by $I(X; Y) = 0$. Therefore,

$$\sum_{x, y, z, u} \hat{p}(x, y, z, u)$$

$$= \sum_{x, y, z, u : p(x)>0, p(y)>0, p(z)>0, p(u)>0} \hat{p}(x, y, z, u)$$

$$= \sum_{x, y, z, u : p(x)>0, p(y)>0, p(z)>0, p(u)>0} \frac{p(x, z)p(x, u)p(y, z)p(y, u)}{p(z)p(u)p(x)p(y)}$$

$$= \sum_{x, y, z, u : p(x)>0, p(y)>0, p(z)>0, p(u)>0} \frac{p(x, y, z)p(x, u)p(y, u)}{p(u)p(x, y)}$$

$$= \sum_{x, y, u : p(x)>0, p(y)>0, p(u)>0} \frac{p(x, u)p(y, u)}{p(u)} = 1.$$

The theorem is proved.   □

ACKNOWLEDGMENT

The authors wish to thank F. Matúš for his careful reading of the manuscript. R. W. Yeung wishes to thank B. Hajek for his useful discussion.

REFERENCES

[1] T. S. Han, "A uniqueness of Shannon's information distance and related non-negativity problems," *J. Comb., Inform. & Syst. Sci.*, vol. 6, pp. 320–321, 1981.
[2] F. Matúš, "Probabilistic conditional independence structures and matroid theory: Background," *Int. J. General Syst.*, vol. 22, pp. 185–196, 1994.
[3] R. W. Yeung, "A new outlook on Shannon's information measures," *IEEE Trans. Inform. Theory*, vol. 37, pp. 466–474, 1991.
[4] ——, "A framework for linear information inequalities," *IEEE Trans. Inform. Theory*, this issue, pp. 1924–1934.

Capacity of Fading Channels with Channel Side Information

Andrea J. Goldsmith, *Member, IEEE*,
and Pravin P. Varaiya, *Fellow, IEEE*

Abstract— We obtain the Shannon capacity of a fading channel with channel side information at the transmitter and receiver, and at the receiver alone. The optimal power adaptation in the former case is "water-pouring" in time, analogous to water-pouring in frequency for time-invariant frequency-selective fading channels. Inverting the channel results in a large capacity penalty in severe fading.

Index Terms— Capacity, channel side information, fading channels, power adaptation.

I. INTRODUCTION

The growing demand for wireless communication makes it important to determine the capacity limits of fading channels. In this correspondence, we obtain the capacity of a single-user fading channel when the channel fade level is tracked by both the transmitter and receiver, and by the receiver alone. In particular, we show that the fading-channel capacity with channel side information at both the transmitter and receiver is achieved when the transmitter adapts its power, data rate, and coding scheme to the channel variation. The optimal power allocation is a "water-pouring" in time, analogous to the water-pouring used to achieve capacity on frequency-selective fading channels [1], [2].

We show that for independent and identically distributed (i.i.d.) fading, using receiver side information only has a lower complexity and the same approximate capacity as optimally adapting to the channel, for the three fading distributions we examine. However, for correlated fading, not adapting at the transmitter causes both a decrease in capacity and an increase in encoding and decoding complexity. We also consider two suboptimal adaptive techniques: channel inversion and truncated channel inversion, which adapt the transmit power but keep the transmission rate constant. These techniques have very simple encoder and decoder designs, but they exhibit a capacity penalty which can be large in severe fading. Our capacity analysis for all of these techniques neglects the effects of estimation error and delay, which will generally degrade capacity.

