On the Capacity of some Channels with Channel State Information

Giuseppe Caire [*]
Mobile Communications Group
Institute Eurecom
06904 Sophia Antipolis, FRANCE
E-mail: `caire@eurecom.fr`

Shlomo Shamai
Department of Electrical Engineering
Technion – Israel Institute of Technology
Haifa 32000, ISRAEL
E-mail: `sshlomo@ee.technion.ac.il`

March 3, 1999

Abstract

In this paper we study the capacity of some channels whose conditional output probability distribution depends on a state process independent of the channel input and where channel state information (CSI) signals are available both at the transmitter (CSIT) and at the receiver (CSIR). When the channel state and the CSI signals are jointly i.i.d., the channel reduces to a case studied by Shannon. In this case, we show that when the CSIT is a deterministic function of the CSIR, optimal coding is particularly simple. When the state process has memory, we provide a general capacity formula and we give some more restrictive conditions under which the capacity has still a simple single-letter characterization, allowing simple optimal coding. Finally, we turn to the AWGN channel with fading and we provide a generalization of some results about capacity with CSI for this channel. In particular, we show that variable-rate coding (or multiplexing of several codebooks) is not needed to achieve capacity and, even when the CSIT is not perfect, the capacity achieving power allocation is of the *waterfilling* type.

Keywords: Channel capacity, channel state information, fading channels, power allocation.

1 Introduction

Channels whose output conditional probability depends on a time-varying state have been widely studied. Depending on the assumptions on the channel state and on the availability of channel state

[*] This work has been done while G. Caire was visiting the Electrical Engineering Dept. of Princeton University, Princeton, NJ 08544, USA, under the CNR grant No. 078484

information (CSI) at the transmitter (CSIT) and at the receiver (CSIR), a whole range of problems rises, each related to some physical situation of interest. A partial list includes channels with CSIT only [1, 2, 3], the Gilbert-Elliot channel [4, 5] and, more in general, the finite-state Markov channels without CSI [6], a number of compound channels studied in [7], the block-interference channel of [8] and various forms of arbitrary-varying channel [9, 10, 11].

More recently, driven by the growing interest in mobile wireless communications, numerous works have been devoted to assessing the information theoretic limits of Gaussian fading channels, which can be modeled as the continuous counterpart of the discrete channels mentioned above (see for example [12, 13, 14, 15, 16, 17, 18, 19]). Also, some recent works have been devoted to the more realistic case of non-perfect CSI [20, 21, 22].

The receiver may have some CSI from the insertion of training symbols in the transmitted signal. Moreover, it can wait until the end of transmission before decoding, so that it has CSI over the whole received sequence. For the CSI at the transmitter, we distinguish between channels where CSIT is *causal* from channels where CSIT is *non-causal*. In the case of causal CSIT, first introduced by Shannon [1], the transmitter at time n knows only the CSI signal from time 1 to n. In the case of non-causal CSIT, introduced by Gelfand and Pinsker [10], the transmitter knows in advance the realization of the state sequence from the start to the end of transmission. Clearly, both the information theoretic problems and the practical applications related with these two classes of channels are rather different. Causal CSIT is more suited to situations where the channel state is measured sequentially. For example, in a fading channel where measures of the instantaneous channel attenuation are obtained at the receiver and sent back to the transmitter via a feedback link, as in power control schemes currently implemented in some cellular standards [23]. Non-causal CSIT is more suited to situations where the transmitter can sound the channel beforehand over the whole transmission span, as in the case of the storage of encoded information in a computer memory with defective cells [24, 25].

In this paper we deal with channels with causal CSIT. In Section 2 we consider the case of i.i.d. states studied by Salehi [20] and we show that the channel reduces to Shannon's channel [1]. We show that when the CSIT is a deterministic function of the CSIR, optimal codes can be constructed directly over the input alphabet, while in general, coding over an expanded alphabet is required (a related result in the case of a discrete additive channel can be found in [26]). In Section 3 we consider the case of states with memory. In the case of no CSIR, the capacity was determined by Jelinek [2, 3] for two classes of channels: the *finite-state Markov indecomposable* and the *strongly indecomposable* channels. The general case with arbitrary memory has been recently treated in [27]. Unfortunately, the capacity formulas in [2, 3, 27] do not provide much intuition on practical good coding schemes. Nevertheless, under some more restrictive conditions, a single-letter characterization of channel capacity is still valid, so that codes constructed directly over the input alphabet are still optimal. In Section 4 we turn to the AWGN channel with fading and we provide a generalization of some results about capacity with CSI for this channel. As a by-product of this analysis, we show that: i) variable-rate coding (or multiplexing of several codebooks) is not needed to achieve capacity; ii) the capacity achieving power allocation is of the *waterfilling* type, even with non-perfect CSIT; iii) constant power allocation is optimal for the case of no CSIT and perfect CSIR, even in the case of non-i.i.d. fading (as mentioned explicitly in [13]). Finally, in Section 5 we provide some numerical examples of the fading AWGN channel with non-perfect CSIT.

Notation conventions are as follows: random variables are denoted by upper case letters (e.g., A); a lower case letter a is used to denote a particular value of A; the short-hand notation A_M^N indicates a sequence of random variables $(A_M, \ldots, A_N)$ and a_M^N denotes a particular value of A_M^N; $\{A_n\}$ denotes a generic sequence of random variables A_M^N, for any arbitrary M and N.

2 Channel model and results for i.i.d. channel states

Consider the channel of Fig. 1, with discrete input $X_n \in \mathcal{X}$, output $Y_n \in \mathcal{Y}$ and state $S_n \in \mathcal{S}$, characterized by a family of conditional output probability distributions $\{p(y|x,s) : s \in \mathcal{S}\}$ such that

$$p(y_1^N | x_1^N, s_1^N) = \prod_{n=1}^{N} p(y_n | x_n, s_n) \tag{1}$$

The transmitter and the receiver are provided with the CSIT signal $U_n \in \mathcal{U}$ and with the CSIR signal $V_n \in \mathcal{V}$, respectively. After conditioning on X_n and on S_n, Y_n is statistically independent of U_m, V_m (for all m) and of S_m, X_m (for all $m \neq n$). Moreover, S_n, U_n and V_n are independent of the past channel inputs (i.e., this model does not take into account intersymbol interference channels [28]). We say that the CSIT (resp. the CSIR) is *perfect* if U_n (resp. V_n) is equal to S_n, and that it is *absent* if U_n (resp. V_n) is statistically independent of S_n.

Encoding and decoding. A block code of length N for the channel of Fig. 1 is defined by a sequence of N encoding functions $f_n : \mathcal{W} \times \mathcal{U}^n \to \mathcal{X}$, for $n = 1, \ldots, N$, such that $x_n = f_n(w, u_1^n)$, where w ranges over the set of possible source messages $\mathcal{W}$ and u_1^n is the realization of the CSIT up to time n. The decoding function is $\phi : \mathcal{Y}^N \times \mathcal{V}^N \to \mathcal{W}$, such that the decoded message is $\widehat{w} = \phi(y_1^N, v_1^N)$.

Channel capacity. In the case where $\{S_n\}$ is an i.i.d. sequence, $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{S}$ are finite alphabets, the CSIT is perfect and the CSIR is absent, the capacity of the above channel was obtained by Shannon [1] and it is given by

$$C = \max_{q(t)} I(T;Y) \tag{2}$$

where $T \in \mathcal{X}^{|\mathcal{S}|}$ is a random vector of length $|\mathcal{S}|$ with elements in $\mathcal{X}$ and probability distribution $q(t)$. A code for this channel is a set of $|\mathcal{W}|$ sequences of length N of vectors $t \in \mathcal{X}^{|\mathcal{S}|}$. For a given source message $w \in \mathcal{W}$, the code word $t_1^N(w)$ is selected. At each time n, the channel input is given by $x_n = t_n(w, s_n)$, where $t_n(w, s)$ denotes the s-th element of the vector $t_n(w)$. A code word $t_1^N(w)$ defines a sequence of N functions $\mathcal{S} \to \mathcal{X}$. Then, this encoding rule is a particular case of the general encoding rule given above. The remarkable fact is that, in this case, this is enough to achieve capacity.

A generalization of Shannon's result has been provided by Salehi [20] in the case where $\{(S_n, U_n, V_n)\}$ is an i.i.d. sequence over $\mathcal{S} \times \mathcal{U} \times \mathcal{V}$, with joint distribution $\omega(s, u, v)$. In this case, the channel capacity is given by

$$C = \max_{q(t)} I(T;Y|V) \tag{3}$$

where $T \in \mathcal{X}^{|\mathcal{U}|}$ is a random vector of length $|\mathcal{U}|$ with elements in $\mathcal{X}$ and $q(t)$ is the probability distribution of T. The above result is proved directly in [20], but next simple argument shows that it follows again from Shannon's result, so that no proof is actually needed.

We can consider the CSIR V_n as an additional channel output. Then, the channel of Fig. 1 is equivalent to the channel of Fig. 2, with state U_n, output (Y_n, V_n) and conditional output probability

$$\begin{aligned} p'(y, v|x, u) &= \sum_s \Pr(y, v|x, u, s) \Pr(s|x, u) \\ &= \sum_s p(y|x, s)\omega(s, u, v)/p(u) \end{aligned} \tag{4}$$

where $p(u) = \sum_{s,v} \omega(s, u, v)$. The channel of Fig. 2 is clearly of the type studied by Shannon, with perfect CSIT and no CSIR. Its capacity is given by $\max_{q(t)} I(T; Y, V)$, but since V_n is independent on T_n we have $I(T;V) = 0$, so that (3) follows immediately.

Optimal codes for the channel of Fig. 1 are constructed over an extended input alphabet $\mathcal{X}^{|\mathcal{U}|}$, or equivalently, are codes defined over the alphabet of functions $\mathcal{U} \to \mathcal{X}$, designated sometimes as *strategy letters* [20]. This might pose some conceptual and practical problems for code construction, especially for large (infinite in the limit) $|\mathcal{U}|$. Nevertheless, optimal codes can be constructed directly over the input alphabet $\mathcal{X}$ in the following special case:

Proposition 1. *Let $U_n = g(V_n)$, where $g(\cdot)$ is a deterministic function $\mathcal{V} \to \mathcal{U}$. Then, the channel capacity is given by*

$$C = \sum_u p(u) \max_{q(x|u)} I(X;Y|V,u) \tag{5}$$

Proof. In order to prove (5) it is sufficient to show that

$$\max_{q(t)} I(T;Y|V) \leq \sum_u p(u) \max_{q(x|u)} I(X;Y|V,u) \tag{6}$$

Since the RHS of (6) corresponds to one possible assignment of the probability of T_n (namely, where the u-th component of T_n is distributed independently of the other components according to $q(x|u)$), (5) follows.

