. On the design of channel shortening demodulators for iterative receivers in linear vector channels. *IEEE Access* **2018**, *6*, 48339–48359. [[CrossRef](#)]
70. Mezghani, A.; Nossek, J.A. Analysis of 1-bit output noncoherent fading channels in the low SNR regime. In Proceedings of the IEEE International Symposium on Information Theory, Seoul, Republic of Korea, 28 June–3 July 2009; pp. 1080–1084. [[CrossRef](#)]
71. Papoulis, A. *Probability, Random Variables, and Stochastic Processes*, 2nd ed.; McGraw-Hill: New York, NY, USA, 1984.
72. Kramer, G. Capacity results for the discrete memoryless network. *IEEE Trans. Inf. Theory* **2003**, *49*, 4–21. [[CrossRef](#)]
73. Verdú, S. Spectral efficiency in the wideband regime. *IEEE Trans. Inf. Theory* **2002**, *48*, 1319–1343. [[CrossRef](#)]
74. Kramer, G.; Ashikhmin, A.; van Wijngaarden, A.; Wei, X. Spectral efficiency of coded phase-shift keying for fiber-optic communication. *IEEE/OSA J. Lightw. Technol.* **2003**, *21*, 2438–2445. [[CrossRef](#)]
75. Hui, J.Y.N. Fundamental Issues of Multiple Accessing. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 1983.
76. Scarlett, J.; Martinez, A.; Fabregas, A.G.i. Mismatched decoding: Error exponents, second-order rates and saddlepoint approximations. *IEEE Trans. Inf. Theory* **2014**, *60*, 2647–2666. [[CrossRef](#)]
77. Asadi Kangarshahi, E.; Guillén i Fabregas, A. A single-letter upper bound to the mismatch capacity. *IEEE Trans. Inf. Theory* **2021**, *67*, 2013–2033. [[CrossRef](#)]
78. Lau, V.K.N.; Liu, Y.; Chen, T.A. Capacity of memoryless channels and block-fading channels with designable cardinality-constrained channel state feedback. *IEEE Trans. Inf. Theory* **2004**, *50*, 2038–2049. [[CrossRef](#)]
79. Shannon, C.E. Geometrische Deutung einiger Ergebnisse bei der Berechnung der Kanalkapazität. *Nachrichtentechnische Z.* **1957**, *10*, 1–4. English version in *Claude Elwood Shannon: Collected Papers*; Sloane, N.J.A., Wyner, A.D., Eds.; IEEE Press: Piscataway, NJ, USA, 1993; pp. 259–264.
80. Farmanbar, H.; Khandani, A.K. Precoding for the AWGN channel with discrete interference. *IEEE Trans. Inf. Theory* **2009**, *55*, 4019–4032. [[CrossRef](#)]
81. Wachsmann, U.; Fischer, R.; Huber, J. Multilevel codes: Theoretical concepts and practical design rules. *IEEE Trans. Inf. Theory* **1999**, *45*, 1361–1391. [[CrossRef](#)]
82. Stolte, N. Rekursive Codes mit der Plotkin-Konstruktion und ihre Decodierung. Ph.D. Thesis, Technische Universität Darmstadt, Darmstadt, Germany, 2002.
83. Arikan, E. Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Trans. Inf. Theory* **2009**, *55*, 3051–3073. [[CrossRef](#)]
84. Seidl, M.; Schenk, A.; Stierstorfer, C.; Huber, J.B. Polar-coded modulation. *IEEE Trans. Commun.* **2013**, *61*, 4108–4119. [[CrossRef](#)]
85. Honda, J.; Yamamoto, H. Polar coding Without alphabet extension for asymmetric models. *IEEE Trans. Inf. Theory* **2013**, *59*, 7829–7838. [[CrossRef](#)]
86. Runge, C.; Wiegart, T.; Lentner, D.; Prinz, T. Multilevel binary polar-coded modulation achieving the capacity of asymmetric channels. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Espoo, Finland, 26 June–1 July 2022; pp. 2595–2600. [[CrossRef](#)]
87. Parthasarathy, K.R. Extreme points of the convex set of joint probability distributions with fixed marginals. *Proc. Math. Sci.* **2007**, *117*, 505–515. [[CrossRef](#)]
88. Nadkarni, M.G.; Navada, K.G. On the number of extreme measures with fixed marginals. *arXiv* **2008**, arXiv:math/0806.1214.