The tradeoff between these adaptive and nonadaptive techniques is therefore one of both capacity and complexity. Assuming that the channel is estimated at the receiver, the adaptive techniques require a feedback path between the transmitter and receiver and some complexity in the transmitter. The optimal adaptive technique uses variable-rate and power transmission, and the complexity of its decoding technique is comparable to the complexity of decoding a sequence of additive white Gaussian noise (AWGN) channels in parallel. For the nonadaptive technique, the code design must make use of the channel correlation statistics, and the decoder complexity is proportional to the channel decorrelation time. The optimal adaptive technique always has the highest capacity, but the increase relative
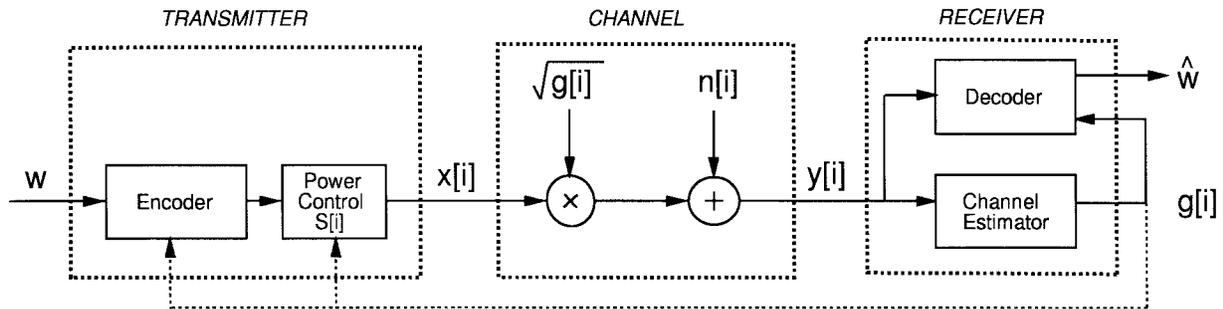
Fig. 1.   System model.

to nonadaptive transmission using receiver side information only is small when the fading is approximately i.i.d. The suboptimal adaptive techniques reduce complexity at a cost of decreased capacity.

This tradeoff between achievable data rates and complexity is examined for adaptive and nonadaptive modulation in [3], where adaptive modulation achieves an average data rate within 7–10 dB of the capacity derived herein (depending on the required error probability), while nonadaptive modulation exhibits a severe rate penalty. Trellis codes can be combined with the adaptive modulation to achieve higher rates [4].

We do not consider the case when the channel fade level is unknown to both the transmitter and receiver. Capacity under this assumption was obtained for the Gilbert–Elliot channel in [5] and for more general Markov channel models in [6]. If the statistics of the channel variation are also unknown, then channels with deep fading will typically have a capacity close to zero. This is because the data must be decoded without error, which is difficult when the location of deep fades are random. In particular, the capacity of a fading channel with arbitrary variation is at most the capacity of a time-invariant channel under the worst case fading conditions. More details about the capacity of time-varying channels under these assumptions can be found in the literature on Arbitrarily Varying Channels [7], [8].

The remainder of this correspondence is organized as follows. The next section describes the system model. The capacity of the fading channel under the different side information conditions is obtained in Section III. Numerical calculation of these capacities in Rayleigh, log-normal, and Nakagami fading is given in Section IV. Our main conclusions are summarized in the final section.

II. SYSTEM MODEL

Consider a discrete-time channel with stationary and ergodic time-varying gain $\sqrt{g[i]}$, $0 \leq g[i]$, and AWGN $n[i]$. We assume that the channel power gain $g[i]$ is independent of the channel input and has an expected value of unity. Let S denote the average transmit signal power, N_0 denote the noise density of $n[i]$, and B denote the received signal bandwidth. The instantaneous received signal-to-noise ratio (SNR) is then $\gamma[i] = Sg[i]/(N_0 B)$, and its expected value over all time is $S/(N_0 B)$.

The system model, which sends an input message w from the transmitter to the receiver, is illustrated in Fig. 1. The message is encoded into the codeword x, which is transmitted over the time-varying channel as $x[i]$ at time i. The channel gain $g[i]$ changes over the transmission of the codeword. We assume perfect instantaneous channel estimation so that the channel power gain $g[i]$ is known to the receiver at time i. We also consider the case when $g[i]$ is known to both the receiver and transmitter at time i, as might be obtained through an error-free delayless feedback path. This allows the transmitter to adapt $x[i]$ to the channel gain at time i, and is a reasonable model for a slowly varying channel with channel estimation and transmitter feedback.

III. CAPACITY ANALYSIS

A. Side Information at the Transmitter and Receiver

Assume that the channel power gain $g[i]$ is known to both the transmitter and receiver at time i. The capacity of a time-varying channel with side information about the channel state at both the transmitter and receiver was originally considered by Wolfowitz for the following model. Let $c[i]$ be a stationary and ergodic stochastic process representing the channel state, which takes values on a finite set $\mathcal{S}$ of discrete memoryless channels. Let C_s denotes the capacity of a particular channel $s \in \mathcal{S}$, and $p(s)$ denote the probability, or fraction of time, that the channel is in state s. The capacity of this time-varying channel is then given by [9, Theorem 4.6.1]

$$C = \sum_{s \in \mathcal{S}} C_s p(s). \tag{1}$$

We now consider the capacity of the fading channel shown in Fig. 1. Specifically, assume an AWGN fading channel with stationary and ergodic channel gain $g[i]$. It is well known that a time-invariant AWGN channel with average received SNR γ has capacity $C_\gamma = B \log(1 + \gamma)$. Let $p(\gamma) = p(\gamma[i] = \gamma)$ denote the probability distribution of the received SNR. From (1), the capacity of the fading channel with transmitter and receiver side information is thus[1]

$$C = \int_\gamma C_\gamma p(\gamma) \, d\gamma = \int_\gamma B \log(1 + \gamma) p(\gamma) \, d\gamma. \tag{2}$$