In order to show (6), we write

$$\begin{aligned} I(T;Y|V) &= \sum_{u,v} \Pr(u,v) I(T;Y|v) \\ &= \sum_{u,v} \Pr(u,v) I(T;Y|v,g(v)) \\ &= \sum_u p(u) \sum_v \Pr(v|u) I(T;Y|v,u) \\ &\overset{a}{\leq} \sum_u p(u) \max_{q(x|u)} \sum_v \Pr(v|u) I(X;Y|v,u) \\ &= \sum_u p(u) \max_{q(x|u)} I(X;Y|V,u) \end{aligned} \tag{7}$$

where (a) follows from the fact that $x_n = t_n(u_n)$, so that if $U_n = u_n$ is given, any probability distribution of T_n induces a probability distribution of X_n. □

Comment. It is easy to show that the capacity of Proposition 1 can be achieved by a multiplexed multiple codebook scheme [17, 22]. For each value $u \in \mathcal{U}$, a codebook of length $p(u)N$ and rate slightly less than $I(X;Y|V,u)$ is generated i.i.d. according to the probability distribution $q(x|u)$. For the message w, a set of $|\mathcal{U}|$ code words is selected, one for each codebook. At time n, if $U_n = u$ the transmitter sends the first not yet transmitted symbol of the u-th code word. Then, the code words are multiplexed according to the CSIT sequence U_1^N. If $g(\cdot)$ is deterministic, the receiver can demultiplex the received sequence before decoding since it can perfectly recover U_1^N from V_1^N. After demultiplexing, the $|\mathcal{U}|$ code words are independently decoded.

If $S_n \to V_n \to U_n$ is a Markov chain but U_n is a random function of V_n, the derivation above is no longer valid and the general coding technique based on vectors $t \in \mathcal{X}^{|\mathcal{U}|}$ must be considered. Intuitively, we see that if $g(\cdot)$ is not deterministic, the decoder is not able to demultiplex correctly the received sequence and the multiplexed multiple codebook scheme cannot be applied in a straightforward way.

The case where U_n is a deterministic function of V_n has an interesting practical application. Namely, it describes a situation where the CSIT is obtained via an error-free low-rate feedback channel from the receiver to the transmitter. [1] For example, V_n might be an accurate measure of the channel state obtained from the received signal. Then, the receiver instructs the transmitter by sending a command $U_n = g(V_n)$, where $g(\cdot)$ is some quantization function in order to reduce the rate of the feedback link.

A very interesting related problem is to maximize the capacity (5) under a constraint on the entropy of the feedback signal, namely, under $H(U) \leq R_f$, where R_f is the rate of the feedback link. We leave this problem for future investigation, and in Example 1 of Section 5 we show that by restricting $g(\cdot)$ to belong to some particular class of functions, the maximum capacity may not correspond to the maximum of $H(U)$.

3 States with memory

In general, the state, CSIT and CSIR processes $\{S_n\}, \{U_n\}, \{V_n\}$ are defined by a sequence of finite-dimensional joint distributions

$$\mathbf{\Omega} = \left\{\omega^{(N)}(s_1^N, u_1^N, v_1^N)\right\}_{N=1}^{\infty}$$

with the only requirement that (S_n, U_n, V_n) is independent of the past channel inputs X_1^{n-1}.

The case of states with memory encompasses also information-unstable channels. Then it should be treated in the general framework of [29]. By following again Shannon's approach, we can consider a new channel without CSIT, whose n-th input symbol is the random vector $T_n \in \mathcal{X}^{|\mathcal{U}|^n}$ and n-th output symbol is the pair (Y_n, V_n). A code for this channel is a set

$$\left\{t_1^N(w) = (t_1(w), \ldots, t_N(w)) \ : \ w \in \mathcal{W}, \ t_n(w) \in \mathcal{X}^{|\mathcal{U}|^n}\right\}$$

of $|\mathcal{W}|$ code words, each formed by concatenating N vectors over $\mathcal{X}$ such that the n-th vector of each code word has length $|\mathcal{U}|^n$. For a given source message $w \in \mathcal{W}$, the code word $t_1^N(w)$ is selected. At each time n, the channel input is given by $x_n = t_n(w, u_1^n)$, i.e., by the u_1^n-th component of the vector $t_n(w)$. [2]

[1] Low-rate feedback links are already implemented in many standards for cellular wireless systems [23].

[2] There is an obvious one-to-one correspondence between sequences $u_1^n \in \mathcal{U}^n$ and the integers from 1 to $|\mathcal{U}|^n$. Then, with a slight abuse of terminology, we indicate as the u_1^n-th component of the vector $t_n \in \mathcal{X}^{|\mathcal{U}|^n}$ the component whose index is the integer corresponding to u_1^n.

Clearly, the original channel and the new channel are completely equivalent both in terms of capacity and in terms of optimal encoding and decoding schemes. Since $\{S_n\}, \{\mathcal{U}_n\}, \{\mathcal{V}_n\}$ are independent of $\{T_n\}$, the N-th order channel transition probability of the new channel is immediately obtained as

$$P(y_1^N, v_1^N | t_1^N) = \sum_{s_1^N, u_1^N} \left[\prod_{n=1}^{N} p(y_n | x_n = t_n(u_1^n), s_n) \right] \omega^{(N)}(s_1^N, u_1^N, v_1^N) \tag{8}$$

An input process for the new channel is defined by a sequence of finite-dimensional distributions

$$\mathbf{T} = \left\{ \eta^{(N)}(t_1^N) \right\}_{N=1}^{\infty}$$

Let $\mathbf{Y}, \mathbf{V}$ denote the sequence of finite-dimensional output distributions induced by $\mathbf{T}$ and by the channel transition probability. The inf-information rate $\underline{\mathbf{I}}(\mathbf{T}; \mathbf{Y}, \mathbf{V})$ is defined as the *lim-inf in probability* [29] for $N \to \infty$ of the normalized information density

$$i_N(T_1^N; Y_1^N, V_1^N) = \frac{1}{N} \log \frac{P(Y_1^N, V_1^N | T_1^N)}{\Pr(Y_1^N, V_1^N)}$$

Then, from the general capacity formula given of [29], we can write:

$$C = \sup_{\mathbf{T}} \underline{\mathbf{I}}(\mathbf{T}; \mathbf{Y}, \mathbf{V}) \tag{9}$$

By summing $P(y_1^N, v_1^N | t_1^N)$ over all $y_1^N \in \mathcal{Y}^N$, we notice that the output V_1^N does not depend on the input T_1^N. Therefore, we can also write [27]

$$C = \sup_{\mathbf{T}} \underline{\mathbf{I}}(\mathbf{T}; \mathbf{Y} | \mathbf{V}) \tag{10}$$

where $\underline{\mathbf{I}}(\mathbf{T}; \mathbf{Y} | \mathbf{V})$ is the lim-inf in probability of the normalized conditional information density

$$i_N(T_1^N; Y_1^N | V_1^N) = \frac{1}{N} \log \frac{\Pr(Y_1^N | T_1^N, V_1^N)}{\Pr(Y_1^N | V_1^N)} \tag{11}$$

The above formula does not tell much in terms of practical coding and decoding schemes. However, by adding some constraints on the state and CSI signals, a simple single-letter capacity formula can still be found:

Proposition 2. *Assume: i) perfect CSIR ($V_n = S_n$); ii) deterministic CSIT (i.e., $U_n = g_n(S_1^n)$ with $g_n : \mathcal{S}^n \to \mathcal{U}$ deterministic); iii) that $\Pr(S_n | U_1^n) = \Pr(S_n | U_n)$; iv) that $\{S_n\}$ and $\{U_n\}$ are jointly stationary and ergodic. Then*

$$C = \sum_{u} p(u) \max_{q(x|u)} I(X; Y | S, u) \tag{12}$$

where $p(u)$ *is the first-order distribution of* U_n.

Proof. The achievability of (12) is easily established by appropriately choosing the input process $\mathbf{T}$. For all N, we consider product input distributions

$$\eta^{(N)}(t_1^N) = \prod_{n=1}^N \eta_n^{(N)}(t_n) \tag{13}$$

Moreover, we choose $\eta_n^{(N)}(t_n)$ such that its u_1^n-th marginal, given by

$$\Pr(T_n(u_1^n) = x) = \sum_{\substack{t_n \in \mathcal{X}^{|\mathcal{U}|^n} \\ t_n(u_1^n) = x}} \eta_n^{(N)}(t_n) \tag{14}$$

depends only on u_n and is independent of n and of u_1^{n-1}, i.e., $\Pr(T_n(u_1^n) = x) = q(x|u_n)$, where $q(\cdot|\cdot)$ does not depend on n.

Under the hypotheses of perfect CSIR and deterministic CSIT, we have

$$\Pr(y_1^N | t_1^N, s_1^N) = \prod_{n=1}^N p(y_n | x_n = t_n(u_1^n), s_n)$$

Moreover, for the product input distributions defined by (13) and (14), we have that also the conditional output distribution has a product form,

$$\begin{aligned} \Pr(y_1^N | s_1^N) &= \sum_{t_1^N} \prod_{n=1}^N p(y_n | x_n = t_n(u_1^n), s_n) \eta_n^{(N)}(t_n) \\ &= \prod_{n=1}^N \sum_{t_n} p(y_n | x_n = t_n(u_1^n), s_n) \eta_n^{(N)}(t_n) \\ &= \prod_{n=1}^N \sum_{x_n} p(y_n | x_n, s_n) q(x_n | u_n) \end{aligned} \tag{15}$$

Then, the normalized information density (11) is given by

$$i_N(T_1^N; Y_1^N | V_1^N) = \frac{1}{N} \sum_{n=1}^N \log \frac{p(Y_n | X_n, S_n)}{\Pr(Y_n | S_n)} \tag{16}$$

and, because of the joint ergodicity of $\{S_n\}$ and $\{U_n\}$, the above sample mean converges in probability, as $N \to \infty$, to the expectation

$$I(X; Y | S, U) = \mathrm{E}\left[\log \frac{p(Y|X,S)}{\Pr(Y|S)}\right]$$

where $(X, Y, S, U) \sim p(y|x,s) q(x|u) \Pr(s,u)$. By choosing $q(x|u)$ such that, for all $u \in \mathcal{U}$, it maximizes

$$I(X; Y | S, u) = \mathrm{E}\left[\log \frac{p(Y|X,S)}{\Pr(Y|S)} \middle| U = u\right]$$

we obtain that (12) is achievable.

