Capacity of Multi-antenna Gaussian Channels

İ. Emre Telatar*

Abstract

We investigate the use of multiple transmitting and/or receiving antennas for single user communications over the additive Gaussian channel with and without fading. We derive formulas for the capacities and error exponents of such channels, and describe computational procedures to evaluate such formulas. We show that the potential gains of such multi-antenna systems over single-antenna systems is rather large under independence assumptions for the fades and noises at different receiving antennas.

1 Introduction

We will consider a single user Gaussian channel with multiple transmitting and/or receiving antennas. We will denote the number of transmitting antennas by t and the number of receiving antennas by r. We will exclusively deal with a linear model in which the received vector $\boldsymbol{y} \in \mathbb{C}^r$ depends on the transmitted vector $\boldsymbol{x} \in \mathbb{C}^t$ via

$$\boldsymbol{y} = H\boldsymbol{x} + \boldsymbol{n} \tag{1}$$

where H is a $r \times t$ complex matrix and $\boldsymbol{n}$ is zero-mean complex Gaussian noise with independent, equal variance real and imaginary parts. We assume $\mathcal{E}[\boldsymbol{n}\boldsymbol{n}^\dagger] = I_r$, that is, the noises corrupting the different receivers are independent. The transmitter is constrained in its total power to P,

$$\mathcal{E}[\boldsymbol{x}^\dagger\boldsymbol{x}] \le P.$$

Equivalently, since $\boldsymbol{x}^\dagger\boldsymbol{x} = \operatorname{tr}(\boldsymbol{x}\boldsymbol{x}^\dagger)$, and expectation and trace commute,

$$\operatorname{tr}(\mathcal{E}[\boldsymbol{x}\boldsymbol{x}^\dagger]) \le P. \tag{2}$$

*Rm. 2C-174, Lucent Technologies, Bell Laboratories, 600 Mountain Avenue, Murray Hill, NJ, USA 07974, `telatar@lucent.com`

This second form of the power constraint will prove more useful in the upcoming discussion.

We will consider several scenarios for the matrix H:

1. H is deterministic.

2. H is a random matrix (for which we shall use the notation $\boldsymbol{H}$), chosen according to a probability distribution, and each use of the channel corresponds to an independent realization of $\boldsymbol{H}$.

3. H is a random matrix, but is fixed once it is chosen.

The main focus of this paper in on the last two of these cases. The first case is included so as to expose the techniques used in the later cases in a more familiar context. In the cases when $\boldsymbol{H}$ is random, we will assume that its entries form an i.i.d. Gaussian collection with zero-mean, independent real and imaginary parts, each with variance 1/2. Equivalently, each entry of $\boldsymbol{H}$ has uniform phase and Rayleigh magnitude. This choice models a *Rayleigh fading environment* with enough separation *within* the receiving antennas and the transmitting antennas such that the fades for each transmitting-receiving antenna pair are independent. In all cases, we will assume that the *realization* of $\boldsymbol{H}$ is known to the receiver, or, equivalently, the channel output consists of the pair $(\boldsymbol{y}, \boldsymbol{H})$, and the *distribution* of $\boldsymbol{H}$ is known at the transmitter.

2 Preliminaries

A complex random vector $\boldsymbol{x} \in \mathbb{C}^n$ is said to be Gaussian if the real random vector $\hat{\boldsymbol{x}} \in \mathbb{R}^{2n}$ consisting of its real and imaginary parts, $\hat{\boldsymbol{x}} = \begin{bmatrix} \mathfrak{Re}(\boldsymbol{x}) \\ \mathfrak{Im}(\boldsymbol{x}) \end{bmatrix}$, is Gaussian. Thus, to specify the distribution of a complex Gaussian random vector $\boldsymbol{x}$, it is necessary to specify the expectation and covariance of $\hat{\boldsymbol{x}}$, namely,

$$\mathcal{E}[\hat{\boldsymbol{x}}] \in \mathbb{R}^{2n} \quad \text{and} \quad \mathcal{E}\left[(\hat{\boldsymbol{x}} - \mathcal{E}[\hat{\boldsymbol{x}}])(\hat{\boldsymbol{x}} - \mathcal{E}[\hat{\boldsymbol{x}}])^\dagger\right] \in \mathbb{R}^{2n \times 2n}.$$

We will say that a complex Gaussian random vector $\boldsymbol{x}$ is *circularly symmetric* if the covariance of the corresponding $\hat{\boldsymbol{x}}$ has the structure

$$\mathcal{E}\left[(\hat{\boldsymbol{x}} - \mathcal{E}[\hat{\boldsymbol{x}}])(\hat{\boldsymbol{x}} - \mathcal{E}[\hat{\boldsymbol{x}}])^\dagger\right] = \frac{1}{2} \begin{bmatrix} \mathfrak{Re}(Q) & -\mathfrak{Im}(Q) \\ \mathfrak{Im}(Q) & \mathfrak{Re}(Q) \end{bmatrix} \tag{3}$$

for some Hermitian non-negative definite $Q \in \mathbb{C}^{n \times n}$. Note that the real part of an Hermitian matrix is symmetric and the imaginary part of an Hermitian matrix is anti-symmetric and thus the matrix appearing in (3) is real and symmetric. In this case $\mathcal{E}\left[(\boldsymbol{x} - \mathcal{E}[\boldsymbol{x}])(\boldsymbol{x} - \mathcal{E}[\boldsymbol{x}])^\dagger\right] = Q$, and thus, a circularly symmetric complex Gaussian random vector $\boldsymbol{x}$ is specified by prescribing $\mathcal{E}[\boldsymbol{x}]$ and $\mathcal{E}\left[(\boldsymbol{x} - \mathcal{E}[\boldsymbol{x}])(\boldsymbol{x} - \mathcal{E}[\boldsymbol{x}])^\dagger\right]$.

For any $z \in \mathbb{C}^n$ and $A \in \mathbb{C}^{n\times m}$ define

$$\hat{z} = \begin{bmatrix} \mathfrak{Re}(z) \\ \mathfrak{Im}(z) \end{bmatrix} \quad \text{and} \quad \hat{A} = \begin{bmatrix} \mathfrak{Re}(A) & -\mathfrak{Im}(A) \\ \mathfrak{Im}(A) & \mathfrak{Re}(A) \end{bmatrix}.$$

LEMMA 1. *The mappings* $z \to \hat{z} = \left[\begin{smallmatrix} \mathfrak{Re}(z) \\ \mathfrak{Im}(z) \end{smallmatrix}\right]$ *and* $A \to \hat{A} = \left[\begin{smallmatrix} \mathfrak{Re}(A) & -\mathfrak{Im}(A) \\ \mathfrak{Im}(A) & \mathfrak{Re}(A) \end{smallmatrix}\right]$ *have the following properties:*

$$C = AB \iff \hat{C} = \hat{A}\hat{B} \tag{4a}$$
$$C = A + B \iff \hat{C} = \hat{A} + \hat{B} \tag{4b}$$
$$C = A^\dagger \iff \hat{C} = \hat{A}^\dagger \tag{4c}$$
$$C = A^{-1} \iff \hat{C} = \hat{A}^{-1} \tag{4d}$$
$$\det(\hat{A}) = |\det(A)|^2 = \det(AA^\dagger) \tag{4e}$$
$$z = x + y \iff \hat{z} = \hat{x} + \hat{y} \tag{4f}$$
$$y = Ax \iff \hat{y} = \hat{A}\hat{x} \tag{4g}$$
$$\mathfrak{Re}(x^\dagger y) = \hat{x}^\dagger \hat{y}. \tag{4h}$$

Proof. The properties (4a), (4b) and (4c) are immediate. (4d) follows from (4a) and the fact that $\hat{I}_n = I_{2n}$. (4e) follows from

$$\det(\hat{A}) = \det\left(\begin{bmatrix} I & iI \\ 0 & I \end{bmatrix} \hat{A} \begin{bmatrix} I & -iI \\ 0 & I \end{bmatrix}\right) = \det\left(\begin{bmatrix} A & 0 \\ \mathfrak{Im}(A) & A^* \end{bmatrix}\right) = \det(A)\det(A)^*.$$

(4f), (4g) and (4h) are immediate. □

COROLLARY 1. $U \in \mathbb{C}^{n\times n}$ *is unitary if and only if* $\hat{U} \in \mathbb{R}^{2n\times 2n}$ *is orthonormal.*

Proof. $U^\dagger U = I_n \iff (\hat{U})^\dagger \hat{U} = \hat{I}_n = I_{2n}$. □

COROLLARY 2. *If* $Q \in \mathbb{C}^{n\times n}$ *is non-negative definite then so is* $\hat{Q} \in \mathbb{R}^{2n\times 2n}$.

Proof. Given $x = [x_1, \ldots, x_{2n}]^\dagger \in \mathbb{R}^{2n}$, let $z = [x_1 + jx_{n+1}, \ldots, x_n + jx_{2n}]^\dagger \in \mathbb{C}^n$, so that $x = \hat{z}$. Then by (4g) and (4h)

$$x^\dagger \hat{Q} x = \mathfrak{Re}(z^\dagger Q z) = z^\dagger Q z \geq 0. \qquad \square$$

The probability density (with respect to the standard Lebesgue measure on $\mathbb{C}^n$) of a circularly symmetric complex Gaussian with mean μ and covariance Q is given by

$$\begin{aligned}\gamma_{\mu,Q}(x) &= \det(\pi\hat{Q})^{-1/2}\exp\bigl(-(\hat{x}-\hat{\mu})^\dagger\hat{Q}^{-1}(\hat{x}-\hat{\mu})\bigr)\\ &= \det(\pi Q)^{-1}\exp\bigl(-(x-\mu)^\dagger Q^{-1}(x-\mu)\bigr)\end{aligned}$$

where the second equality follows from (4d)–(4h). The differential entropy of a complex Gaussian $\boldsymbol{x}$ with covariance Q is given by

$$\begin{aligned}\mathcal{H}(\gamma_Q) &= \mathcal{E}_{\gamma_Q}[-\log\gamma_Q(\boldsymbol{x})]\\ &= \log\det(\pi Q) + (\log e)\,\mathcal{E}[\boldsymbol{x}^\dagger Q^{-1}\boldsymbol{x}]\\ &= \log\det(\pi Q) + (\log e)\,\mathrm{tr}(\mathcal{E}[\boldsymbol{x}\boldsymbol{x}^\dagger]Q^{-1})\\ &= \log\det(\pi Q) + (\log e)\,\mathrm{tr}(I)\\ &= \log\det(\pi e Q).\end{aligned}$$

For us, the importance of the circularly symmetric complex Gaussians is due to the following lemma: circularly symmetric complex Gaussians are entropy maximizers.