89. Farmanbar, H.; Gharan, S.O.; Khandani, A.K. Channel code design with causal side information at the encoder. *Eur. Trans. Telecommun.* **2010**, *21*, 337–351. [[CrossRef](#)]
90. Birkhoff, G. Three observations on linear algebra. *Univ. Nac. Tucumán. Revista A* **1946**, *5*, 147–151.
91. Wolfowitz, J. *Coding Theorems of Information Theory*, 2nd ed.; Springer: Berlin, Germany, 1964.
92. Kim, T.T.; Skoglund, M. On the Expected Rate of Slowly Fading Channels With Quantized Side Information. *IEEE Trans. Commun.* **2007**, *55*, 820–829. [[CrossRef](#)]
93. Rosenzweig, A.; Steinberg, Y.; Shamai, S. On channels with partial channel state information at the transmitter. *IEEE Trans. Inf. Theory* **2005**, *51*, 1817–1830. [[CrossRef](#)]
94. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. 623–656. Reprinted in *Claude Elwood Shannon: Collected Papers*; Sloane, N.J.A., Wyner, A.D., Eds.; IEEE Press: Piscataway, NJ, USA, 1993; pp. 5–83. [[CrossRef](#)]
95. Goldsmith, A.J.; Varaiya, P.P. Capacity of fading channels with channel side information. *IEEE Trans. Inf. Theory* **1997**, *43*, 1986–1992. [[CrossRef](#)]
96. Abou-Faycal, I.; Trott, M.; Shamai, S. The capacity of discrete-time memoryless Rayleigh-fading channels. *IEEE Trans. Inf. Theory* **2001**, *47*, 1290–1301. [[CrossRef](#)]
97. Lapidoth, A.; Moser, S.M. Capacity bounds via duality with applications to multiple-antenna systems on flat fading channels. *IEEE Trans. Inf. Theory* **2003**, *49*, 2426–2467. [[CrossRef](#)]
98. Taricco, G.; Elia, M. Capacity of fading channel with no side information. *Elec. Lett.* **1997**, *33*, 1368–1370. [[CrossRef](#)]
99. Marzetta, T.L.; Hochwald, B.M. Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading. *IEEE Trans. Inf. Theory* **1999**, *45*, 139–157. [[CrossRef](#)]
100. Zheng, L.; Tse, D. Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel. *IEEE Trans. Inf. Theory* **2002**, *48*, 359–383. [[CrossRef](#)]
101. Gursoy, M.; Poor, H.; Verdu, S. The noncoherent Rician fading channel—part I: Structure of the capacity-achieving input. *IEEE Trans. Wireless Commun.* **2005**, *4*, 2193–2206. [[CrossRef](#)]
102. Chowdhury, M.; Goldsmith, A. Capacity of block Rayleigh fading channels without CSI. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Barcelona, Spain, 10–15 July 2016; pp. 1884–1888. [[CrossRef](#)]
103. Goldsmith, A.J.; Médard, M. Capacity of time-varying channels with causal channel side information. *IEEE Trans. Inf. Theory* **2007**, *53*, 881–899. [[CrossRef](#)]
104. Jelinek, F. Indecomposable channels with side information at the transmitter. *Inf. Control* **1965**, *8*, 36–55. [[CrossRef](#)]
105. Das, A.; Narayan, P. Capacities of time-varying multiple-access channels with side information. *IEEE Trans. Inf. Theory* **2002**, *48*, 4–25. [[CrossRef](#)]
106. Van Veen, B.; Buckley, K. Beamforming: A versatile approach to spatial filtering. *IEEE ASSP Mag.* **1988**, *5*, 4–24. [[CrossRef](#)]
107. Liaskos, C.; Nie, S.; Tsioliariidou, A.; Pitsillides, A.; Ioannidis, S.; Akyildiz, I. A new wireless communication paradigm through software-controlled metasurfaces. *IEEE Commun. Mag.* **2018**, *56*, 162–169. [[CrossRef](#)]
108. Renzo, M.; Debbah, M.; Phan-Huy, D.T.; Zappone, A.; Alouini, M.S.; Yuen, C.; Sciancalepore, V.; Alexandropoulos, G.C.; Hoydis, J.; Gacanin, H.; et al. Smart radio environments empowered by reconfigurable AI meta-surfaces: An idea whose time has come. *J. Wirel. Com. Netw.* **2019**, *2019*, 129. [[CrossRef](#)]
109. Thangaraj, A.; Kramer, G.; Böcherer, G. Capacity bounds for discrete-time, amplitude-constrained, additive white Gaussian noise channels. *IEEE Trans. Inf. Theory* **2017**, *63*, 4172–4182. [[CrossRef](#)]
110. Csiszár, I.; Körner, J. *Information Theory: Coding Theorems for Discrete Memoryless Channels*; Akadémiai Kiadó: Budapest, Hungary, 1981.
111. Cody, W.J.; Thacher, H.C., Jr. Rational Chebyshev approximations for the exponential integral $E_1(x)$. *Math. Comp.* **1968**, *22*, 641–649.
112. Nantomah, K. On Some Bounds for the Exponential Integral Function. *J. Nepal Mathem. Soc.* **2021**, *4*, 28–34. [[CrossRef](#)]
113. Sofotasios, P.C.; Muhaidat, S.; Karagiannidis, G.K.; Sharif, B.S. Solutions to integrals involving the Marcum Q-function and applications. *IEEE Signal Proc. Lett.* **2015**, *22*, 1752–1756. [[CrossRef](#)]
114. Neeser, F.; Massey, J. Proper complex random processes with applications to information theory. *IEEE Trans. Inf. Theory* **1993**, *39*, 1293–1302. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.