For the converse, from Fano's inequality [28] we can write

$$H(W|\widehat{W}) \leq P_e \log|W| + h(P_e) = N\epsilon_N \tag{17}$$

where W and $\widehat{W}$ are the transmitted and the decoded messages and where $P_e = \Pr(W \neq \widehat{W})$. The decoder with perfect CSI is a mapping $\phi : (Y_1^N, S_1^N) \mapsto \widehat{W}$, then we have

$$\begin{aligned} H(W|\widehat{W}) &= H(W|\phi(Y_1^N, S_1^N)) \\ &\geq H(W|Y_1^N, S_1^N) \\ &= H(W|S_1^N) - I(W; Y_1^N|S_1^N) \\ &= NR - I(W; Y_1^N|S_1^N) \end{aligned} \tag{18}$$

By combining (17) and (18) we obtain

$$R \leq \frac{1}{N} I(W; Y_1^N|S_1^N) + \epsilon_N \tag{19}$$

where $\epsilon_N \to 0$ as $N \to \infty$. Then we have

$$\begin{aligned} I(W; Y_1^N|S_1^N) &= \sum_{n=1}^N I(W; Y_n|Y_1^{n-1}, S_1^N) \\ &= \sum_{n=1}^N H(Y_n|Y_1^{n-1}, S_1^N) - H(Y_n|Y_1^{n-1}, S_1^N, W) \\ &\stackrel{a}{\leq} \sum_{n=1}^N H(Y_n|S_n, U_1^n) - H(Y_n|Y_1^{n-1}, S_1^N, U_1^n, W, T_n) \\ &\stackrel{b}{\leq} \sum_{n=1}^N H(Y_n|S_n, U_1^n) - H(Y_n|S_n, U_1^n, X_n = T_n(U_1^n)) \\ &= \sum_{n=1}^N I(X_n; Y_n|S_n, U_1^n) \end{aligned} \tag{20}$$

where (a) follows from the deterministic CSIT assumption and (b) follows from the fact that, for given $U_1^n = u_1^n$, any probability distribution of T_n induces a probability distribution $q(x_n|u_1^n)$ of X_n.

Now, we let $\mathbf{U}_n = U_1^{n-1}$ and write

$$\begin{aligned} I(X_n; Y_n|S_n, U_1^n) &= I(X_n; Y_n|S_n, U_n, \mathbf{U}_n) \\ &= \sum_{u_n, s_n, \mathbf{u}_n} \Pr(s_n|u_n, \mathbf{u}_n) \Pr(\mathbf{u}_n|u_n) p(u_n) I(X_n; Y_n|s_n, u_n, \mathbf{u}_n) \\ &\stackrel{a}{=} \sum_{u_n, s_n} p(u_n) \Pr(s_n|u_n) \sum_{\mathbf{u}_n} \Pr(\mathbf{u}_n|u_n) I(X_n; Y_n|s_n, u_n, \mathbf{u}_n) \end{aligned}$$

$$
\begin{aligned}
&\stackrel{b}{\leq} \sum_{u_n,s_n} p(u_n)\Pr(s_n|u_n) I(\overline{X}_n;Y_n|s_n,u_n) \\
&= \sum_{u_n} p(u_n) I(\overline{X}_n;Y_n|S_n,u_n)
\end{aligned}
\tag{21}
$$

where (a) follows from hypothesis (iii) and (b) follows from the concavity of mutual information with respect to the input distribution, and where $\overline{X}_n$ denotes an input distributed according to $q(x_n|u_n)$, defined by

$$
q(x_n|u_n) = \sum_{\mathbf{u}_n} q(x_n|u_n,\mathbf{u}_n)\Pr(\mathbf{u}_n|u_n)
$$

Since U_n is stationary and ergodic, the last line of (21) does not depend on n.

The mutual information in the last line of (21) can be maximized by choosing, for each $u_n \in \mathcal{U}$, the input distribution $q(x|u)$ maximizing $I(X;Y|S,u)$. Then, by using (21) and (20) in (19), we obtain the converse as desired. □

Comment. The above proposition has some interesting particular cases. With perfect transmitter and receiver CSI (i.e., $U_n = g_n(S_1^n) = S_n$), the capacity is given by

$$
C = \sum_s p(s) \max_{q(x|s)} I(X;Y|s)
$$

which is the same expression given in [7] for a compound channel with memoryless state. Then, as it is well-known, perfect CSI makes the cases of i.i.d. states and of states with memory equivalent. [3] Also, with perfect CSI, interleaving does not incur any loss of optimality. If U_n is a d-step delayed version of S_n, namely,

$$
U_n = g_n(S_1^n) = \begin{cases} 0 & \text{for} \quad 1 \leq n \leq d \\ S_{n-d} & \text{for} \quad n \geq d+1 \end{cases}
$$

and $\{S_n\}$ is Markov, we find the result of [22]. Given the similarity of (5) and (12), it is immediate to show that a multiplexed multiple codebook scheme can achieve (12) (see the achievability part of [22]). Finally, under mild regularity conditions which guarantee the convergence of the finite alphabet result when the alphabet cardinality is taken to infinity, another interesting case is when S_n is a Gaussian process and $U_n = g_n(S_1^{n-d})$ is its Minimum Mean-Square Error (MMSE) estimate, based on the past measurements S_1^{n-d}. In this case we can write $S_n = U_n + E_n$, where the prediction error E_n is orthogonal to all functions of S_1^{n-d}. Then, S_n is in fact independent of U_1^{n-1} given U_n, as required by Proposition 2.

[3]This conclusion does not hold if a constraint on the transmission delay is taken into account. In this case, the so called delay-limited capacity [30, 31, 32] and/or the information outage probability [13] should be considered, since ergodicity or, more in general, information stability cannot be used.

4 AWGN channel with fading

In this section we consider the case of a real scalar AWGN channel with fading given by

$$Y_n = \sqrt{S_n} X_n + Z_n \tag{22}$$

where $S_n \in \mathbb{R}_+$ is assumed to be stationary and ergodic and $Z_n \sim \mathcal{N}(0,1)$ i.i.d.. The CSIT U_n is a noisy estimate of S_n. We assume that S_n is independent of U_1^{n-1} given U_n and that the receiver has perfect knowledge of the conditional *received* average signal-to-noise ratio (SNR)

$$\begin{aligned} V_n &= \mathrm{E}\left[\left. \frac{|\mathrm{E}[Y_n|X_n, S_n]|^2}{\mathrm{var}(Y_n|X_n, S_n)} \right| S_n, U_1^n \right] \\ &= S_n \mathrm{E}[|X_n|^2 | U_1^n] \end{aligned} \tag{23}$$

for all n (notice that, in general, the conditional second-order moment of the input $\mathrm{E}[|X_n|^2|U_1^n]$ depends on n because of the CSIT sequence U_1^n). As before, $\{S_n\}$ and $\{U_n\}$ are assumed jointly asymptotically ergodic and stationary. An average input power constraint $\mathrm{E}[|X_n|^2] \leq \mathcal{P}$ is imposed. [4]

This channel is a particular case of the model of Fig. 1 (provided that, under mild regularity conditions, it can be extended to the case of continuous alphabets), where the dependence of the CSIR signal V_n on S_n and on U_1^n is explicitly given by (23). This model applies for example to the case of time-division duplex (TDD) [23], with frequency non-selective block-fading [13, 14], assumed to be constant over each TDMA slot. The transmitter estimates the fading state on the current slot from the measurements in the previous slots of a pilot signal inserted in the reverse link. Then, the CSIT is a sequence of *predicted* fading states. The receiver measures the received SNR over the current frame directly from the received signal. It is reasonable to assume that the CSIR quality is very good (almost perfect), since no prediction is needed.

In general, U_n is not known explicitly by the receiver. Then, error-free demultiplexing of the received sequence as required by the multiplexed multiple codebook scheme of [17] is not possible. Also, since U_n is not in general a deterministic function of V_n, Propositions 1 and 2 cannot be extended directly to this case. Nevertheless, we have the following:

Proposition 3. *The AWGN channel with fading described by (22), with state S_n, CSIT U_n and CSIR $V_n = S_n \mathrm{E}[|X_n|^2|U_1^n]$, with the assumption that $\Pr(S_n|U_1^n) = \Pr(S_n|U_n)$ and subject to the input*

[4] The results for this real model can be immediately translated into results for the circularly-symmetric complex model with average energy per complex input symbol E_s and one-sided noise power spectral density N_0 by letting $\mathcal{P} = E_s/N_0$ and by doubling the information rates (expressed in nat per complex symbol). The input power constraint should be regarded as an average *transmit* SNR constraint.

power constraint $\mathrm{E}[|X_n|^2] \leq \mathcal{P}$, *has capacity given by*

$$C = \max_{\gamma} \mathrm{E}\left[\frac{1}{2}\log(1 + S\gamma(U))\right] \tag{24}$$

where expectation is with respect to the first-order joint distribution of S_n *and* U_n*, and the maximization is over the power allocation functions* $\gamma : \mathcal{U} \to \mathbb{R}_+$ *such that* $\mathrm{E}[\gamma(U)] \leq \mathcal{P}$.

Proof. We cannot use the multiplexed multiple codebook scheme since the decoder does not know U_n exactly. In order to show achievability, we construct a new channel and we show that it has at least the same capacity of the original one. The new channel is again a real scalar AWGN channel with fading $V_n \in \mathbb{R}_+$, input T_n and output $Y_n = \sqrt{V_n}T_n + Z_n$, with $Z_n \sim \mathcal{N}(0,1)$. The fading is defined by $V_n = S_n\gamma(U_n)$ where $\gamma(\cdot)$ is a given time-invariant deterministic function $\mathcal{U} \to \mathbb{R}_+$, such that $\mathrm{E}[\gamma(U_n)] \leq \mathcal{P}$. The new channel has no CSIT, perfect CSIR (i.e., V_n is known to the receiver), and input power constraint $\mathrm{E}[|T_n|^2] = 1$.