LEMMA 2. *Suppose the complex random vector $\boldsymbol{x} \in \mathbb{C}^n$ is zero-mean and satisfies $\mathcal{E}[\boldsymbol{x}\boldsymbol{x}^\dagger] = Q$, i.e., $\mathcal{E}[\boldsymbol{x}_i\boldsymbol{x}_j^*] = Q_{ij}$, $1 \le i,j \le n$. Then the entropy of $\boldsymbol{x}$ satisfies $\mathcal{H}(\boldsymbol{x}) \le \log\det(\pi e Q)$ with equality if and only if $\boldsymbol{x}$ is a circularly symmetric complex Gaussian with*

$$\mathcal{E}[\boldsymbol{x}\boldsymbol{x}^\dagger] = Q$$

Proof. Let p be any density function satisfying $\int_{\mathbb{C}^n} p(x)x_i x_j^*\,dx = Q_{ij}$, $1 \le i,j \le n$. Let

$$\gamma_Q(x) = \det(\pi Q)^{-1}\exp\bigl(-x^\dagger Q^{-1}x\bigr).$$

Observe that $\int_{\mathbb{C}^n}\gamma_Q(x)x_i x_j^*\,dx = Q_{ij}$, and that $\log\gamma_Q(x)$ is a linear combination of the terms $x_i x_j^*$. Thus $\mathcal{E}_{\gamma_Q}[\log\gamma_Q(\boldsymbol{x})] = \mathcal{E}_p[\log\gamma_Q(\boldsymbol{x})]$. Then,

$$\begin{aligned}\mathcal{H}(p) - \mathcal{H}(\gamma_Q) &= -\int_{\mathbb{C}^n} p(x)\log p(x)\,dx + \int_{\mathbb{C}^n}\gamma_Q(x)\log\gamma_Q(x)\,dx\\ &= -\int_{\mathbb{C}^n} p(x)\log p(x)\,dx + \int_{\mathbb{C}^n} p(x)\log\gamma_Q(x)\,dx\\ &= \int_{\mathbb{C}^n} p(x)\log\frac{\gamma_Q(x)}{p(x)}\,dx\\ &\le 0,\end{aligned}$$

with equality only if $p = \gamma_Q$. Thus $\mathcal{H}(p) \le \mathcal{H}(\gamma_Q)$. □

LEMMA 3. *If $\boldsymbol{x} \in \mathbb{C}^n$ is a circularly symmetric complex Gaussian then so is $\boldsymbol{y} = A\boldsymbol{x}$ for any $A \in \mathbb{C}^{m\times n}$.*

Proof. We may assume $\boldsymbol{x}$ is zero-mean. Let $Q = \mathcal{E}[\boldsymbol{x}\boldsymbol{x}^\dagger]$. Then $\boldsymbol{y}$ is zero-mean, $\hat{\boldsymbol{y}} = \hat{A}\hat{\boldsymbol{x}}$, and

$$\mathcal{E}[\hat{\boldsymbol{y}}\hat{\boldsymbol{y}}^\dagger] = \hat{A}\,\mathcal{E}[\hat{\boldsymbol{x}}\hat{\boldsymbol{x}}^\dagger]\hat{A}^\dagger = \tfrac{1}{2}\hat{A}\hat{Q}\hat{A}^\dagger = \tfrac{1}{2}\hat{K}$$

where $K = AQA^\dagger$. □

LEMMA 4. *If $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent circularly symmetric complex Gaussians, then $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{y}$ is a circularly symmetric complex Gaussian.*

Proof. Let $A = \mathcal{E}[\boldsymbol{x}\boldsymbol{x}^\dagger]$ and $B = \mathcal{E}[\boldsymbol{y}\boldsymbol{y}^\dagger]$. Then $\mathcal{E}[\hat{\boldsymbol{z}}\hat{\boldsymbol{z}}^\dagger] = \frac{1}{2}\hat{C}$ with $C = A + B$. □

3 The Gaussian channel with fixed transfer function

We will start by reminding ourselves the case of deterministic H. The results of this section can be inferred from [1, Ch. 8]

3.1 Capacity

We will first derive an expression for the capacity $C(H, P)$ of this channel. To that end, we will maximize the average mutual information $\mathcal{I}(\boldsymbol{x}; \boldsymbol{y})$ between the input and the output of the channel over the choice of the distribution of $\boldsymbol{x}$.

By the singular value decomposition theorem, any matrix $H \in \mathbb{C}^{r\times t}$ can be written as

$$H = UDV^\dagger$$

where $U \in \mathbb{C}^{r\times r}$ and $V \in \mathbb{C}^{t\times t}$ are unitary, and $D \in \mathbb{R}^{r\times t}$ is non-negative and diagonal. In fact, the diagonal entries of D are the non-negative square roots of the eigenvalues of $HH^\dagger$, the columns of U are the eigenvectors of $HH^\dagger$ and the columns of V are the eigenvectors of $H^\dagger H$. Thus, we can write (1) as

$$\boldsymbol{y} = UDV^\dagger\boldsymbol{x} + \boldsymbol{n}.$$

Let $\tilde{\boldsymbol{y}} = U^\dagger\boldsymbol{y}$, $\tilde{\boldsymbol{x}} = V^\dagger\boldsymbol{x}$, $\tilde{\boldsymbol{n}} = U^\dagger\boldsymbol{n}$. Note that U and V are invertible, $\tilde{\boldsymbol{n}}$ has the same distribution as $\boldsymbol{n}$ and, $\mathcal{E}[\tilde{\boldsymbol{x}}^\dagger\tilde{\boldsymbol{x}}] = \mathcal{E}[\boldsymbol{x}^\dagger\boldsymbol{x}]$. Thus, the original channel is equivalent to the channel

$$\tilde{\boldsymbol{y}} = D\tilde{\boldsymbol{x}} + \tilde{\boldsymbol{n}} \tag{5}$$

where $\tilde{\boldsymbol{n}}$ is zero-mean, Gaussian, with independent, identically distributed real and imaginary parts and $\mathcal{E}[\tilde{\boldsymbol{n}}\tilde{\boldsymbol{n}}^\dagger] = I_r$. Since H is of rank at most $\min\{r, t\}$, at most $\min\{r, t\}$ of the singular values of it are non-zero. Denoting these by $\lambda_i^{1/2}$, $i = 1, \ldots, \min\{r, t\}$, we can write (5) component-wise, to get

$$\tilde{y}_i = \lambda_i^{1/2}\tilde{x}_i + \tilde{n}_i, \quad 1 \le i \le \min\{r, t\},$$

and the rest of the components of $\tilde{y}$ (if any) are equal to the corresponding components of $\tilde{n}$. We thus see that $\tilde{y}_i$ for $i > \min\{t, r\}$ is independent of the transmitted signal and that $\tilde{x}_i$ for $i > \min\{t, r\}$ don't play any role. To maximize the mutual information, we need to choose $\{\tilde{x}_i : 1 \le i \le \min\{r, t\}\}$ to be independent, with each $\tilde{x}_i$ having independent Gaussian, zero-mean real and imaginary parts. The variances need to be chosen via "water-filling" as

$$\mathcal{E}[\mathfrak{Re}(\tilde{x}_i)^2] = \mathcal{E}[\mathfrak{Im}(\tilde{x}_i)^2] = \tfrac{1}{2}(\mu - \lambda_i^{-1})^+$$

where μ is chosen to meet the power constraint. Here, a^+ denotes $\max\{0, a\}$. The power P and the maximal mutual information can thus be parametrized as

$$P(\mu) = \sum_i \left(\mu - \lambda_i^{-1}\right)^+, \qquad C(\mu) = \sum_i \left(\ln(\mu\lambda_i)\right)^+.$$

Remark 1 (Reciprocity). Since the non-zero eigenvalues of $H^\dagger H$ are the same as those of $HH^\dagger$, we see that the capacities of channels corresponding to H and $H^\dagger$ are the same.

Example 1. Take $H_{ij} = 1$ for all i, j. We can write H as

$$H = \begin{bmatrix} \sqrt{1/r} \\ \vdots \\ \sqrt{1/r} \end{bmatrix} \left(\sqrt{rt}\right) \left[\sqrt{1/t} \ldots \sqrt{1/t}\right]$$

and we thus see that in the singular value decomposition of H the diagonal matrix D will have only one non-zero entry, $\sqrt{rt}$. (We also see that the first column of U is $\sqrt{1/r}[1, \ldots, 1]^\dagger$ and the first column of V is $\sqrt{1/t}[1, \ldots, 1]^\dagger$.) Thus,

$$C = \log(1 + rtP).$$

The $\boldsymbol{x} = V\tilde{\boldsymbol{x}}$ that achieves this capacity satisfies $\mathcal{E}[x_i x_j^*] = P/t$ for all i, j, i.e., the transmitters are all sending the same signal. Note that, even though each transmitter is sending a power of P/t, since their signals add coherently at the receiver, the power received at each receiver is Pt. Since each receiver sees the same signal and the noises at the receivers are uncorrelated the overall signal to noise ratio is Prt.

Example 2. Take $r = t = n$ and $H = I_n$. Then

$$C = n \log(1 + P/n)$$

For $\boldsymbol{x}$ that achieves this capacity $\mathcal{E}[x_i x_j^*] = \delta_{ij} P/n$, i.e, the components of $\boldsymbol{x}$ are i.i.d. However, it is incorrect to infer from this conclusion that to achieve capacity one has to do independent coding for each transmitter. It is true that the capacity of this channel can be achieved by splitting the incoming data stream into t streams, coding and modulating these schemes separately, and then sending the t modulated signals over the different transmitters. But, suppose Nt bits are going to be transmitted, and we will either separate them into t groups of N bits each and use each group to select one of 2^N signals for each transmitter, or, we will use all all Nt bits to select one of 2^{Nt} signal vectors. The second of these alternatives will yield a probability of error much smaller than the first, at the expense of much greater complexity. Indeed, the log of the error probability in the two cases will differ by a factor of t. (See the error exponents of parallel channels in [1, pp. 149–150].)