By Jensen's inequality, (2) is always less than the capacity of an AWGN channel with the same average power. Suppose now that we also allow the transmit power $S(\gamma)$ to vary with $\gamma[i]$, subject to an average power constraint S

$$\int_\gamma S(\gamma) p(\gamma) \, d\gamma \leq S. \tag{3}$$

With this additional constraint, we cannot apply (2) directly to obtain the capacity. However, we expect that the capacity with this average power constraint will be the average capacity given by (2) with the power optimally distributed over time. This motivates the following definition for the fading channel capacity, for which we subsequently prove the channel coding theorem and converse.

Definition: Given the average power constraint (3), define the time-varying channel capacity by

$$C(S) = \max_{S(\gamma): \int S(\gamma) p(\gamma) \, d\gamma = S} \int_\gamma B \log\left(1 + \frac{S(\gamma)\gamma}{S}\right) p(\gamma) \, d\gamma. \tag{4}$$

The channel coding theorem shows that this capacity is achievable, and the converse shows that no code can achieve a higher rate with arbitrarily small error probability. These two theorems are stated below and proved in the Appendix.

---

[1] Wolfowitz's result was for γ ranging over a finite set, but it can be extended to infinite sets, as we show in the Appendix.
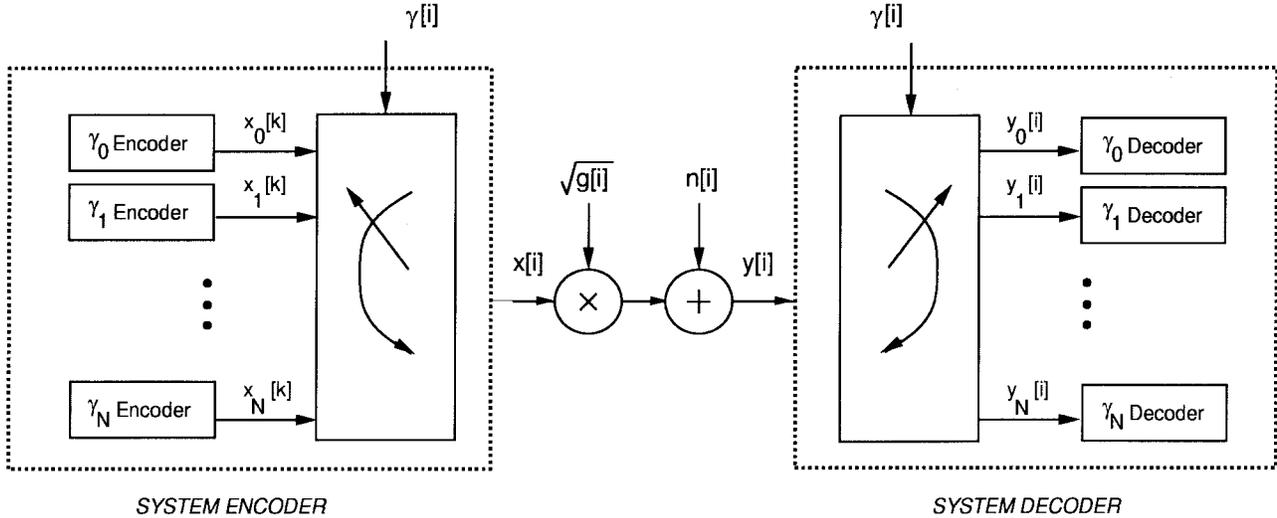
Fig. 2.   Multiplexed coding and decoding.

Coding Theorem: There exists a coding scheme with average power S that achieves any rate $R < C(S)$ with arbitrarily small probability of error.

Converse: Any coding scheme with rate $R > C(S)$ and average power S will have a probability of error bounded away from zero.