For this channel, we consider a conventional (i.e., constant-rate and constant-power) encoder, with codebook of rate R generated with i.i.d. components according a distribution $q(t)$ such that $\mathrm{E}[T_n] = 0$ and $\mathrm{E}[|T_n|^2] = 1$. By letting $q(t) = \mathcal{N}(0,1)$ and by conditioning on V_1^N, we have that Y_1^N is Gaussian with conditionally independent components $Y_n \sim \mathcal{N}(0, 1 + V_n)$. The corresponding normalized conditional information density is given by

$$\frac{1}{N} i_N(T_1^N; Y_1^N | V_1^N) = \frac{1}{N}\sum_{n=1}^{N} \frac{1}{2}\left[\log(1 + V_n) - |Z_n|^2 + \frac{|Y_n|^2}{1 + V_n}\right] \tag{25}$$

Both $\{V_n\}$ and $\{Y_n\}$ are stationary and ergodic, because of the joint stationarity and ergodicity of $\{S_n\}$ and $\{U_n\}$ and since $\gamma(\cdot)$ is time-invariant and $\{T_n\}$ is i.i.d.. Then, the above sample mean converges in probability, as $N \to \infty$, to the expectation

$$\mathrm{E}\left[\frac{1}{2}\log(1 + V)\right]$$

(notice that $\mathrm{E}[|Z_n|^2] = \mathrm{E}[|Y_n|^2/(1 + V_n)] = 1$). By substituting $V = S\gamma(U)$ and by maximizing over all functions $\gamma(\cdot)$ such that $\mathrm{E}[\gamma(U)] \leq \mathcal{P}$, we get that (24) is achievable.

For the converse, we go back to the original channel. From Fano's inequality, by recalling that the decoder with CSI V_n is a mapping $\phi : (Y_1^N, V_1^N) \mapsto \widehat{W}$ and by repeating the steps in (18), we obtain

$$R \leq \frac{1}{N} I(W; Y_1^N | V_1^N) + \epsilon_N \tag{26}$$

where $\epsilon_N \to 0$ as $N \to \infty$. Then we have

$$I(W; Y_1^N | V_1^N) \quad = \quad \sum_{n=1}^{N} I(W; Y_n | V_1^N, Y_1^{n-1})$$

$$
\begin{aligned}
&= \sum_{n=1}^{N} h(Y_n|V_1^N, Y_1^{n-1}) - h(Y_n|V_1^N, Y_1^{n-1}, W) \\
&\leq \sum_{n=1}^{N} h(Y_n|V_n) - h(Y_n|V_1^N, Y_1^{n-1}, W, X_n, S_n) \\
&= \sum_{n=1}^{N} h(Y_n|V_n) - h(Y_n|X_n, S_n) \\
&\stackrel{a}{\leq} \sum_{n=1}^{N} \mathrm{E}\left[\frac{1}{2}\log(1+V_n)\right] \qquad (27)
\end{aligned}
$$

where (a) follows from the fact that $h(Y_n|X_n, S_n) = h(Z_n) = \frac{1}{2}\log 2e\pi$ and that $h(Y_n|V_n) \leq \mathrm{E}[\frac{1}{2}\log(2e\pi(1+V_n))]$. The above upper bound to mutual information is achieved if X_1^N is a sequence with zero-mean Gaussian components X_n, statistically independent conditionally on U_1^N. Since a Gaussian distribution is determined only by its mean and covariance, and the mean is fixed to zero, without loss of generality we can write

$$
X_n = \sqrt{g_n(U_1^n)}T_n
$$

where T_n is i.i.d. $\sim \mathcal{N}(0,1)$. Then, we need only to prove that under the assumption that S_n is independent of U_1^{n-1} given U_n,

$$
\frac{1}{N}\sum_{n=1}^{N} \mathrm{E}\left[\frac{1}{2}\log(1+S_n g_n(U_1^n))\right]
$$

is maximized by $g_n(U_1^n) = \gamma(U_n)$, a time invariant function of U_n only. We have

$$
\begin{aligned}
\mathrm{E}\left[\frac{1}{2}\log(1+S_n g_n(U_1^n))\right] &= \mathrm{E}\left[\mathrm{E}\left[\frac{1}{2}\log(1+S_n g_n(U_1^n))\middle| S_n, U_n\right]\right] \\
&\leq \mathrm{E}\left[\frac{1}{2}\log(1+\mathrm{E}[S_n g_n(U_1^n)|S_n, U_n])\right] \\
&= E\left[\frac{1}{2}\log(1+S_n \mathrm{E}[g_n(U_1^n)|S_n, U_n])\right] \\
&= E\left[\frac{1}{2}\log(1+S_n \mathrm{E}[g_n(U_1^n)|U_n])\right] \\
&= E\left[\frac{1}{2}\log(1+S_n \gamma_n(U_n))\right] \qquad (28)
\end{aligned}
$$

where we have used Jensen's inequality, the fact that $\Pr(U_1^{n-1}|S_n, U_n) = \Pr(U_1^{n-1}|U_n)$ and we have defined $\gamma_n(U_n) = \mathrm{E}[g_n(U_1^n)|U_n]$. Now, since $\{S_n\}$ and $\{U_n\}$ are jointly ergodic and stationary, they have a stationary first-order joint distribution $p(s,u)$. By using the upper bound (28) into (27) we get

$$
\begin{aligned}
\frac{1}{N}I(W; Y_1^N|V_1^N) &\leq \frac{1}{N}\sum_{n=1}^{N} \mathrm{E}\left[\frac{1}{2}\log(1+S\gamma_n(U))\right] \\
&\leq \mathrm{E}\left[\frac{1}{2}\log\left(1+S\frac{1}{N}\sum_{n=1}^{N}\gamma_n(U)\right)\right]
\end{aligned}
$$

$$= \mathrm{E}\left[\frac{1}{2}\log\left(1+S\gamma(U)\right)\right] \tag{29}$$

where we have defined $\gamma(U) = \frac{1}{N}\sum_{n=1}^{N}\gamma_n(U)$. The above bound is achieved for every finite N by choosing $\gamma_n(\cdot)$ to be independent of n. Finally, by using (29) in (26) and letting $N \to \infty$ we get

$$R \leq E\left[\frac{1}{2}\log(1+S\gamma(U))\right]$$

where the RHS can be maximized over all the power allocation functions $\gamma(\cdot)$ such that $\mathrm{E}[\gamma(U)] \leq \mathcal{P}$. □

Comment. The above proof applies immediately to the cases of no CSIT and of perfect CSIT. In the first case (S_n statistically independent of U_n), by using again Jensen's inequality it is immediate to show that the constant function $\gamma(U_n) = \mathcal{P}$ is the capacity achieving power allocation, as argued in [13]. In the case of perfect CSIT ($S_n = U_n$) we obtain the result of [17] (see Section 4.1). In [17], achievability is proved by using the multiplexed multiple codebook scheme, assuming that the CSIR is $V_n = S_n$. Our result shows that this scheme (or, more in general, variable-rate variable-power schemes [33]) is not needed in order to achieve capacity. On the contrary, a simple conventional (i.e., constant-rate constant-power) Gaussian codebook is sufficient, provided that the code symbols are dynamically scaled by the appropriate power allocation function before transmission. We refer to this scheme as *single codebook with dynamic power allocation.* From a practical implementation point of view, we argue that the single codebook scheme with dynamic power allocation might be a simpler way to achieve the same capacity, without requiring multiple code books and variable-rate coding. Moreover, this scheme achieves capacity also in the case of non-perfect CSIT, under the conditions of Proposition 3, while the multiplexed multiple codebook scheme is inherently hard to implement, because the receiver is not able to demultiplex exactly the received sequence.

Results along this line are shown in [31, 32] and in the multiuser case in [34, 30, 35]. Nevertheless, for finite-complexity and limited decoding delay a combination of dynamic power allocation and variable coding schemes may be very effective, as demonstrated in [33].

4.1 Optimal power allocation

For simplicity we assume $\mathcal{U}$ discrete and $p(u) > 0$ for all $u \in \mathcal{U}$. The generalization to $\mathcal{U}$ continuous is rather straightforward and it is based on standard continuity arguments.

It is immediate to see that the optimal power allocation function $\gamma(u)$ must satisfy the power constraint with equality. Then, from the Lagrange multipliers and the Kuhn-Tucker conditions [28] we get that $\gamma(u)$ is the solution of (24) if and only if

$$\int_0^\infty \frac{s}{1+s\gamma(u)} p(s|u)\, ds \leq \lambda \tag{30}$$

for all $u \in \mathcal{U}$, with equality for all u such that $\gamma(u) > 0$, where λ is a given positive constant whose value is fixed in order to satisfy the power constraint with equality. Let $f_u(\gamma)$ denote the LHS of (30) as a function of $\gamma \geq 0$, parameterized by u. For given u, $f_u(\gamma)$ is a positive decreasing function of γ with maximum value $s(u) = \mathrm{E}[S|U=u]$, attained at $\gamma = 0$. For each u, $s(u)$ is assumed to be finite since it is physically reasonable to assume that an infinite conditional average channel gain occurs only with zero probability (the assumption $p(u) > 0$ rules out this case). Then, the solution $\gamma(u)$ is found in general as

$$\gamma(u) = \begin{cases} f_u^{-1}(\lambda) & \text{if } \; 0 < \lambda < s(u) \\ 0 & \text{if } \; \lambda \geq s(u) \end{cases} \tag{31}$$

The actual value of λ is determined by solving $\sum_u p(u)\gamma(u) = \mathcal{P}$. In practical computations we can parameterize both the average transmitted power $\mathcal{P}$ and the solution $\gamma(u)$ in terms of $\lambda \in [0, \max_u s(u)]$. Since $f_u^{-1}(\lambda)$ is decreasing in λ, $\mathcal{P}$ is also a decreasing function of λ. For a given λ (i.e., for a given $\mathcal{P}$), positive power is allocated only to the values $u \in \mathcal{U}$ such that $s(u) > \lambda$. In this sense, the optimal power allocation $\gamma(u)$ has a waterfilling nature, similar to the optimal power allocation in the case of perfect CSIT, found in [17]. It is immediate to see that $\gamma(u)$ in (31) coincides with the power allocation given in [17] in the case $p(s|u) = \delta(s-u)$, i.e., for perfect CSIT. In this case we have

$$\gamma_{\text{perf.}}(u) = \left[\frac{1}{\lambda} - \frac{1}{u}\right]_+$$

where $[\cdot]_+$ denotes the positive part.

5 Examples

In this section we consider some examples of AWGN channel with fading. Under the assumptions of the previous section, the determination of the channel capacity reduces to the computation of the solution (31) of the constrained maximization problem given by Proposition 3.