3.2 Alternative Derivation of the Capacity

The mutual information $\mathcal{I}(\boldsymbol{x}; \boldsymbol{y})$ can be written as

$$\mathcal{I}(\boldsymbol{x}; \boldsymbol{y}) = \mathcal{H}(\boldsymbol{y}) - \mathcal{H}(\boldsymbol{y}|\boldsymbol{x}) = \mathcal{H}(\boldsymbol{y}) - \mathcal{H}(\boldsymbol{n}),$$

and thus maximizing $\mathcal{I}(\boldsymbol{x}; \boldsymbol{y})$ is equivalent to maximizing $\mathcal{H}(\boldsymbol{y})$. Note that if $\boldsymbol{x}$ satisfies $\mathcal{E}[\boldsymbol{x}^\dagger \boldsymbol{x}] \le P$, so does $\boldsymbol{x} - \mathcal{E}[\boldsymbol{x}]$, so we can restrict our attention to zero-mean $\boldsymbol{x}$. Furthermore, if $\boldsymbol{x}$ is zero-mean with covariance $\mathcal{E}[\boldsymbol{x}\boldsymbol{x}^\dagger] = Q$, then $\boldsymbol{y}$ is zero-mean with covariance $\mathcal{E}[\boldsymbol{y}\boldsymbol{y}^\dagger] = HQH^\dagger + I_r$, and by Lemma 2 among such $\boldsymbol{y}$ the entropy is largest when $\boldsymbol{y}$ is circularly symmetric complex Gaussian, which is the case when $\boldsymbol{x}$ is circularly symmetric complex Gaussian (Lemmas 3 and 4). So, we can further restrict our attention to circularly symmetric complex Gaussian $\boldsymbol{x}$. In this case the mutual information is given by

$$\mathcal{I}(\boldsymbol{x}; \boldsymbol{y}) = \log\det(I_r + HQH^\dagger) = \log\det(I_t + QH^\dagger H)$$

where the second equality follows from the determinant identity $\det(I + AB) = \det(I + BA)$, and it only remains to choose Q to maximize this quantity subject to the constraints $\operatorname{tr}(Q) \le P$ and that Q is non-negative definite. The quantity $\log\det(I + HQH^\dagger)$ will occur in this document frequently enough that we will let

$$\Psi(Q, H) = \log\det(I + HQH^\dagger)$$

to denote it. Since $H^\dagger H$ is Hermitian it can be diagonalized, $H^\dagger H = U^\dagger \Lambda U$, with unitary U and non-negative diagonal $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_t)$. Applying the determinant identity again we see that

$$\det(I_r + HQH^\dagger) = \det(I_t + \Lambda^{1/2} U Q U^\dagger \Lambda^{1/2}).$$

Observe that $\tilde{Q} = UQU^\dagger$ is non-negative definite when and only when Q is, and that $\operatorname{tr}(\tilde{Q}) = \operatorname{tr}(Q)$; thus the maximization over Q can be carried equally well over $\tilde{Q}$.

Note also that for any non-negative definite matrix A, $\det(A) \le \prod_i A_{ii}$, thus

$$\det(I_r + \Lambda^{1/2}\tilde{Q}\Lambda^{1/2}) \le \prod_i (1 + \tilde{Q}_{ii}\lambda_i)$$

with equality when $\tilde{Q}$ is diagonal. Thus we see that the maximizing $\tilde{Q}$ is diagonal, and the optimal diagonal entries can be found via "water-filling" to be

$$\tilde{Q}_{ii} = \bigl(\mu - \lambda_i^{-1}\bigr)^+, \quad i = 1, \dots, t$$

where μ is chosen to satisfy $\sum_i \tilde{Q}_{ii} = P$. The corresponding maximum mutual information is given by

$$\sum_i \bigl(\log(\mu\lambda_i)\bigr)^+$$

as before.

3.3 Error Exponents

Knowing the capacity of a channel is not always sufficient. One may be interested in knowing how hard it is to get close to this capacity. Error exponents provide a partial answer to this question by giving an upper bound to the probability of error achievable by block codes of a given length n and rate R. The upper bound is known as the random coding bound and is given by

$$\mathcal{P}(\text{error}) \le \exp(-nE_r(R)),$$

where the random coding exponent $E_r(R)$ is given by

$$E_r(R) = \max_{0 \le \rho \le 1} E_0(\rho) - \rho R,$$

where, in turn, $E_0(\rho)$ is given by the supremum over all input distributions $q_{\boldsymbol{x}}$ satisfying the energy constraint of

$$E_0(\rho, q_{\boldsymbol{x}}) = -\log \int \left[\int q_{\boldsymbol{x}}(x) p(y|x)^{1/(1+\rho)}\, dx \right]^{1+\rho} dy.$$

In our case $p(y|x) = \det(\pi I_r)^{-1} \exp\bigl(-(y-x)^\dagger(y-x)\bigr)$. If we choose $q_{\boldsymbol{x}}$ as the Gaussian distribution γ_Q we get (after some algebra)

$$E_0(\rho, Q) = \rho \log\det(I_r + (1+\rho)^{-1} HQH^\dagger) = \rho\Psi\bigl((1+\rho)^{-1})Q, H\bigr).$$

The maximization of E_0 over Q is thus same same problem as maximizing the mutual information, and we get $E_0(\rho) = \rho C(P/(1+\rho), H)$.

To choose $q_{\boldsymbol{x}}$ as Gaussian is not optimal, and a distribution concentrated on a "thin spherical shell" will give better results as in [1, §7.3]—nonetheless, the above expression is a convenient lower bound to E_0 and thus yields an upper bound to the probability of error.

4 The Gaussian Channel with Rayleigh Fading

Suppose now that the matrix H is not fixed, but is a random matrix $\boldsymbol{H}$ independent of both $\boldsymbol{x}$ and $\boldsymbol{n}$. The realization of H of $\boldsymbol{H}$ is assumed to be known at the receiver, but not at the transmitter. The channel is thus with input $\boldsymbol{x}$ and output $(\boldsymbol{y}, \boldsymbol{H}) = (\boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}, \boldsymbol{H})$. We will assume that the entries of $\boldsymbol{H}$ are independent and each entry is zero-mean, Gaussian, with independent real and imaginary parts, each with variance 1/2. Equivalently, each entry of $\boldsymbol{H}$ has uniformly distributed phase and Rayleigh distributed magnitude, with expected magnitude square equal to unity. This is intended to model a Rayleigh fading channel with enough physical separation within the transmitting and the receiving antennas to achieve independence in the entries of $\boldsymbol{H}$. We will first show that such an $\boldsymbol{H}$ is invariant under unitary transformations.

Lemma 5. *Suppose $\boldsymbol{H} \in \mathbb{C}^{r\times t}$ is a complex Gaussian matrix with independent identically distributed entries, each entry with independent real and imaginary parts with zero-mean and equal variance. Then for any unitary $U \in \mathbb{C}^{r\times r}$, and $V \in \mathbb{C}^{t\times t}$, the distribution of $U\boldsymbol{H}V^\dagger$ is the same as the distribution $\boldsymbol{H}$.*

Proof. It suffices to show that $\boldsymbol{G} = U\boldsymbol{H}$ has the same distribution as $\boldsymbol{H}$. The lemma then follows from an application of this to $\boldsymbol{G}^\dagger$. Since columns of $\boldsymbol{H}$ are independent, the columns of $\boldsymbol{G}$ are independent also. It remains to check that each column of $\boldsymbol{G}$ has the same distribution as that of $\boldsymbol{H}$. Since the columns of $\boldsymbol{H}$ are circularly symmetric complex Gaussian vectors, so are those of $\boldsymbol{G}$. If $\boldsymbol{g}_j$ and $\boldsymbol{h}_j$ are the j^{th} column of $\boldsymbol{G}$ and $\boldsymbol{H}$ respectively, then

$$\mathcal{E}[\boldsymbol{g}_j\boldsymbol{g}_j^\dagger] = U\,\mathcal{E}[\boldsymbol{h}_j\boldsymbol{h}_j^\dagger]U^\dagger = \mathcal{E}[\boldsymbol{h}_j\boldsymbol{h}_j^\dagger]$$

where the last equality holds because $\mathcal{E}[\boldsymbol{h}_j\boldsymbol{h}_j^\dagger]$ is a multiple of the identity matrix. □

In this section we will assume that the channel is memoryless: for each use of the channel an independent realization of $\boldsymbol{H}$ is drawn. In this case we are on familiar ground and the capacity can be computed as the maximum mutual information. However, the results that follow are valid verbatim for channels for which $\boldsymbol{H}$ is generated by an ergodic process: as long as the receiver observes the $\boldsymbol{H}$ process only the first order statistics are needed to determine channel capacity.

4.1 Capacity

Since the receiver knows the realization of $\boldsymbol{H}$, the channel output is the pair $(\boldsymbol{y}, \boldsymbol{H}) = (\boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}, \boldsymbol{H})$. The mutual information between input and output is then

$$\begin{aligned}\mathcal{I}\big(\boldsymbol{x}; (\boldsymbol{y}, \boldsymbol{H})\big) &= \mathcal{I}(\boldsymbol{x}; \boldsymbol{H}) + \mathcal{I}(\boldsymbol{x}; \boldsymbol{y}|\boldsymbol{H}) \\ &= \mathcal{I}(\boldsymbol{x}; \boldsymbol{y}|\boldsymbol{H}) \\ &= \mathcal{E}_{\boldsymbol{H}}[\mathcal{I}(\boldsymbol{x}; \boldsymbol{y}|\boldsymbol{H} = H)].\end{aligned}$$

We know from the previous section that if $\boldsymbol{x}$ is constrained to have covariance Q, the choice of $\boldsymbol{x}$ that maximizes $\mathcal{I}(\boldsymbol{x}; \boldsymbol{y}|\boldsymbol{H} = H)$ is the circularly symmetric complex Gaussian of covariance Q, and $\Psi(Q, H) = \log\det(I_r + HQH^\dagger)$ is the corresponding maximal mutual information. We thus need to maximize

$$\Psi(Q) = \mathcal{E}[\Psi(Q, \boldsymbol{H})] = \mathcal{E}\big[\log\det(I_r + \boldsymbol{H}Q\boldsymbol{H}^\dagger)\big]$$

over the choice of non-negative definite Q subject to $\operatorname{tr}(Q) \le P$.