Example 1. Assume S_n i.i.d. uniformly distributed on $[0, A]$ and let U_n be the 1-bit quantization feedback information

$$U_n = \begin{cases} 0 & \text{if } \; S_n < a \\ 1 & \text{if } \; S_n \geq a \end{cases}$$

where $0 \leq a \leq A$ is a suitable quantization threshold. Moreover, assume that the receiver knows exactly the received SNR $V_n = S_n \mathrm{E}[|X_n|^2 | U_1^n]$, so that Proposition 3 applies. From (30) we find

$$\begin{aligned} f_0(\gamma) &= \frac{1}{a}\left[\frac{a}{\gamma(0)} - \frac{\log(1+a\gamma(0))}{\gamma(0)^2}\right] \\ f_1(\gamma) &= \frac{1}{A-a}\left[\frac{A-a}{\gamma(1)} - \frac{\log(1+A\gamma(1)) - \log(1+a\gamma(1))}{\gamma(1)^2}\right] \end{aligned} \quad (32)$$

Fig. 3 shows $f_0(\gamma)$ and $f_1(\gamma)$ for $a = 1$ and $A = 2$. For $0 \leq \lambda < a/2$ both $\gamma(0)$ and $\gamma(1)$ are positive. For $a/2 \leq \lambda < (A+a)/2$ only $\gamma(1)$ is positive and $\gamma(0)$ is zero. The value of λ is obtained by solving the constraint equation $a\gamma(0)/A + (A-a)\gamma(1)/A = \mathcal{P}$. If $a = 0$, the first inequality becomes irrelevant, while if $a = A$, the second inequality has always a positive solution, but the probability of transmitting with power $\gamma(1)$ is zero, so that the value of $\gamma(1)$ is irrelevant. Both these two extreme cases correspond to the constant power transmission case, where in the former $\gamma(1) = \mathcal{P}$ and in the latter $\gamma(0) = \mathcal{P}$. This agrees with the fact that for both $a = 0$ and $a = A$, the CSIT does not provide any information. Fig. 4 shows $\gamma(0)$ and $\gamma(1)$ vs. $\mathcal{P}$ in the case $a = 1$ and $A = 2$.

The resulting average capacity can be written as

$$\begin{aligned} C_{\mathrm{csi}}(\mathcal{P}) &= \frac{1}{2A}\left[\frac{\log(1+a\gamma(0))}{\gamma(0)} + a(\log(1+a\gamma(0)) - \log(1+a\gamma(1)) + \right. \\ &\left. + \frac{\log(1+A\gamma(1)) - \log(1+a\gamma(1))}{\gamma(1)} + A\log(1+A\gamma(1)) - A\right] \end{aligned}$$

This can be compared with the capacity in the case of no CSIT (constant power)

$$C_{\mathrm{const}}(\mathcal{P}) = \frac{1}{2A}\left[\frac{\log(1+A\mathcal{P})}{\mathcal{P}} + A\log(1+A\mathcal{P}) - A\right]$$

and with the capacity with perfect CSIT (i.e., when $U_n = S_n$). This is obtained from the general solution of [17] as

$$C_{\mathrm{perf}}(\mathcal{P}) = \frac{1}{2}(\log(A/\rho_0) + \rho_0/A - 1)$$

where ρ_0 is the solution of the constraint equation, that it this case can be written as

$$\frac{1}{A}\left[\frac{A-\rho_0}{\rho_0} - \log\frac{A}{\rho_0}\right] = \mathcal{P}$$

Finally, the capacity of an AWGN channel without fading and with the same average channel gain and transmitted power is given by $C_{\mathrm{awgn}}(\mathcal{P}) = \frac{1}{2}\log(1+A\mathcal{P}/2)$. Fig. 5 shows capacity vs. $\mathcal{P}$ for Example 1, with $A = 2$ and $a = 1$. For this fading statistics the 1-bit CSIT provides almost optimal performance, in fact C_{csi} and C_{perf} are very close. In general, CSIT provides a performance improvement over constant power transmission only in the low-SNR region (i.e., for low rates). In this example, for

$R = 0.1$ bit/symbol the gain over constant power transmission is about 1.7 dB while for higher rates it disappears.

Fig. 6 shows capacity versus the quantization threshold a, for different SNRs and $A = 2$. It is clearly visible that for low SNR the optimal threshold is close to A and for high SNR the optimal threshold moves closer to 0. This has a nice intuitive explanation: for low SNR, the transmitter cannot waste power, so that it must know when the channel is good to transmit. Then, the optimal threshold should be large, in order to reveal when the channel is "exceptionally" good. For large SNR, the transmitter has a lot of power and can compensate for poor channel gains, so that the threshold should be small, in order to reveal when the channel is "exceptionally" bad and needs to be compensated for. In general, both for high and for low SNR, the capacity is maximized for $H(U) < 1$ bit. Then, this example shows that by constraining $g(\cdot)$ to be in a particular class of functions, the solution to the capacity maximization problem under an entropy constraint for U may not be reached when the constraint is satisfied with equality, in general.

Example 2. Consider a Low-Earth-Orbit Satellite system and a mobile terminal in urban environment. Due to the terminal motion, the line-of-sight (LOS) path between the terminal and the satellite may be either present or blocked. A simplified channel state model is a process S_n which can be either constant (LOS present) or exponentially distributed (LOS blocked) [36], so that the channel is AWGN or Rayleigh, respectively. Assume that $V_n = S_n \mathrm{E}[|X_n|^2 | U_1^n]$ is known to the receiver and that $S_n = |U_n + (1 - U_n)G_n|^2$, where U_n is an ergodic process that takes on values 0 or 1 with given first-order probability $p(u)$ and G_n is a complex white Gaussian process with i.i.d. real and imaginary parts $\sim \mathcal{N}(0, A^2/2)$. Clearly, S_n is independent of U_1^{n-1} given U_n, so that Proposition 3 applies. Before entering the details of the calculation, we would like to point out that this case models the realistic scenario where the receiver has a very accurate (perfect) CSI while the transmitter knows only the statistics of the channel (AWGN or Rayleigh), which vary slowly due to the terminal motion, but ignores the actual value of the channel gain. Since the channel statistics changing rate is usually much slower than the signaling rate, a low rate feedback link that instructs the transmitter about the channel statistics can be implemented.

The conditional pdf $p(s|u)$ is given by $p(s|1) = \delta(s - 1)$ and $p(s|0) = \exp(-s/A)/A$. The solution of (30) in this case yields

$$\gamma(1) = \left[\frac{1}{\lambda} - 1\right]_+$$

and $\gamma(0)$ solution of

$$\frac{1}{\gamma(0)} - \frac{e^{1/(A\gamma(0))}}{A\gamma(0)^2} E_i(1, 1/(A\gamma(0))) = \lambda$$

for $\lambda < A$ and $\gamma(0) = 0$ for $\lambda(0) \geq A$, where $E_i(n,x) = \int_1^\infty t^{-n} e^{-xt} dt$ [37]. The capacity is given by

$$C_{\text{csi}}(\mathcal{P}) = p(1)\frac{1}{2}\log(1+\gamma(1)) + p(0)\frac{1}{2}e^{1/(A\gamma(0))} E_i(1, 1/(A\gamma(0)))$$

For comparison, the capacity with constant power transmission is

$$C_{\text{const}}(\mathcal{P}) = p(1)\frac{1}{2}\log(1+\mathcal{P}) + p(0)\frac{1}{2}e^{1/(A\mathcal{P})} E_i(1, 1/(A\mathcal{P}))$$

the capacity with perfect CSIT is

$$C_{\text{perf}}(\mathcal{P}) = \begin{cases} p(1)\frac{1}{2}\log(1/\rho_0) + p(0)\frac{1}{2}E_i(1,\rho_0/A) & \text{if} \quad \rho_0 \leq 1 \\ p(0)\frac{1}{2}E_i(1,\rho_0/A) & \text{if} \quad \rho_0 > 1 \end{cases}$$

where ρ_0 is the solution of the constraint equation

$$p(1)\left[\frac{1}{\rho_0} - 1\right]_+ + p(0)\left(\frac{e^{-\rho_0/A}}{\rho_0} - \frac{1}{A}E_i(1,\rho_0/A)\right) = \mathcal{P}$$

and the capacity of an AWGN with the same average channel gain and the same transmitted average power is

$$C_{\text{awgn}}(\mathcal{P}) = \frac{1}{2}\log(1 + (p(1) + Ap(0))\mathcal{P})$$

Fig. 7 and 8 show capacity vs. $\mathcal{P}$ for Example 2, in the case $A = 0.1$ with $p(0) = 0.5$ and $p(0) = 0.9$, respectively. The average channel gain is 0.55 in the first and 0.19 is the second case. For this channel model, the knowledge of the channel statistics at the transmitter provides large performance gains for low SNR. In the case of $p(0) = 0.5$, the gain with respect to constant power transmission is 2.5 dB at $R = 0.1$ and 1.9 dB at $R = 0.5$ bit/symbol. In the case of $p(0) = 0.9$, the gain with respect to constant power transmission is 5 dB at $R = 0.1$ and about 2.0 dB at $R = 0.5$ bit/symbol.

The intuition behind this result is that higher average rates are achievable if the transmitter sends high-rate and high-power "bursts" when the channel is "good" (LOS present), and basically turns transmission off when the channel is "bad" (LOS absent). This effect is more visible in the low-SNR region, since in this case power must be used more efficiently. This result may be appealing in a TDMA network for data transmission, where mobile terminals are likely to experience different channel conditions. In this setting, as shown in [35] in the case of perfect CSIT, a protocol that allocates resources according to the users' channel condition may provide substantial improvements in terms of total bandwidth efficiency.