Since Q is non-negative definite, we can write it as $Q = UDU^\dagger$ where U is unitary and D is non-negative and diagonal. With this substitution

$$\Psi(Q) = \mathcal{E}\big[\log\det\big(I_r + (\boldsymbol{H}U)D(\boldsymbol{H}U)^\dagger\big)\big]$$

By Lemma 5 the distribution of $\boldsymbol{H}U$ is the same as that of $\boldsymbol{H}$, and thus $\Psi(Q) = \Psi(D)$. We can thus restrict our attention to non-negative diagonal Q. Given any such Q and any permutation matrix Π, consider $Q^\Pi = \Pi Q \Pi^\dagger$. Since $\boldsymbol{H}\Pi$ has the same distribution as $\boldsymbol{H}$, $\Psi(Q^\Pi) = \Psi(Q)$. Note that for any H, the mapping $Q \mapsto I_r + HQH^\dagger$ is linear and preserves positive definiteness. Since $\log\det$ is concave on the set of positive definite matrices, $Q \mapsto \Psi(Q, H) = \log\det(I_r + HQH^\dagger)$ is concave. It then follows that $Q \mapsto \Psi(Q)$ is concave. Thus

$$\tilde{Q} = \frac{1}{t!}\sum_{\Pi} Q^\Pi$$

satisfies $\Psi(\tilde{Q}) \ge \Psi(Q)$ and $\operatorname{tr}(\tilde{Q}) = \operatorname{tr}(Q)$. Note that $\tilde{Q}$ is a multiple of the identity matrix and we conclude that the optimal Q must be of the form αI. It is clear that the maximum is achieved when α is the largest possible, namely P/t. To summarize, we have shown the following:

Theorem 1. *The capacity of the channel is achieved when $\boldsymbol{x}$ is a circularly symmetric complex Gaussian with zero-mean and covariance $(P/t)I_t$. The capacity is given by $\mathcal{E}\big[\log\det\big(I_r + (P/t)\boldsymbol{H}\boldsymbol{H}^\dagger\big)\big]$.*

Note that for fixed r, by the law of large numbers $\frac{1}{t}\boldsymbol{H}\boldsymbol{H}^\dagger \to I_r$ almost surely as

t gets large. Thus, the capacity in the limit of large t equals

$$r \log(1 + P). \tag{6}$$

4.2 Evaluation of the Capacity

Although the expectation $\mathcal{E}\big[\log\det\big(I_r + (P/t)\boldsymbol{H}\boldsymbol{H}^\dagger\big)\big]$ is easy to evaluate for either $r = 1$ or $t = 1$, its evaluation gets rather involved for r and t larger than 1. We will now show how to do this evaluation. Note that

$$\det\big(I_r + (P/t)\boldsymbol{H}\boldsymbol{H}^\dagger\big) = \det\big(I_t + (P/t)\boldsymbol{H}^\dagger\boldsymbol{H}\big)$$

and define

$$\boldsymbol{W} = \begin{cases} \boldsymbol{H}\boldsymbol{H}^\dagger & r < t \\ \boldsymbol{H}^\dagger\boldsymbol{H} & r \geq t, \end{cases}$$

$n = \max\{r, t\}$ and $m = \min\{r, t\}$. Then $\boldsymbol{W}$ is an $m \times m$ random non-negative definite matrix and thus has real, non-negative eigenvalues. We can write the capacity in terms of the eigenvalues $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_m$ of $\boldsymbol{W}$:

$$\mathcal{E}\Bigg[\sum_{i=1}^{m} \log\big(1 + (P/t)\boldsymbol{\lambda}_i\big)\Bigg] \tag{7}$$

The distribution law of $\boldsymbol{W}$ is called the *Wishart distribution with parameters m, n* and the joint density of the ordered eigenvalues is known to be (see e.g. [2] or [3, p. 37])

$$p_{\boldsymbol{\lambda},\text{ordered}}(\lambda_1, \ldots, \lambda_m) = K_{m,n}^{-1} e^{-\sum_i \lambda_i} \prod_i \lambda_i^{n-m} \prod_{i<j} (\lambda_i - \lambda_j)^2, \qquad \lambda_1 \geq \cdots \geq \lambda_m \geq 0$$

where $K_{m,n}$ is a normalizing factor. The unordered eigenvalues then have the density

$$p_{\boldsymbol{\lambda}}(\lambda_1, \ldots, \lambda_m) = (m! K_{m,n})^{-1} e^{-\sum_i \lambda_i} \prod_i \lambda_i^{n-m} \prod_{i<j} (\lambda_i - \lambda_j)^2.$$

The expectation we wish to compute

$$\begin{aligned} \mathcal{E}\Bigg[\sum_{i=1}^{m} \log\big(1 + (P/t)\boldsymbol{\lambda}_i\big)\Bigg] &= \sum_{i=1}^{m} \mathcal{E}\big[\log\big(1 + (P/t)\boldsymbol{\lambda}_i\big)\big] \\ &= m\,\mathcal{E}\big[\log\big(1 + (P/t)\boldsymbol{\lambda}_1\big)\big] \end{aligned}$$

depends only on the distribution of one of the unordered eigenvalues. To compute the density of $\boldsymbol{\lambda}_1$ we only need to integrate out the $\lambda_2, \dots, \lambda_m$:

$$p_{\boldsymbol{\lambda}_1}(\lambda_1) = \int \cdots \int p_{\boldsymbol{\lambda}}(\lambda_1, \dots, \lambda_m)\, d\lambda_2 \cdots d\lambda_m.$$

To that end, note that $\prod_{i<j}(\lambda_i - \lambda_j)$ is the determinant of a Vandermonde matrix

$$D(\lambda_1, \dots, \lambda_m) = \begin{bmatrix} 1 & \dots & 1 \\ \lambda_1 & \dots & \lambda_m \\ \vdots & & \vdots \\ \lambda_1^{m-1} & \dots & \lambda_m^{m-1} \end{bmatrix}$$

and we can write $p_{\boldsymbol{\lambda}}$ as

$$p_{\boldsymbol{\lambda}}(\lambda_1, \dots, \lambda_m) = (m! K_{m,n})^{-1} \det\bigl(D(\lambda_1 \dots, \lambda_m)\bigr)^2 \prod_i \lambda_i^{n-m} e^{-\lambda_i}.$$

With row operations we can transform $D(\lambda_1, \dots, \lambda_m)$ into

$$\tilde{D}(\lambda_1, \dots, \lambda_m) = \begin{bmatrix} \varphi_1(\lambda_1) & \dots & \varphi_1(\lambda_m) \\ \vdots & & \vdots \\ \varphi_m(\lambda_1) & \dots & \varphi_m(\lambda_m) \end{bmatrix}$$

where $\varphi_1, \dots, \varphi_m$ is the result of applying the Gram–Schmidt orthogonalization procedure to the sequence

$$1,\, \lambda,\, \lambda^2,\, \dots,\, \lambda^{m-1}$$

in the space of real valued functions with inner product

$$\langle f, g \rangle = \int_0^\infty f(\lambda) g(\lambda) \lambda^{n-m} e^{-\lambda}\, d\lambda.$$

Thus $\int_0^\infty \varphi_i(\lambda)\varphi_j(\lambda)\lambda^{n-m}e^{-\lambda}\, d\lambda = \delta_{ij}$. The determinant of D then equals (modulo multiplicative constants picked up from the row operations) the determinant of $\tilde{D}$, which in turn, by the definition of the determinant, equals

$$\det\bigl(\tilde{D}(\lambda_1, \dots, \lambda_m)\bigr) = \sum_\alpha (-1)^{\mathrm{per}(\alpha)} \prod_i \tilde{D}_{\alpha_i, i} = \sum_\alpha (-1)^{\mathrm{per}(\alpha)} \prod_i \varphi_{\alpha_i}(\lambda_i)$$

where the summation is over all permutations of $\{1,\ldots,m\}$, and $\operatorname{per}(\alpha)$ is 0 or 1 depending on the permutation α being even or odd. Thus

$$p_{\boldsymbol{\lambda}}(\lambda_1,\ldots,\lambda_m) = C_{m,n}\sum_{\alpha,\beta}(-1)^{\operatorname{per}(\alpha)+\operatorname{per}(\beta)}\prod_i \varphi_{\alpha_i}(\lambda_i)\varphi_{\beta_i}(\lambda_i)\lambda_i^{n-m}e^{-\lambda_i}.$$

Integrating over $\lambda_2,\ldots,\lambda_m$ we get

$$\begin{aligned}
p_{\boldsymbol{\lambda}_1}(\lambda_1) &= C_{m,n}\sum_{\alpha,\beta}(-1)^{\operatorname{per}(\alpha)+\operatorname{per}(\beta)}\varphi_{\alpha_1}(\lambda_1)\varphi_{\beta_1}(\lambda_1)\lambda_1^{n-m}e^{-\lambda_1}\prod_{i\geq 2}\delta_{\alpha_i\beta_i}\\
&= C_{m,n}(m-1)!\sum_{i=1}^{m}\varphi_i(\lambda_1)^2\lambda_1^{n-m}e^{-\lambda_1}\\
&= \frac{1}{m}\sum_{i=1}^{m}\varphi_i(\lambda_1)^2\lambda_1^{n-m}e^{-\lambda_1}
\end{aligned}$$

where the second equality follows from the fact that if $\alpha_i=\beta_i$ for $i\geq 2$ then $\alpha_1=\beta_1$ also (since both α and β are permutations of $\{1,\ldots,m\}$) and thus $\alpha=\beta$, and the last equality follows from the fact that $\varphi_i(\lambda_1)^2\lambda_1^{n-m}e^{-\lambda_1}$ integrates to unity and thus $C_{m,n}$ must equal $1/m!$. Observe now that the Gram–Schmidt orthonormalization yields

$$\varphi_{k+1}(\lambda) = \left[\tfrac{k!}{(k+n-m)!}\right]^{1/2} L_k^{n-m}(\lambda), \qquad k=0,\ldots,m-1$$

where $L_k^{n-m}(x) = \frac{1}{k!}e^x x^{m-n}\frac{d^k}{dx^k}(e^{-x}x^{n-m+k})$ is the associated Laguerre polynomial of order k. (See [4, §8.90,8.97].)

To summarize:

Theorem 2. *The capacity of the channel with t transmitters and r receivers under power constraint P equals*

$$\int_0^\infty \log(1+P\lambda/t)\sum_{k=0}^{m-1}\tfrac{k!}{(k+n-m)!}\big[L_k^{n-m}(\lambda)\big]^2\lambda^{n-m}e^{-\lambda}\,d\lambda \tag{8}$$

where $m=\min\{r,t\}$ and $n=\max\{r,t\}$, and L_j^i are the associated Laguerre polynomials.

Figure 1 shows the value of the integral in (8) for $1\leq r,t\leq 20$ and $P=20$dB.

Example 3. Consider $t=1$. In this case $m=1$ and $n=r$. Noting that $L_0^{n-m}(\lambda)=1$, an application of (8) yields the capacity as

$$\frac{1}{\Gamma(r)}\int_0^\infty \log(1+P\lambda)\lambda^{r-1}e^{-\lambda}\,du. \tag{9}$$

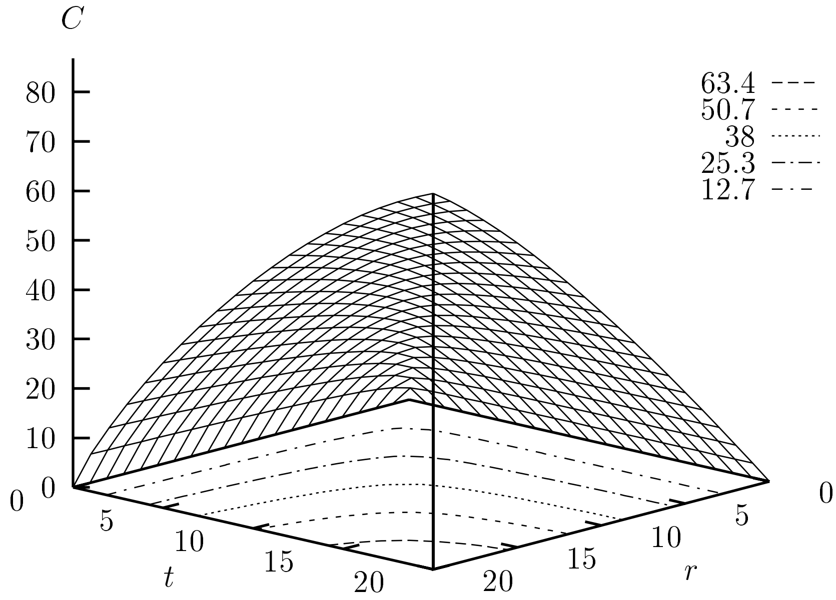


Figure 1: Capacity (in nats) vs. r and t for $P = 20$dB

The values of this integral are tabulated in Table 1 for $1 \le r \le 10$ and P from 0dB to 35dB in 5dB increments. See also Figure 2. Note that as r gets large, so does the capacity. For large r, the capacity is asymptotic to $\log(1 + Pr)$, in the sense that the difference goes to zero.

Example 4. Consider $r = 1$. As in the previous example, applying (8) yields the capacity as

$$\frac{1}{\Gamma(t)} \int_0^\infty \log\bigl(1 + P\lambda/t\bigr) \lambda^{t-1} e^{-\lambda} \, du. \tag{10}$$

As noted in (6), the capacity approaches $\log(1 + P)$ as t gets large. The values of the capacity are shown in Table 2 for various values of t and P. See also Figure 3.

Example 5. Consider $r = t$. In this case $n = m = r$, and an application of (8) yields the capacity as

$$\int_0^\infty \log(1 + P\lambda/r) \sum_{k=0}^{r-1} L_k(\lambda)^2 e^{-\lambda} \, d\lambda, \tag{11}$$

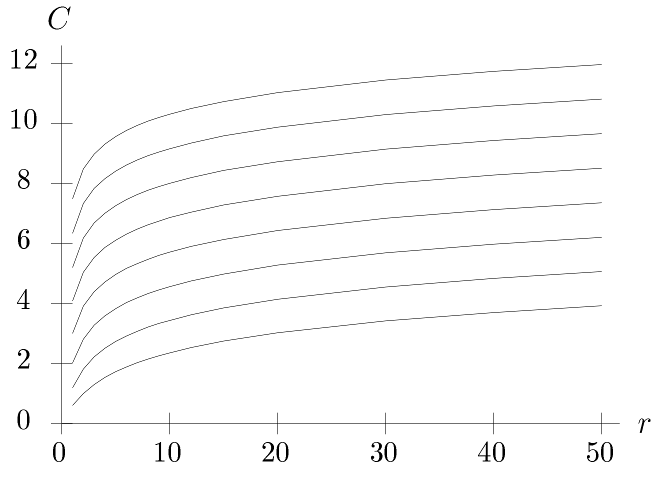
where $L_k = L_k^0$ is the Laguerre polynomial of order k.

Figure 4 shows this capacity for various values of r and P. It is clear from the figure that the capacity is very well approximated by a linear function of r. Indeed,

r \ P	0dB	5dB	10dB	15dB	20dB	25dB	30dB	35dB
1	0.5963	1.1894	2.0146	3.0015	4.0785	5.1988	6.3379	7.4845
2	1.0000	1.8133	2.8132	3.9066	5.0377	6.1824	7.3315	8.4822
3	1.2982	2.2146	3.2732	4.3922	5.5329	6.6808	7.8310	8.9820
4	1.5321	2.5057	3.5913	4.7204	5.8646	7.0136	8.1642	9.3153
5	1.7236	2.7327	3.8333	4.9679	6.1138	7.2634	8.4141	9.5652
6	1.8853	2.9183	4.0285	5.1663	6.3133	7.4632	8.6141	9.7652
7	2.0250	3.0752	4.1919	5.3319	6.4796	7.6298	8.7807	9.9319
8	2.1479	3.2110	4.3324	5.4740	6.6222	7.7726	8.9235	10.075
9	2.2576	3.3306	4.4556	5.5985	6.7471	7.8975	9.0485	10.200
10	2.3565	3.4375	4.5654	5.7091	6.8580	8.0086	9.1596	10.311

The capacity in nats of a multiple receiver, single transmitter fading channel. The path gain from the transmitter to any receiver has uniform phase and Rayleigh amplitude of unit mean square. The gains to different receivers are independent. The number of receivers is r, and P is the signal to noise ratio.

Table 1: Values of the integral in (9)



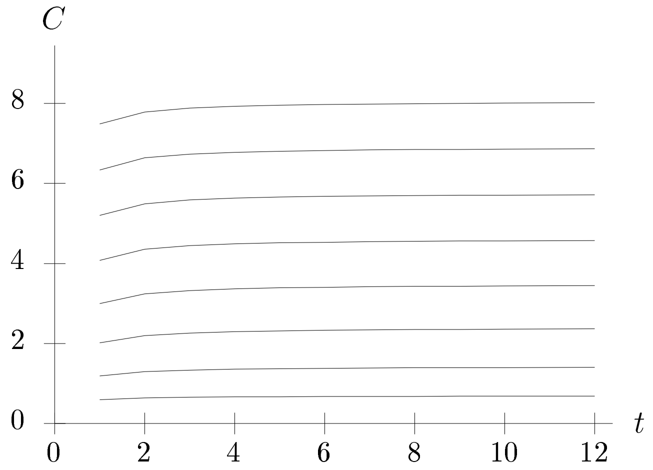
The value of the capacity (in nats) as found from (9) vs. r for $0\text{dB} \leq P \leq 35\text{dB}$ in 5dB increments.

Figure 2: Capacity vs. r for $t = 1$ and various values of P.

t \ P	0dB	5dB	10dB	15dB	20dB	25dB	30dB	35dB
1	0.5963	1.1894	2.0146	3.0015	4.0785	5.1988	6.3379	7.4845
2	0.6387	1.2947	2.1947	3.2411	4.3540	5.4923	6.6394	7.7893
3	0.6552	1.3354	2.2608	3.3236	4.4441	5.5854	6.7334	7.8837
4	0.6640	1.3570	2.2947	3.3646	4.4882	5.6305	6.7789	7.9293
5	0.6695	1.3702	2.3152	3.3891	4.5142	5.6571	6.8057	7.9561
6	0.6733	1.3793	2.3289	3.4053	4.5314	5.6746	6.8233	7.9738
7	0.6760	1.3858	2.3388	3.4169	4.5436	5.6870	6.8358	7.9863
8	0.6781	1.3907	2.3462	3.4255	4.5527	5.6963	6.8451	7.9956
9	0.6797	1.3946	2.3519	3.4322	4.5598	5.7034	6.8523	8.0028
10	0.6810	1.3977	2.3565	3.4375	4.5654	5.7091	6.8580	8.0086

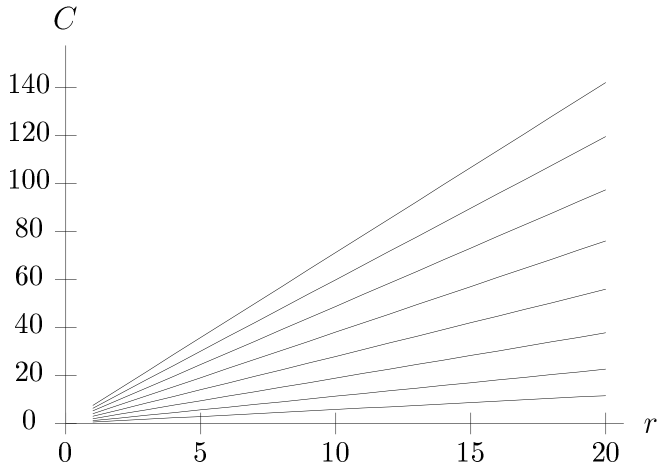
The capacity in nats of a multiple transmitter, single receiver fading channel. The path gain from any transmitter to the receiver has uniform phase and Rayleigh amplitude with unit mean square. The fades for each path gain is independent. The number of transmitters is t and P is the signal to noise ratio.

Table 2: Values of the integral in (10).



The value of the capacity (in nats) as found from (10) vs. t for $0\text{dB} \leq P \leq 35\text{dB}$ in 5dB increments.

Figure 3: Capacity vs. t for $r = 1$ and various values of P.



The value of the capacity (in nats) as found from (11) vs. r for 0dB $\leq P \leq$ 35dB in 5dB increments.

Figure 4: Capacity vs. r for $r = t$ and various values of P.

first rewrite (7) as

$$C = \mathcal{E} \int_0^\infty \log\Big(1 + \frac{Pm}{t}\nu\Big) m \, dF^{\frac{1}{m}\boldsymbol{W}}(\nu)$$

where $F_A(x)$ is the empirical distribution of the eigenvalues of an $m \times m$ Hermitian matrix A:

$$F^A(x) = \frac{\text{the number of eigenvalues of } A \text{ less than } x}{m}.$$

A very general result from the theory of random matrices (see, e.g., [5]) says that for $\boldsymbol{W}$ defined as above, as $n = \max\{r, t\}$ and $m = \min\{r, t\}$ are increased with n/m approaching a limit $\tau \geq 1$,

$$\frac{dF^{\frac{1}{m}\boldsymbol{W}}(\nu)}{d\nu} \to \begin{cases} \frac{1}{2\pi}\sqrt{\left(\frac{\nu_+}{\nu} - 1\right)\left(1 - \frac{\nu_-}{\nu}\right)} & \text{for } \nu \in [\nu_-, \nu_+] \\ 0 & \text{otherwise,} \end{cases} \tag{12}$$

with $\nu_\pm = (\sqrt{\tau} \pm 1)^2$. Thus, in the limit of large r and t,

$$\frac{C}{m} \to \frac{1}{2\pi} \int_{\nu_-}^{\nu_+} \log\Big(1 + \frac{Pm}{t}\nu\Big) \sqrt{\Big(\frac{\nu_+}{\nu} - 1\Big)\Big(1 - \frac{\nu_-}{\nu}\Big)} \, d\nu. \tag{13}$$

For the case under consideration, $m = n = r = t$, for which $\nu_- = 0$, $\nu_+ = 4$, and

$$C \sim r \int_0^4 \log(1 + P\nu) \frac{1}{\pi} \sqrt{\frac{1}{\nu} - \frac{1}{4}}\, d\nu$$

which is linear in r as observed before from the figure.

Remark 2. The result from the theory of random matrices used in Example 5 applies to random matrices that are not necessarily Gaussian. For equation (13) to hold it is sufficient for $\boldsymbol{H}$ to have i.i.d. entries of unit variance.