Example 3. Consider a TDD system. Let the channel complex amplitude gain α_n be a stationary ergodic complex Gaussian process with independent real and imaginary parts $\sim \mathcal{N}(0, 1/2)$ and assume that the channel gain estimator at the transmitter, β_n, is the output of a MMSE linear prediction filter, whose input is the sequence of past noisy measurements of α_n. Then, α_n is independent of β_1^{n-1} given β_n and Proposition 3 applies. We let $S_n = |\alpha_n|^2$ (exponentially distributed with mean 1) and we let $U_n = |\beta_n|^2$. Let ϵ denote the estimation Mean-Square Error. Then, it is straightforward to show that the joint pdf of S_n and U_n is given by (see also [38, Appendix A])

$$p(s, u) = \frac{1}{\epsilon(1-\epsilon)} I_0\left(\frac{2\sqrt{su}}{\epsilon}\right) \exp\left(-\frac{s}{\epsilon} - \frac{u}{\epsilon(1-\epsilon)}\right)$$

Eq. (30) in this case becomes

$$\frac{1}{\gamma(u)}\left[1 - \frac{e^{-u/\epsilon}}{\epsilon} \mathcal{I}\left(\gamma(u), \frac{2\sqrt{u}}{\epsilon}, \frac{1}{\epsilon}\right)\right] \leq \lambda$$

where $\mathcal{I}(a, b, c)$ is defined as

$$\mathcal{I}(a, b, c) = \int_0^\infty \frac{1}{1 + ax} I_0(b\sqrt{x}) e^{-cx}\, dx \tag{33}$$

The conditional average capacity given U_n is given by

$$\int_0^\infty \frac{1}{2}\log(1 + s\gamma(u)) \frac{1}{\epsilon} I_0\left(\frac{2\sqrt{su}}{\epsilon}\right) \exp\left(-\frac{s}{\epsilon} - \frac{u}{\epsilon}\right)\, ds \tag{34}$$

Both these integrals are evaluated in Appendix A. The numerical calculation of $\mathcal{C}_{\rm csi}(\mathcal{P})$ has been carried out by discretizing the variable u, by solving for $\gamma(u)$ for each discrete value u and by averaging the resulting conditional average capacities over all u's. Since the channel is Rayleigh with average gain 1, the capacity with constant power transmission is given by [12] $\mathcal{C}_{\rm const}(\mathcal{P}) = \frac{1}{2} e^{1/\mathcal{P}} E_i(1, 1/\mathcal{P})$. The capacity with perfect CSIT is $\mathcal{C}_{\rm perf}(\mathcal{P}) = \frac{1}{2} E_i(1, \rho_0)$ where ρ_0 is the solution of

$$\frac{1}{\rho_0} e^{-\rho_0} - E_i(1, \rho_0) = \mathcal{P}$$

and the capacity of the AWGN channel with the same average gain and transmitted average power is $\mathcal{C}_{\rm awgn}(\mathcal{P}) = \frac{1}{2}\log(1 + \mathcal{P})$.

Fig 9 shows capacity vs. $\mathcal{P}$ for Example 3 for estimation errors $\epsilon = 0.1, 0.3, 0.5, 0.7$ and 0.9. As expected, $\mathcal{C}_{\rm csi}(\mathcal{P})$ converges to $\mathcal{C}_{\rm perf}(\mathcal{P})$ as $\epsilon \to 0$ and to $\mathcal{C}_{\rm const}(\mathcal{P})$ as $\epsilon \to 1$. In fact, $\epsilon = 1$ corresponds to the case where $U_n = |\beta_n|^2 = 0$, i.e., no CSIT. The gain with respect to constant power transmission at rate $R = 0.1$ is 2.0 dB for $\epsilon = 0.3$ and 2.5 dB for $\epsilon = 0.1$.

By relaxing the assumption that S_n is independent of U_1^{n-1} given U_n, we get that the rate given by Proposition 3 is still achievable, even if it might not be the channel capacity (in fact, the achievability in the proof of Proposition 3 does not make use of the above assumption). Then, calculations like the example above allow simple comparisons of CSI estimation techniques in terms of achievable mutual information versus estimation mean-square error.

6 Conclusions

In this paper we investigated the capacity of some time-varying channels with transmitter and receiver CSI. Conditions for simple optimal coding have been derived in the case of i.i.d. states and states with memory. A number of known results have been obtained as particular cases of this analysis. For the AWGN channel with fading, we found a general simple expression for capacity under some assumptions on the receiver and transmitter CSI. This holds also in the case of states with memory and yields as special cases the results of [17] (for perfect CSI), of [22] (for delayed perfect CSI) and the optimality of constant power transmission [13] (for no transmitter CSI and perfect receiver CSI). Some more intuition on optimal coding strategies for the AWGN fading channel have been given. In particular, we showed that variable-rate variable-power coding schemes are not needed to achieve the capacity of this channel, and a simple single codebook scheme with dynamic power allocation may be a more viable solution. Finally, a number of examples have been provided for the AWGN fading channel. These examples show that even with non-perfect CSI, the optimal power allocation is still of the waterfilling type. Hopefully, this can be extended to the more interesting case of a multiuser channel where time and bandwidth is allocated dynamically to the users in order to maximize their rate sum, as done in [35, 34, 30] in the case of perfect transmitter CSI.

A Evaluation of some definite integrals.

We want to evaluate $\mathcal{I}(a, b, c)$ defined by (33). By using the power series representation of $I_0(z)$ and the definite integral [37, 3.383.10, pag. 319], we can rewrite (33) as

$$\mathcal{I}(a, c, b) = \frac{1}{a} \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{b^2}{4a} \right)^k e^{c/a} \Gamma \left(-k, \frac{c}{a} \right)$$

where $\Gamma(\nu, x) = \int_x^\infty t^{\nu-1} e^{-t} dt$ is the incomplete Gamma function. Then, we use the identity

$$\Gamma(-k, x) = x^{-k} E_i(k+1, x) \quad (k \text{ positive integer})$$

and the recursion formula [37, pag. xxxiii]

$$E_i(k, x) = \frac{1}{k-1} \left(e^{-x} - x E_i(k-1, x) \right) \quad k > 1$$

to obtain

$$\begin{aligned}\mathcal{I}(a,b,c) &= \frac{1}{a}e^{c/a}J_0\left(\frac{b}{\sqrt{a}}\right)E_i\left(1,\frac{c}{a}\right)+\\ &+\frac{1}{a}\sum_{k=0}^{\infty}\frac{1}{k!}\left(\frac{-b^2}{4a}\right)^k\sum_{i=1}^{k}(i-1)!\left(\frac{-a}{c}\right)^i\end{aligned}$$

The infinite sum above does not seem to be amenable to a closed form formulation and it can be evaluated by truncation.

The evaluation of integral (34) is carried out in a similar way. First, we use integration by parts and we rewrite (34) as

$$\frac{1}{2}\int_0^{\infty}\frac{a}{1+ax}Q\left(\frac{b}{\sqrt{2c}},\sqrt{2cx}\right)\,dx$$

where $Q(\nu,\mu)$ is the Marcum Q-function [38, Appendix A] and where we let $a=\gamma(u)$, $b=2\sqrt{u}/\epsilon$ and $c=1/\epsilon$. Then, we use the expansion of the Q-function given by [38, Appendix A]

$$Q(\nu,\mu)=e^{-(\nu^2+\mu^2)/2}\sum_{k=0}^{\infty}\left(\frac{\mu}{\nu}\right)^k I_k(\nu\mu)$$

and we apply the same technique used before to all the terms of the summation. The final result is

$$\frac{1}{2}e^{-b^2/4c}\sum_{k=0}^{\infty}\left(\frac{-c}{a}\right)^k\left\{e^{c/a}\left(\frac{2\sqrt{a}}{b}\right)^k J_k\left(\frac{b}{\sqrt{a}}\right)E_i\left(1,\frac{c}{a}\right)+\sum_{\ell=0}^{\infty}\frac{1}{\ell!(k+\ell)!}\left(\frac{-b^2}{4a}\right)^{\ell}\sum_{i=1}^{k+\ell}(i-1)!\left(\frac{-a}{c}\right)^i\right\}$$

Acknowledgement

The authors wish to thank R. Zamir, U. Erez, A. Das and P. Narayan, and two anonymous reviewers for their helpful comments.

References

[1] C. Shannon, "Channels with side information at the transmitter," *IBM J. Res. & Dev.,* 2, pp. 289 – 293, 1958.

[2] F. Jelinek, "Indecomposable channels with side information at the transmitter," *Information and Control,* 8, pp. 36 – 55, 1965.

[3] F. Jelinek, "Determination of capacity achieving input probabilities for a class of finite state channels with side information," *Information and Control,* 9, pp. 101 – 129, 1966.

[4] E. Gilbert, "Capacity of a burst-noise channel," *Bell Syst. Tech. Journal,* Vol. 39, pp. 1253 – 1265, Sept. 1960.

[5] M. Mushkin and I. Bar-David, "Capacity and coding for the Gilbert-Elliot channel," *IEEE Trans. on Inform. Theory,* Vol. 35, No. 6, pp. 1277 – 1290, Nov. 1989.

[6] A. Goldsmith and P. Varaiya, "Capacity, mutual information and coding for finite-state Markov channels," *IEEE Trans. on Inform. Theory,* Vol. 42, No. 3, pp. 868 – 886, May 1996.

[7] J. Wolfovitz, *Coding theorems of information theory,* New York: Springer-Verlag, 1978.

[8] R. McEliece and W. Stark, "Channels with block interference," *IEEE Trans. on Inform. Theory,* Vol. 30, No. 1, pp. 44 – 53, Jan. 1984.

[9] J. Keifer and J. Wolfowitz, "Channels with arbitrarily varying channel probability functions," *Inform. Contr.,* Vol. 5, pp. 44 – 54, 1962.

[10] S. Gelfand and M. Pinsker, "Coding for channels with random parameters," *Prob. Contr. Inform. Theory,* Vol. 9, No. 1, pp. 19 – 31, 1980.

[11] R. Ahlswede, "Arbitrarily varying channels with states sequence known to the sender," *IEEE Trans. on Inform. Theory,* Vol. IT-32, No. 5, pp. 621 – 629, Sept. 1986.

[12] T. Ericsson, "A Gaussian channel with slow fading," *IEEE Trans. on Inform. Theory,* Vol. IT-16, No. 3, pp. 353 – 356, May 1970.

[13] L. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. on Veh. Tech.,* Vol. 43, No. 2, May 1994.

[14] G. Kaplan and S. Shamai (Shitz), "Error probabilities for the block-fading Gaussian channel," *A.E.U.,* Vol. 49, No. 4, pp. 192 – 205, 1995.

[15] M. Khansari and M. Vetterli, "Source coding and transmission of signals over time-varying channels with side information," *International Symposium on Information Theory, ISIT'95* p. 140, Whistler, Canada, Sept. 1995.

[16] J. Larrea-Arrieta and D. Tait, "Adaptive forward error control schemes in channels with side information at the transmitter," *International Symposium on Information Theory, ISIT'95* p. 218, Whistler, Canada, Sept. 1995.

[17] A. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," proceedings of the *International Conference on Communications, ICC '93,* pp. 600 – 604, Geneva, Switzerland, June 1993.

[18] M. Médard, A. Goldsmith, "Capacity of time-varying channels with channel side information," *International Symposium on Information Theory, ISIT '97*, pp. 372, Ulm, Germany, June 29 – July 4, 1997.

[19] I. Abou-Faycal, M. Trott and S. Shamai, "Capacity of discrete-time Rayleigh fading channels," *International Symposium on Information Theory, ISIT '97*, pp. 473, Ulm, Germany, June 29 – July 4, 1997.

[20] M. Salehi, "Capacity and coding for memories with real-time noisy defect information at encoder and decoder," *IEE Proceedings-I,* Vol. 139, No. 2, pp. 113 – 117, April 1992.

[21] T. Duman and M. Salehi, "Optimal quantization for finite-state channels," *IEEE Trans. on Inform. Theory,* Vol. 43, No. 2, pp. 758 – 765, March 1997.

[22] H. Viswanathan "Capacity of time-varying channels with delayed feedback," submitted to *IEEE Trans. on Inform. Theory,* April 1997.

[23] T. Rappaport, *Wireless Communications.* Englewood Cliffs, NJ: Prentice-Hall, 1996.

[24] A. Kusnetsov and B. Tsybakov, "Coding in a memory with defective cells," *Probl. Pered. Inform.* Vol. 10, No. 2, pp. 52 – 60, 1974.

[25] C. Heegard and A. El Gamal, "On the capacity of computer memory with defects," *IEEE Trans. on Inform. Theory,* Vol. IT-29, No. 5, pp. 731 – 739, Sept. 1983.

[26] U. Erez and R. Zamir, "Capacity and coding for symmetric channels with side information at the transmitter," presented at the *1998 IEEE International Symposium on Information Theory, ISIT '98,* MIT, Cambridge, MA USA, 16–21 Aug. 1998.

[27] A. Das and P. Narayan, "On the capacities of a class of finite-state channels with side information," proceedings of *CISS '98*, Pronceton, NJ, March 18 – 20, 1998.

[28] R. Gallager, *Information theory and reliable communication.* New York: Wiley, 1968.

[29] S. Verdú and T.S. Han, "A general formula for channel capacity," *IEEE Trans. on Inform. Theory,* Vol. 40, No. 4, pp. 1147 – 1157, July 1994.

[30] D. Tse and V. Hanly, "Multi-access fading channels: Part II, delay-limited capacities," to appear on *IEEE Trans. on Inform. Theory,* 1998.

[31] G. Caire, G. Taricco and E. Biglieri, "Information outage rate and power control," submitted to *IEEE Trans. on Inform. Theory,* Jan. 1998.

[32] G. Caire, G. Taricco and E. Biglieri, "Optimal power allocation for minimum outage rate in wireless communications," proceedings of *ICC '98,* Atlanta, GA USA, 7–11 Aug. 1997.

[33] A. Goldsmith and S. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. on Comm.,* Vol. 45, No. 10, pp. 1218 – 1230, Oct. 1997.

[34] D. Tse and V. Hanly, "Multi-access fading channels: Part I, polymatroid structure, optimal resource allocation and throughput capacities," proceedings of *1996 Information Theory Workshop*, pp. 87, Haifa, Israel, June 9 – 13, 1996, to appear on *IEEE Trans. on Inform. Theory,* 1998.

[35] R. Knopp, P. A. Humblet, "Information capacity and power-control in single-cell multiuser communications," proceedings of the *IEEE International Conference on Communications, ICC '95*, pp. 331 – 335, Seattle, Wa. June 18 – 22, 1995.

[36] E. Lutz, D. Cygan, M. Dippold F. Dolainsky, W. Papke "The land mobile satellite communication channel – Recording, statistics, and channel model," *IEEE Trans. on Vehic. Tech.,* Vol. 40, No. 2, May 1991.

[37] I. Gradshteyn and I. Ryzhik, *Tables of Integrals, Series, and Products,* New York: Academic Press, 1980.

[38] M. Schwartz, W. Bennet and S. Stein, *Communication Systems and Techniques,* New York: McGraw-Hill, 1966.

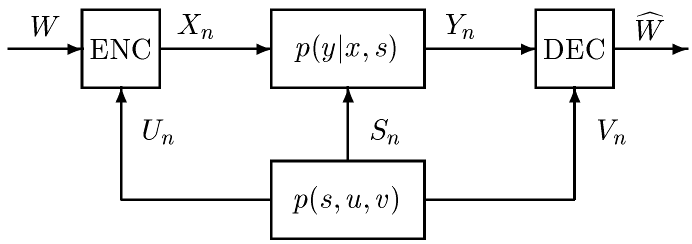

Figure 1: Block diagram of the channel with time-varying state, CSIT and CSIR.

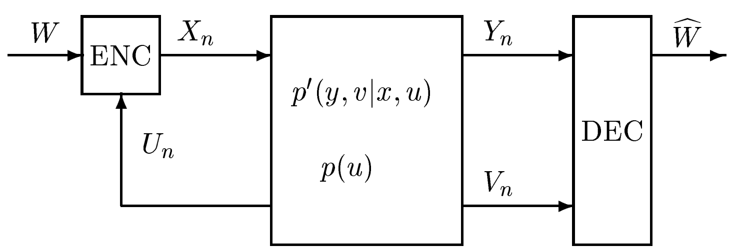

Figure 2: Block diagram of the equivalent channel with CSIT only and output (Y_n, V_n).

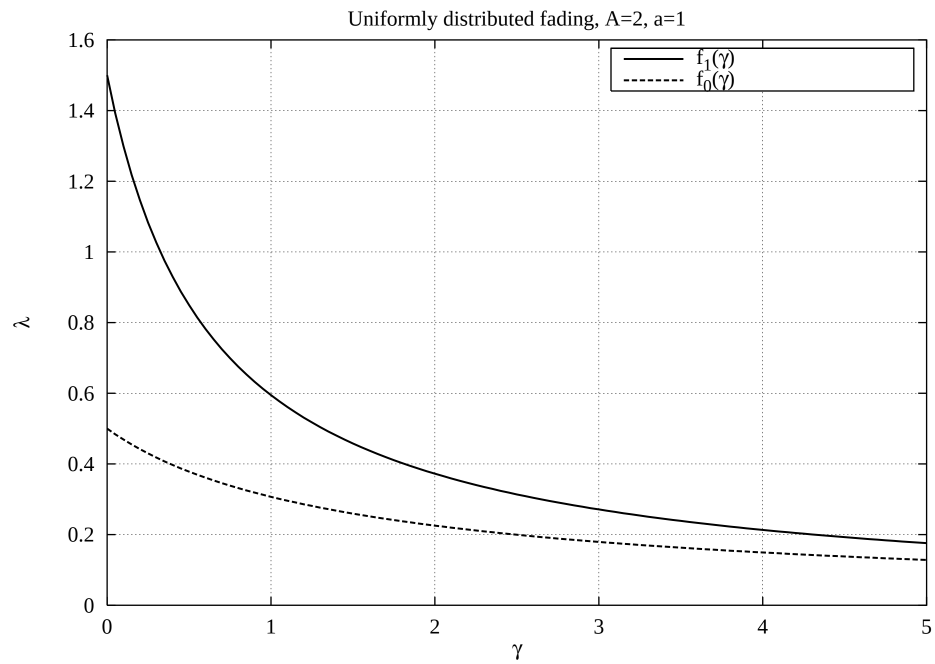

Figure 3: Functions $f_1(\gamma)$ and $f_0(\gamma)$ of Example 1.

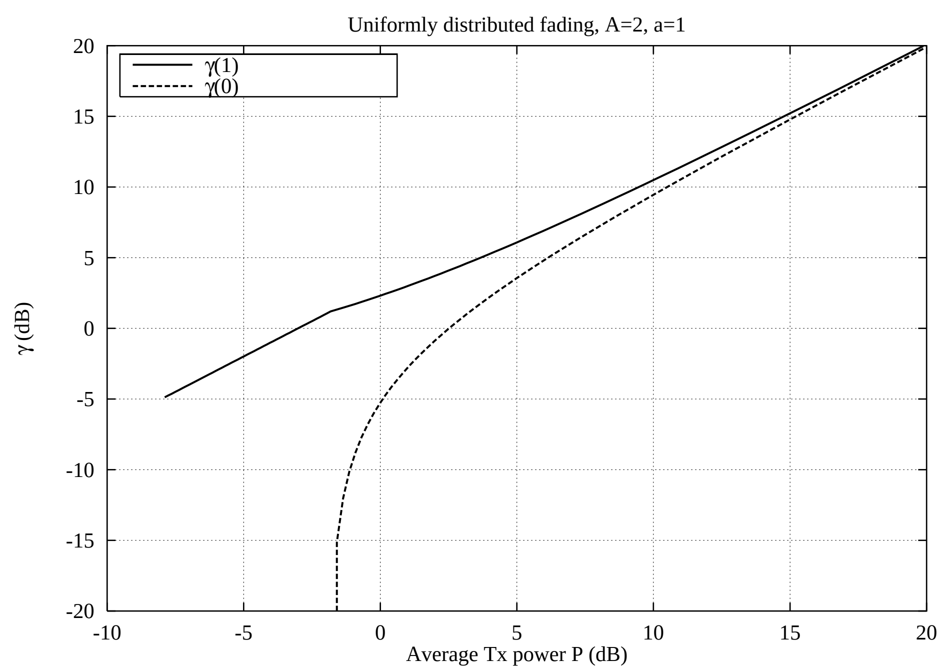

Figure 4: Solutions $\gamma(1)$ and $\gamma(0)$ of the constrained maximization problem of Example 1, plotted versus the average transmitter power $\mathcal{P}$.

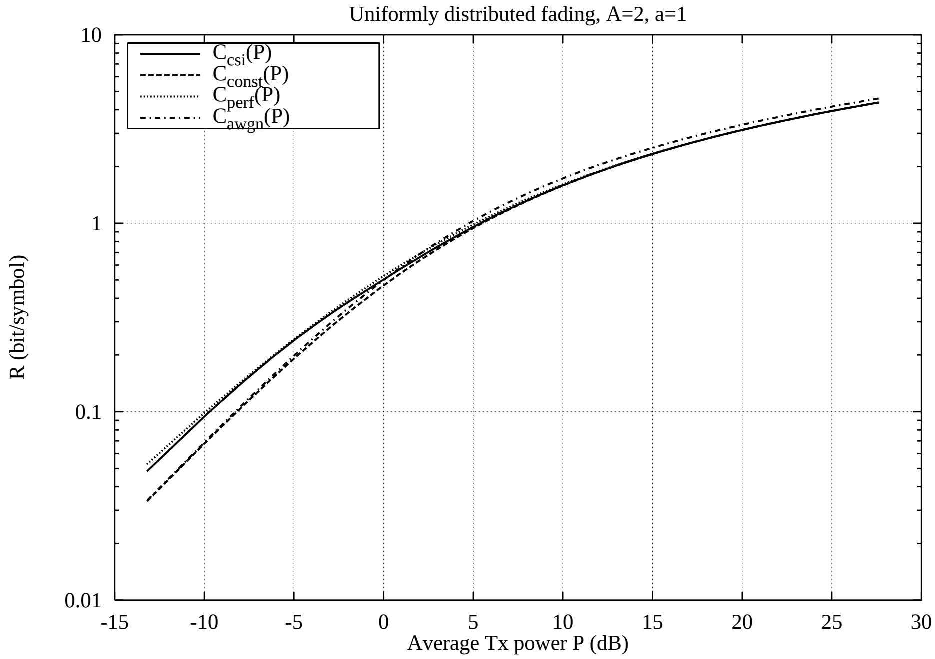

Figure 5: Capacity vs. $\mathcal{P}$ of Example 1, for $A = 2$ and $a = 1$.