Remark 3. The reciprocity property that we observed for deterministic H does not hold for random H: Compare Examples 3 and 4 where the corresponding H's are transposes of each other. In Example 3, capacity increases without bound as r gets large, whereas in Example 4 the capacity is bounded from above.

Nonetheless, interchanging r and t does not change the matrix $\boldsymbol{W}$, and the capacity depends only on P/t and the eigenvalues of $\boldsymbol{W}$. Thus, if $C(r, t, P)$ denotes the capacity of a channel with r receivers, t transmitters and total transmitter power P, then

$$C(a, b, Pb) = C(b, a, Pa).$$

Remark 4. In the computation preceding Theorem 2 we obtained the density of one of the unordered eigenvalues of the complex Wishart matrix $\boldsymbol{W}$. Using the identity (19) in the appendix we can find the joint density of any number k of unordered eigenvalues of $\boldsymbol{W}$:

$$p_{\boldsymbol{\lambda}_1,\dots,\boldsymbol{\lambda}_k}(\lambda_1, \dots, \lambda_k) = \frac{(m-k)!}{m!} \det\left(D_k(\lambda_1, \dots, \lambda_k)^\dagger D_k(\lambda_1, \dots, \lambda_k)\right) \prod_{i=1}^k \lambda_i^{n-m} e^{-\lambda_i}$$

where

$$D_k(\lambda_1, \dots, \lambda_k) = \begin{bmatrix} \varphi_1(\lambda_1) & \dots & \varphi_1(\lambda_k) \\ \vdots & & \vdots \\ \varphi_m(\lambda_1) & \dots & \varphi_m(\lambda_k) \end{bmatrix}.$$

4.3 Error Exponents

As we did in the case of deterministic H we can compute the error exponent in the case of fading channel. To that end, note first that

$$E_0(\rho, q_{\boldsymbol{x}}) = -\log \iint \left[\int q_{\boldsymbol{x}}(x) p(y, H|x)^{1/(1+\rho)}\, dx \right]^{1+\rho} dy\, dH.$$

Since $\boldsymbol{H}$ is independent of $\boldsymbol{x}$, $p(y, H|x) = p_{\boldsymbol{H}}(H)p(y|x, H)$ and thus

$$E_0(\rho, q_{\boldsymbol{x}}) = -\log \mathcal{E}\left[\int \left[\int q_{\boldsymbol{x}}(x)p(y|x, \boldsymbol{H})^{1/(1+\rho)}\, dx\right]^{1+\rho} dy\right].$$

Note that

$$p(y|x, H) = \det(\pi I_r)^{-1} \exp\bigl(-(y - Hx)^\dagger(y - Hx)\bigr).$$

and for $q_{\boldsymbol{x}} = \gamma_Q$, the Gaussian distribution with covariance Q, we can use the results for the deterministic H case to conclude

$$E_0(\rho, \gamma_Q) = -\log \mathcal{E}\bigl[\det(I_r + (1 + \rho)^{-1}\boldsymbol{H}Q\boldsymbol{H}^\dagger)^{-\rho}\bigr].$$

Noting that $A \to \det(A)^{-\rho}$ is a convex function, the argument we used previously to show that $Q = (P/t)I_t$ maximizes the mutual information applies to maximizing E_0 as well, and we obtain

$$E_0(\rho) = -\log \mathcal{E}\left[\det\left(I_r + \frac{P}{t(1+\rho)}\boldsymbol{H}\boldsymbol{H}^\dagger\right)^{-\rho}\right]. \tag{14}$$

To efficiently compute E_0, one would represent the Wishart eigenvalue density as a Vandermonde determinant, (just as in the previous section), and orthonormalize the monomials $1, \lambda, \lambda^2, \ldots, \lambda^{m-1}$, with respect to the inner product

$$\langle f, g\rangle = \int_0^\infty f(\lambda)g(\lambda)\lambda^{n-m}\left(1 + \tfrac{P}{t(1+\rho)}\lambda\right)^{-\rho} e^{-\lambda}\, d\lambda.$$

The multiplicative factor picked up in the orthonormalization is the value of the expectation in (14).

As before, the restriction of $q_{\boldsymbol{x}}$ to Gaussian distributions is suboptimal, but this choice leads to simpler expressions.

5 Non-ergodic channels

We had remarked at the beginning of the previous section that the maximum mutual information has the meaning of capacity when the channel is memoryless, i.e., when each use of the channel employs an independent realization of $\boldsymbol{H}$. This is not the only case when the maximum mutual information is the capacity of the channel. In particular, if the process that generates $\boldsymbol{H}$ is ergodic, then too, we can achieve rates arbitrarily close to the maximum mutual information.

In contrast, for the case in which $\boldsymbol{H}$ is chosen randomly at the beginning of all time and is held fixed for all the uses of the channel, the maximum mutual information is in general not equal to the channel capacity. In this section we will focus on such a case when the entries of $\boldsymbol{H}$ are i.i.d., zero-mean circularly symmetric complex Gaussians with $\mathcal{E}[|\boldsymbol{h}_{ij}|^2] = 1$, the same distribution we have analyzed in the previous section.

5.1 Capacity

In the case described above, the Shannon capacity of the channel is zero: however small the rate we attempt to communicate at, there is a non-zero probability that the realized $\boldsymbol{H}$ is incapable of supporting it no matter how long we take our code length. On the other hand one can talk about a tradeoff between *outage probability* and supportable rate. Namely, given a rate R, and power P, one can find $P_{\text{out}}(R,P)$ such that for any rate less than R and any δ there exists a code satisfying the power constraint P for which the error probability is less than δ for all but a set of H whose total probability is less than $P_{\text{out}}(R,P)$:

$$P_{\text{out}}(R,P) = \inf_{\substack{Q:Q\geq 0\\ \text{tr}(Q)\leq P}} \mathcal{P}\big(\Psi(Q,\boldsymbol{H}) < R\big) \tag{15}$$

where

$$\Psi(Q,\boldsymbol{H}) = \log\det(I_r + \boldsymbol{H}Q\boldsymbol{H}^\dagger).$$

This approach is taken in [6] in a similar problem.

In this section, as in the previous section we will take the distribution of $\boldsymbol{H}$ to be such that the entries of $\boldsymbol{H}$ are independent zero-mean Gaussians, each with independent real and imaginary parts with variance $1/2$.

Example 6. Consider $t = 1$. In this case, it is clear that the $Q = P$ is optimal. The outage probability is then

$$\mathcal{P}\big(\log\det(I_r + \boldsymbol{H}P\boldsymbol{H}^\dagger) < R\big) = \mathcal{P}\big(\log(1 + P\boldsymbol{H}^\dagger\boldsymbol{H}) < R\big)$$

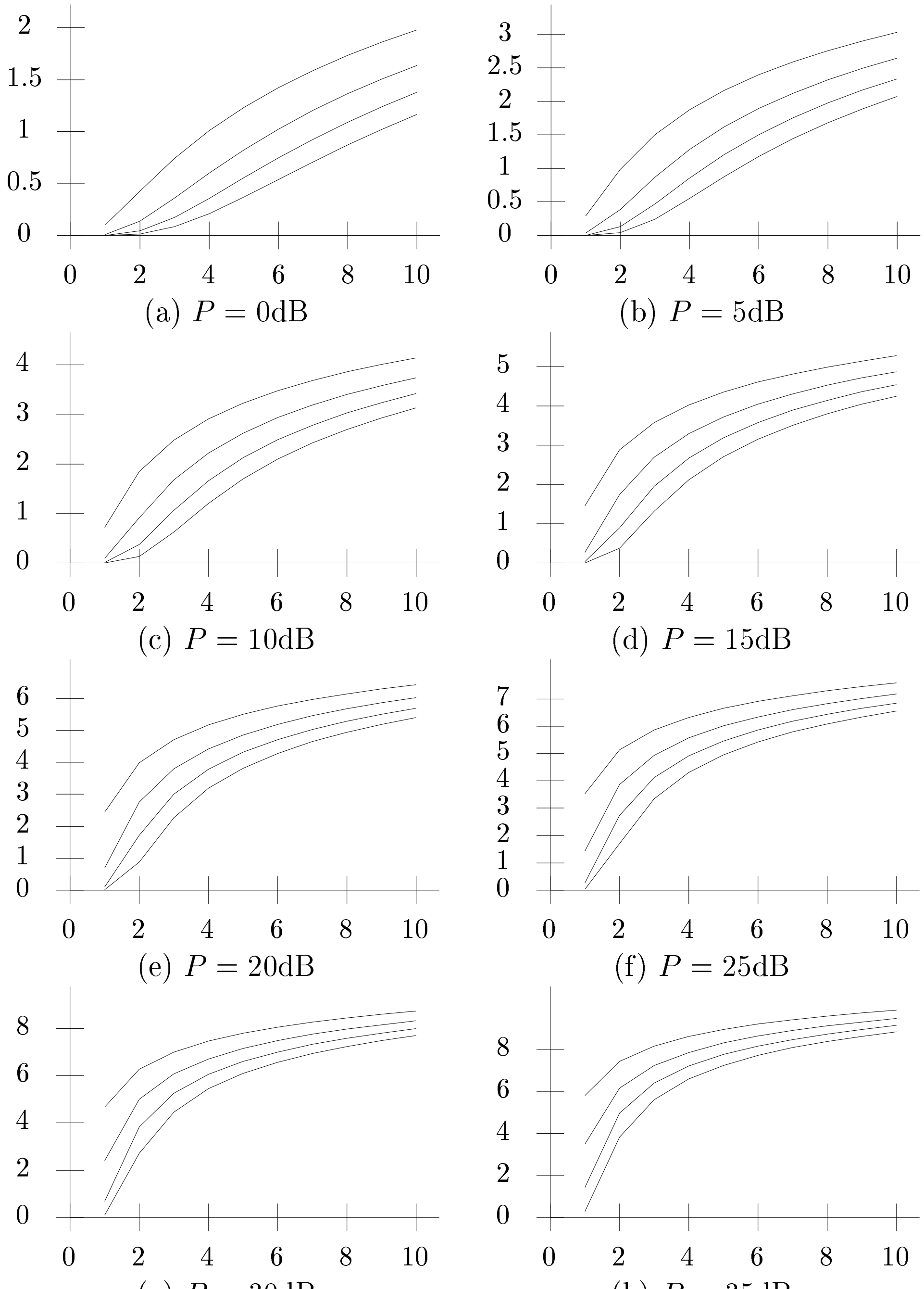
Since $\boldsymbol{H}^\dagger\boldsymbol{H}$ is a χ^2 random variable with $2r$ degrees of freedom and mean r, we can compute the outage probability as

$$P_{\text{out}}(R,P) = \frac{\gamma\big(r, (e^R - 1)/P\big)}{\Gamma(r)}, \tag{16}$$

where $\gamma(a,x)$ is the incomplete gamma function $\int_0^x u^{a-1}e^{-u}\,du$. Let $\psi(P,\epsilon)$ be the value of R that satisfies

$$\mathcal{P}(\Psi(P,\boldsymbol{H}) \leq R) = \epsilon. \tag{17}$$

Figure 5 shows $\psi(P,\epsilon)$ as a function of r for various values of ϵ and P.