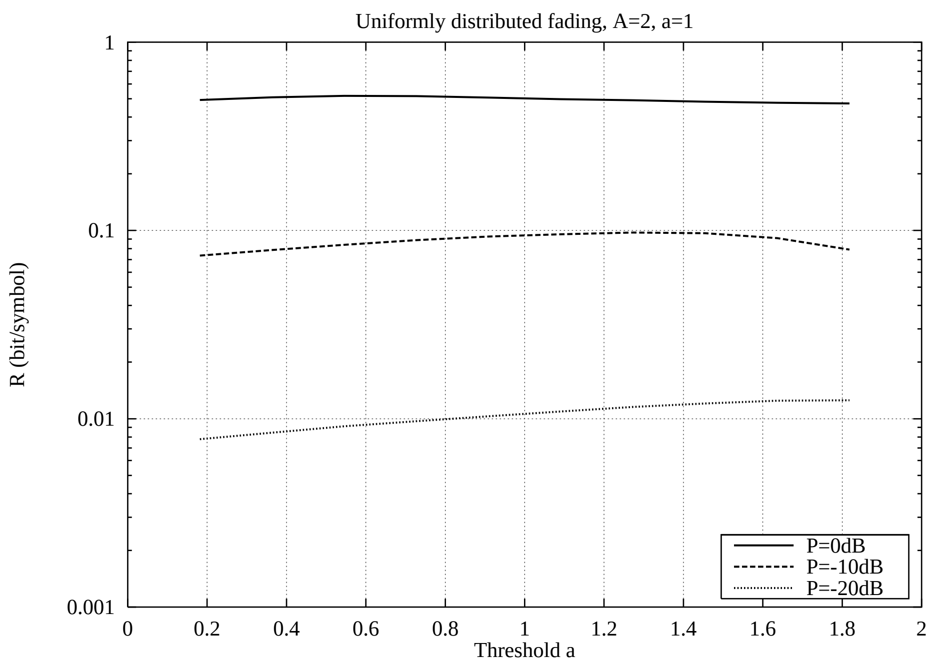

Figure 6: Capacity vs. a for Example 1, for $A = 2$ and SNR$= 0, -10$ and -20 dB.

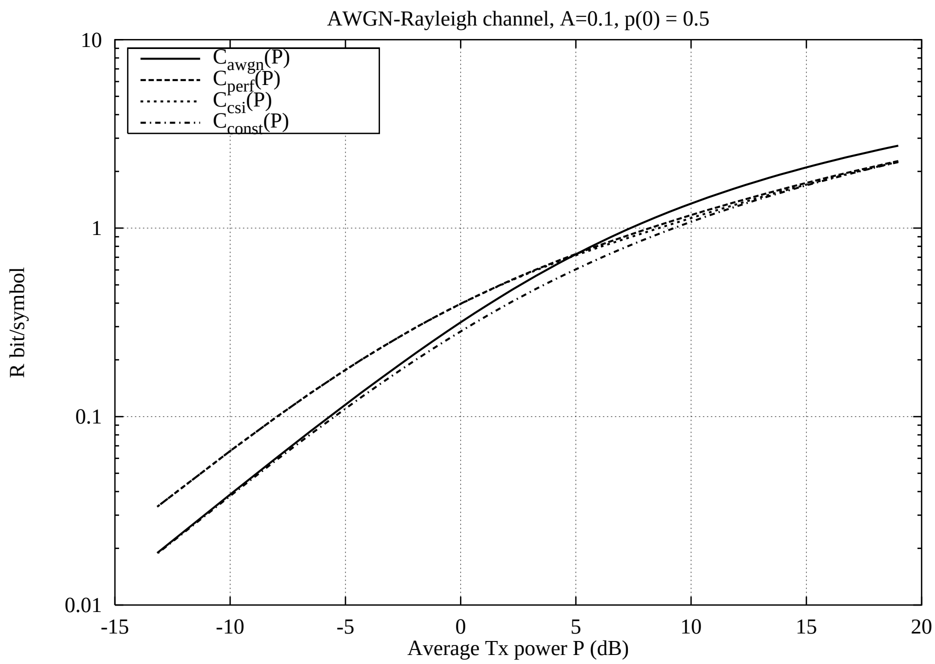

Figure 7: Capacity vs. $\mathcal{P}$ for Example 2, with the probability $p(0)$ of having a Rayleigh channel equal to 0.5.

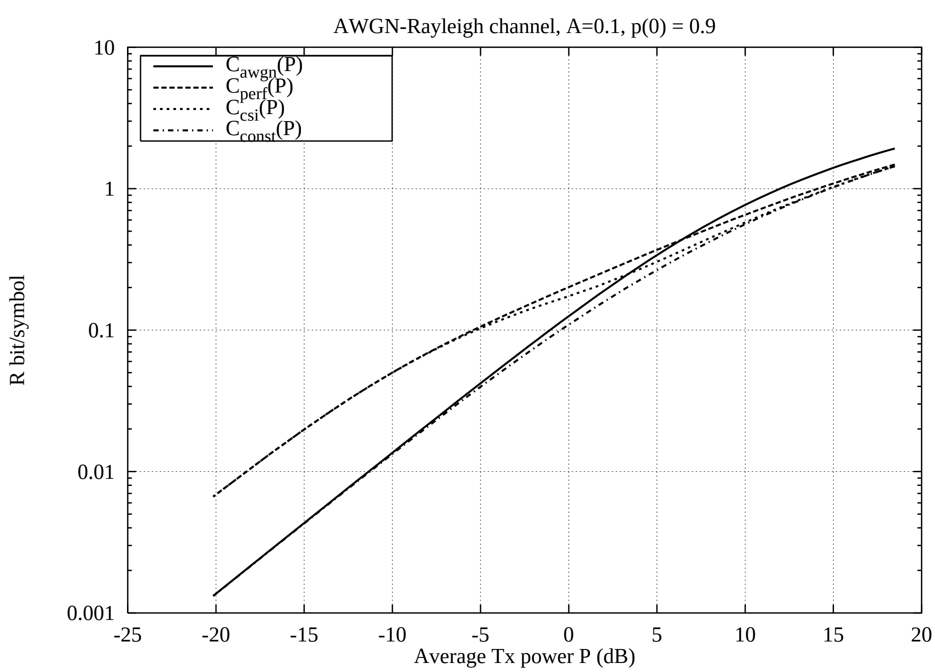

Figure 8: Capacity vs. $\mathcal{P}$ for Example 2, with the probability $p(0)$ of having a Rayleigh channel equal to 0.9.

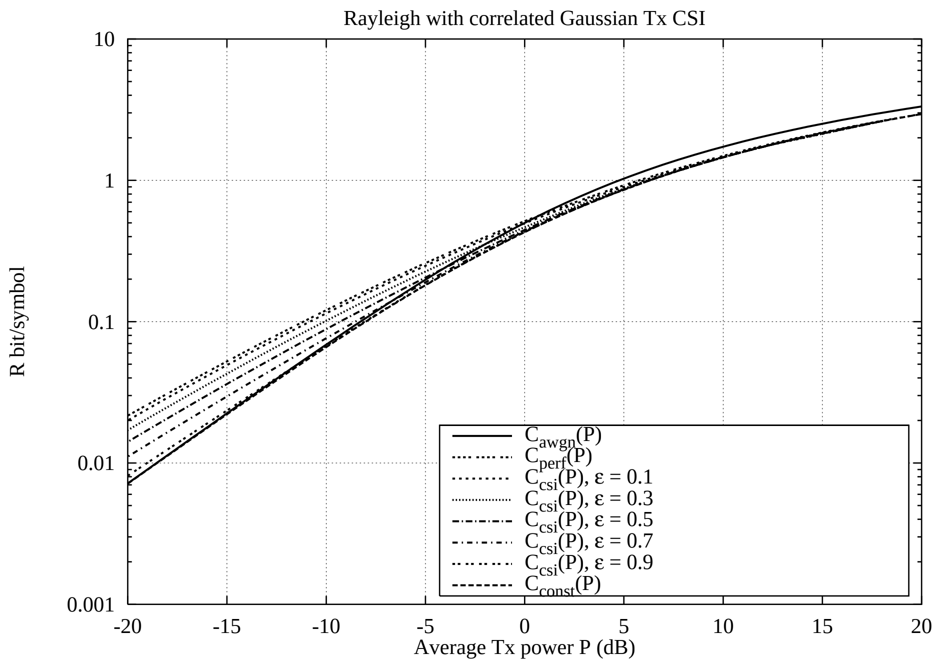

Figure 9: Capacity vs. $\mathcal{P}$ for Example 3, with channel gain estimation error $\epsilon = 0.1, 0.3, 0.5, 0.7, 0.9$.

List of Figures

Footnotes

Note 1 Low-rate feedback links are already implemented in many standards for cellular wireless systems [23].

Note 2 There is an obvious one-to-one correspondence between sequences $u_1^n \in \mathcal{U}^n$ and the integers from 1 to $|\mathcal{U}|^n$. Then, with a slight abuse of terminology, we indicate as the u_1^n-th component of the vector $t_n \in \mathcal{X}^{|\mathcal{U}|^n}$ the component whose index is the integer corresponding to u_1^n.

Note 3 This conclusion does not hold if a constraint on the transmission delay is taken into account. In this case, the so called delay-limited capacity [30, 31, 32] and/or the information outage probability [13] should be considered, since ergodicity or, more in general, information stability cannot be used.

Note 4 The results for this real model can be immediately translated into results for the circularly-symmetric complex model with average energy per complex input symbol E_s and one-sided noise power spectral density N_0 by letting $\mathcal{P} = E_s/N_0$ and by doubling the information rates (expressed in nat per complex symbol). The input power constraint should be regarded as an average *transmit* SNR constraint.