(a) $P = 0$dB (b) $P = 5$dB (c) $P = 10$dB (d) $P = 15$dB (e) $P = 20$dB (f) $P = 25$dB (g) $P = 30$dB (h) $P = 35$dB

$\psi(P, \epsilon)$ vs. r at $t = 1$ for various values of P and ϵ. Recall that $\psi(P, \epsilon)$ is the highest rate for which the outage probability is less than ϵ. Each set of curves correspond to the P indicated below it. Within each set the curves correspond, in descending order, to $\epsilon = 10^{-1}$, 10^{-2}, 10^{-3}, 10^{-4}.

Figure 5: The ϵ-capacity for $t = 1$ as defined by (17).

Note that by Lemma 5 the distribution of $\boldsymbol{H}U$ is the same as that of $\boldsymbol{H}$ for unitary U. Thus, we can conclude that

$$\Psi(UQU^\dagger, \boldsymbol{H})$$

has the same distribution as $\Psi(Q, \boldsymbol{H})$. By choosing U to diagonalize Q we can restrict our attention to diagonal Q.

The symmetry in the problem suggests the following conjecture.

CONJECTURE. *The optimal Q is of the form*

$$\frac{P}{k}\,\mathrm{diag}(\underbrace{1,\ldots,1}_{k \text{ ones}},\underbrace{0,\ldots,0}_{t-k \text{ zeros}})$$

for some $k = 1, \ldots, t$. The value of k depends on the rate: higher the rate (i.e., higher the outage probability), smaller the k.

As one shares the power equally between more transmitters, the expectation of Ψ increases, but the tails of its distribution decay faster. To minimize the probability of outage, one has to maximize the probability mass of Ψ that lies to the right of the rate of interest. If one is interested in achieving rates higher than the expectation of Ψ, then it makes sense to use a small number of transmitters to take advantage of the slow decay of the tails of the distribution of Ψ. Of course, the corresponding outage probability will still be large (larger than $\frac{1}{2}$, say).

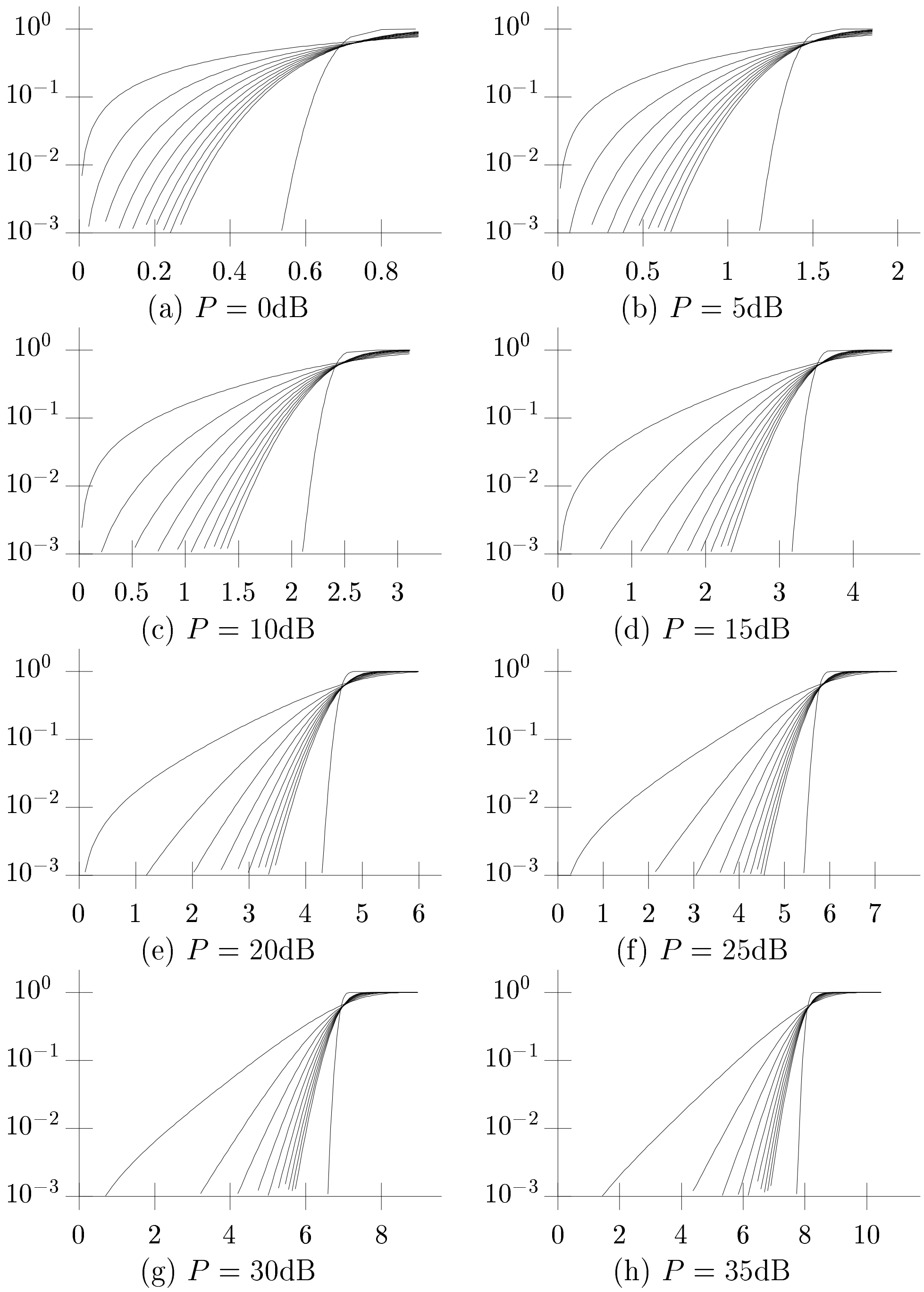
Example 7. Consider $r = 1$. With the conjecture above, it suffices to compute $\mathcal{P}\big(\Psi\big((P/t)I_t, \boldsymbol{H}\big) < R\big)$ for all values of t; if the actual number of transmitters is, say, τ, then the outage probability will be the minimum of the probabilities for $t = 1, \ldots, \tau$. As in Example 6 we see that $\boldsymbol{H}\boldsymbol{H}^\dagger$ is a χ^2 statistic with $2t$ degrees of freedom and mean t, thus

$$\mathcal{P}\big(\Psi\big((P/t)I_t, \boldsymbol{H}\big) \le R\big) = \frac{\gamma(t, t(e^R - 1)/P)}{\Gamma(t)}.$$

Figure 6 shows this distribution for various values of t and P. It is clear from the figure that large t performs better at low R and small t performs better at high R, in keeping with the conjecture. As in Example 3, let $\psi(P, \epsilon)$ be the value of R satisfying

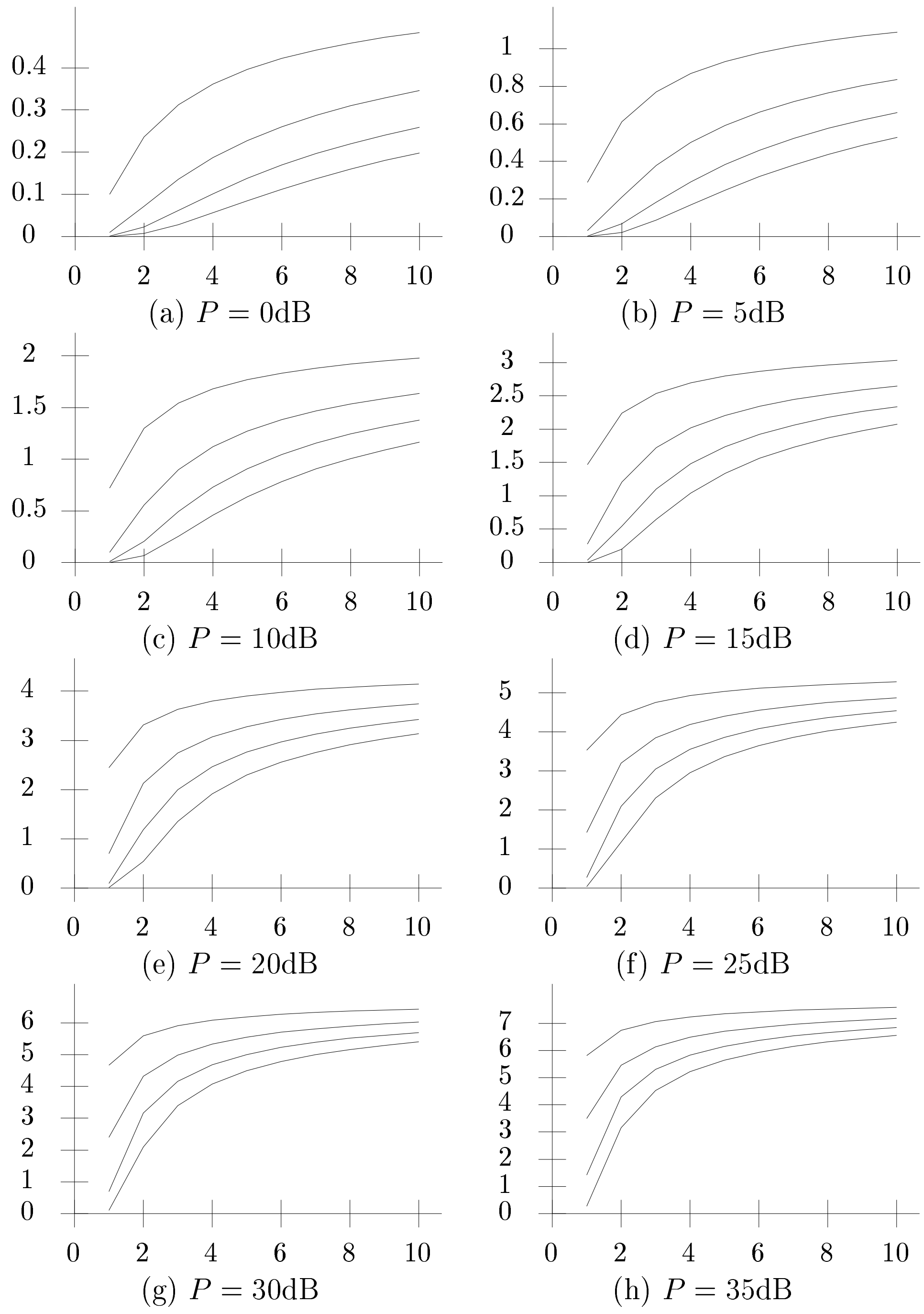
$$\mathcal{P}\big(\Psi\big((P/t)I_t, \boldsymbol{H}\big) \le R\big) = \epsilon. \tag{18}$$

Figure 7 shows $\psi(P, \epsilon)$ vs. t for various values of P and ϵ. For the small ϵ values considered in the figure, using all available transmitters is always better than using a subset.



$\mathcal{P}\big(\Psi\big(P/t)I_t, \boldsymbol{H}\big) \leq R\big)$ vs. R for various values of P and t. Each set of curves corresponds to the P indicated below it. Within each set, the curves correspond, in the order of increasing sharpness, to $t = 1$, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 100.

Figure 6: Distribution of $\Psi\big((P/t)I_t, \boldsymbol{H}\big)$ for $r = 1$.



(a) $P = 0$dB (b) $P = 5$dB (c) $P = 10$dB (d) $P = 15$dB (e) $P = 20$dB (f) $P = 25$dB (g) $P = 30$dB (h) $P = 35$dB

$\psi(P, \epsilon)$ vs. t for various values of P and ϵ. Recall that $\psi(P, \epsilon)$ is the highest rate for which the outage probability remains less than ϵ. Each set of curves corresponds to the P indicated below it. Within each set the curves correspond, in descending order, to $\epsilon = 10^{-1}$, 10^{-2}, 10^{-3}, 10^{-4}.

Figure 7: The ϵ-capacity for $r = 1$, as defined by (18).

6 Multiaccess Channels

Consider now a number of transmitters, say M, each with t transmitting antennas, and each subject to a power constraint P. There is a single receiver with r antennas. The received signal $\boldsymbol{y}$ is given by

$$\boldsymbol{y} = [\boldsymbol{H}_1 \ldots \boldsymbol{H}_M] \begin{bmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_M \end{bmatrix} + \boldsymbol{n}$$

where $\boldsymbol{x}_m$ is the signal transmitted by the m^{th} transmitter, $\boldsymbol{n}$ is Gaussian noise as in (1), and $\boldsymbol{H}_m$, $m = 1, \ldots, M$ are $r \times t$ complex matrices. We assume that the receiver knows all the $\boldsymbol{H}_m$'s, and that these have independent circularly symmetric complex Gaussian entries of zero mean and unit variance. The multiuser capacity for this communication scenario can be evaluated easily by exploting the nature of the solution to the single user scenario discussed above. Namely, since the capacity achieving distribution for the single user scenario yields an i.i.d. solution for each antenna, that the users in the multiuser scenario cannot cooperate becomes immaterial. A rate vector $(R_1, \ldots, R_M)$ will be achievable if

$$\sum_{i=1}^{m} R_{[i]} \leq C(r, mt, mP), \quad \text{for all } m = 1, \ldots, M$$

where $(R_{[1]}, \ldots, R_{[M]})$ is the ordering of the rate vector from the largest to the smallest, and $C(a, b, P)$ is the single user a receiver b transmitter capacity under power constraint P.

7 Conclusion

The use of multiple antennas will greatly increase the achievable rates on fading channels if the channel parameters can be estimated at the receiver and if the path gains between different antenna pairs behave independently. The second of these requirements can be met with relative ease and is somewhat technical in nature. The first requirement is a rather tall order, and can be justified in certain communication scenarios and not in others. Since the original writing of this monograph in late 1994 and early 1995, there has been some work in which the assumption of the availability of channel state information is replaced with the assumption of a slowly varying channel, see e.g., [7].

Appendix

Theorem. *Given m functions $\varphi_1, \dots, \varphi_m$, orthonormal with respect to F, i.e.,*

$$\int \varphi_i(\lambda)\varphi_j(\lambda)\, dF(\lambda) = \delta_{ij},$$

let

$$D_k(\lambda_1, \dots, \lambda_k) = \begin{bmatrix} \varphi_1(\lambda_1) & \dots & \varphi_1(\lambda_k) \\ \vdots & & \vdots \\ \varphi_m(\lambda_1) & \dots & \varphi_m(\lambda_k) \end{bmatrix}$$

and $A_k(\lambda_1, \dots, \lambda_k) = D_k(\lambda_1, \dots, \lambda_k)^\dagger D_k(\lambda_1, \dots, \lambda_k)$. Then

$$\int \det\big(A_k(\lambda_1, \dots, \lambda_k)\big)\, dF(\lambda_k) = (m - k + 1) \det\big((A_{k-1}(\lambda_1, \dots, \lambda_{k-1})\big). \tag{19}$$

Proof. Let $\Phi(\lambda) = [\varphi_1(\lambda), \dots, \varphi_m(\lambda)]^\dagger$. Then the (i, j)th element of $A_k(\lambda_1, \dots, \lambda_k)$ is $\Phi(\lambda_i)^\dagger\Phi(\lambda_j)$. Note that $\int \Phi(\lambda)^\dagger\Phi(\lambda)\, dF(\lambda) = m$ and $\int \Phi(\lambda)\Phi(\lambda)^\dagger\, dF(\lambda) = I_m$. By the definition of the determinant

$$\det(A_k(\lambda_1, \dots , \lambda_k)) = \sum_\alpha (-1)^{\mathrm{per}(\alpha)} \textstyle\prod_{i=1}^k \Phi(\lambda_i)^\dagger\Phi(\lambda_{\alpha_i})$$

where the sum is over all permutations α of $\{1, \dots, k\}$. Let us separate the summation over α into k summations, those for which $\alpha_j = k$, $j = 1, \dots, k$, and consider each sum in turn. For the j^{th} sum, $j = 1, \dots, k-1$, $\alpha_j = k$ for $j \neq k$. For such an α we can define β as $\beta_i = \alpha_i$ for $i \neq j, k$, and $\beta_j = \alpha_k$. Note that β ranges over all permutations of $\{1, \dots, k-1\}$ and that $\mathrm{per}(\beta)$ differs from $\mathrm{per}(\alpha)$ by 1.

$$\begin{aligned} &\sum_{\alpha:\alpha_j=k} (-1)^{\mathrm{per}(\alpha)} \textstyle\prod_{i=1}^k \Phi(\lambda_i)^\dagger\Phi(\lambda_{\alpha_i}) \\ &\qquad = \sum_{\alpha:\alpha_j=k} (-1)^{\mathrm{per}(\alpha)} \Big(\textstyle\prod_{i\neq j,k} \Phi(\lambda_i)^\dagger\Phi(\lambda_{\alpha_i})\Big) \Phi(\lambda_j)^\dagger\Phi(\lambda_{\alpha_j})\Phi(\lambda_k)^\dagger\Phi(\lambda_{\alpha_k}) \\ &\qquad = -\sum_\beta (-1)^{\mathrm{per}(\beta)} \Big(\textstyle\prod_{i\neq j} \Phi(\lambda_i)^\dagger\Phi(\lambda_{\beta_i})\Big) \Phi(\lambda_j)^\dagger\Phi(\lambda_k)\Phi(\lambda_k)^\dagger\Phi(\lambda_{\beta_j}). \end{aligned}$$

Integrating over λ_k, and recalling $\int \Phi(\lambda)\Phi(\lambda)^\dagger\, dF(\lambda) = I_m$,

$$\begin{aligned} \int \sum_{\alpha:\alpha_j=k} (-1)^{\mathrm{per}(\alpha)} \textstyle\prod_{i=1}^k \Phi(\lambda_i)^\dagger\Phi(\lambda_{\alpha_i})\, dF(\lambda_k) &= -\sum_\beta (-1)^{\mathrm{per}(\beta)} \textstyle\prod_{i=1}^{k-1} \Phi(\lambda_i)^\dagger\Phi(\lambda_{\beta_i}) \\ &= -\det(A_{k-1}(\lambda_1, \dots, \lambda_{k-1})) \end{aligned}$$

So, the contribution of the first $k-1$ sums to the integral in (19) is $-(k-1)\det(A_{k-1})$. For the last sum $\alpha_k = k$. Define β as $\beta_i = \alpha_i$ for $i \neq k$. As before β ranges over the permutations of $\{1, \ldots, k-1\}$, but now $\mathrm{per}(\beta) = \mathrm{per}(\alpha)$.

$$\sum_{\alpha:\alpha_k=k} (-1)^{\mathrm{per}(\alpha)} \prod_{i=1}^{k} \Phi(\lambda_i)^\dagger \Phi(\lambda_{\alpha_i}) = \sum_{\beta} (-1)^{\mathrm{per}(\beta)} \left(\prod_{i=1}^{k-1} \Phi(\lambda_i)^\dagger \Phi(\lambda_{\beta_i}) \right) \Phi(\lambda_k)^\dagger \Phi(\lambda_k).$$

Integrating over λ_k, and recalling $\int \Phi(\lambda)^\dagger \Phi(\lambda)\, dF(\lambda) = m$,

$$\begin{aligned} \int \sum_{\alpha:\alpha_k=k} (-1)^{\mathrm{per}(\alpha)} \prod_{i=1}^{k} \Phi(\lambda_i)^\dagger \Phi(\lambda_{\alpha_i})\, dF(\lambda_k) &= m \sum_{\beta} (-1)^{\mathrm{per}(\beta)} \prod_{i=1}^{k-1} \Phi(\lambda_i)^\dagger \Phi(\lambda_{\beta_i}) \\ &= m \det(A_{k-1}(\lambda_1, \ldots, \lambda_{k-1})) \end{aligned}$$

And so, the contribution of the last sum to the integral in (19) is $m \det(A_{k-1})$. The result now follows by adding the two contributions. $\square$

Acknowledgements

I would like to thank S. Hanly, C. Mallows, J. Mazo, A. Orlitsky, D. Tse and A. Wyner for helpful comments and discussions.

References

[1] R. G. Gallager, *Information Theory and Reliable Communication.* New York: John Wiley & Sons, 1968.

[2] A. T. James, "Distributions of matrix variates and latent roots derived from normal samples," *Annals of Mathematical Statistics*, vol. 35, pp. 475–501, 1964.

[3] A. Edelman, *Eigenvalues and Condition Numbers of Random Matrices.* PhD thesis, Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, 1989.

[4] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products.* New York: Academic Press, corrected and enlarged ed., 1980.

[5] J. W. Silverstein, "Strong Convergence of the Empirical Distribution of Eigenvalues of Large Dimensional Random Matrices," *Journal of Multivariate Analysis*, vol. 55, pp 331-339, 1995.

[6] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Transactions on Vehicular Technology*, vol. 43, pp. 359–378, May 1994.

[7] B. Hochwald and T. L. Marzetta, "Capacity of a mobile multiple-antenna communication link in a Rayleigh flat-fading environment," to appear in *IEEE Transactions on Information Theory.*