It is easily shown that the power adaptation which maximizes (4) is

$$\frac{S(\gamma)}{S} = \begin{cases} \dfrac{1}{\gamma_0} - \dfrac{1}{\gamma}, & \gamma \geq \gamma_0 \\ 0, & \gamma < \gamma_0 \end{cases} \quad (5)$$

for some "cutoff" value γ_0. If $\gamma[i]$ is below this cutoff then no data is transmitted over the ith time interval. Since γ is time-varying, the maximizing power adaptation policy of (5) is a "water-pouring" formula in time [1] that depends on the fading statistics $p(\gamma)$ only through the cutoff value γ_0.

Substituting (5) into (3), we see that γ_0 must satisfy

$$\int_{\gamma_0}^{\infty} \left(\frac{1}{\gamma_0} - \frac{1}{\gamma} \right) p(\gamma)\, d\gamma = 1. \quad (6)$$

Substituting (5) into (4) then yields a closed-form capacity formula

$$C(S) = \int_{\gamma_0}^{\infty} B \log \left(\frac{\gamma}{\gamma_0} \right) p(\gamma)\, d\gamma. \quad (7)$$

The channel coding and decoding which achieves this capacity is described in the Appendix, but the main idea is a "time diversity" system with multiplexed input and demultiplexed output, as shown in Fig. 2. Specifically, we first quantize the range of fading values to a finite set $\{\gamma_j : 0 \leq j \leq N\}$. Given a blocklength n, we then design an encoder/decoder pair for each γ_j with codewords $x_j \in \{x_{w_j}[k]\}$, $w_j = 1, \cdots, 2^{n_j R_j}$ of average power $S(\gamma_j)$ which achieve rate $R_j \approx C_j$, where C_j is the capacity of a time-invariant AWGN channel with received SNR $S(\gamma_j)\gamma_j/S$ and $n_j = np(\gamma \approx \gamma_j)$. These encoder/decoder pairs correspond to a set of input and output ports associated with each γ_j. When $\gamma[i] \approx \gamma_j$, the corresponding pair of ports are connected through the channel. The codewords associated with each γ_j are thus multiplexed together for transmission, and demultiplexed at the channel output. This effectively reduces the time-varying channel to a set of time-invariant channels in parallel, where the jth channel only operates when $\gamma[i] \approx \gamma_j$. The average rate on the channel is just the sum of rates R_j associated with each of the γ_j channels weighted by $p(\gamma \approx \gamma_j)$. This sketches the proof of the coding theorem. Details

can be found in the Appendix, along with the converse theorem that no other coding scheme can achieve a higher rate.

B. Side Information at the Receiver

In [10], it was shown that if the channel variation satisfies a compatibility constraint then the capacity of the channel with side information at the receiver only is also given by the average capacity formula (2). The compatibility constraint is satisfied if the channel sequence is i.i.d. and if the input distribution which maximizes mutual information is the same regardless of the channel state. In this case, for a constant transmit power the side information at the transmitter does not increase capacity, as we now show.

If $g[i]$ is known at the decoder then by scaling, the fading channel with power gain $g[i]$ is equivalent to an AWGN channel with noise power $N_0 B/g[i]$. If the transmit power is fixed at S and $g[i]$ is i.i.d. then the input distribution at time i which achieves capacity is an i.i.d. Gaussian distribution with average power S. Thus without power adaptation, the fading AWGN channel satisfies the compatibility constraint of [10]. The channel capacity with i.i.d. fading and receiver side information only is thus given by

$$C(S) = \int B \log (1 + \gamma) p(\gamma)\, d\gamma \quad (8)$$

which is the same as (2), the capacity with transmitter and receiver side information but no power adaptation. The code design in this case chooses codewords $\{x_{w_j}[i]\}_{i=1}^{n}$, $w_j = 1, \cdots, 2^{nR}$ at random from an i.i.d. Gaussian source with variance equal to the signal power. The maximum-likelihood decoder then observes the channel output vector $y[\cdot]$ and chooses the codeword x_{w_j} which minimizes the Euclidean distance

$$\| (y[1], \cdots, y[n]) - (x_{w_j}[1]g[1], \cdots, x_{w_j}[n]g[n]) \|.$$

Thus for i.i.d. fading and constant transmit power, side information at the transmitter has no capacity benefit, and the encoder/decoder pair based on receiver side information alone is simpler than the adaptive multiplexing technique shown in Fig. 2.

However, most physical channels exhibit correlated fading. If the fading is not i.i.d. then (8) is only an upper bound to channel capacity. In addition, without transmitter side information, the code design must incorporate the channel correlation statistics, and the complexity of the maximum-likelihood decoder will be proportional to the channel decorrelation time.
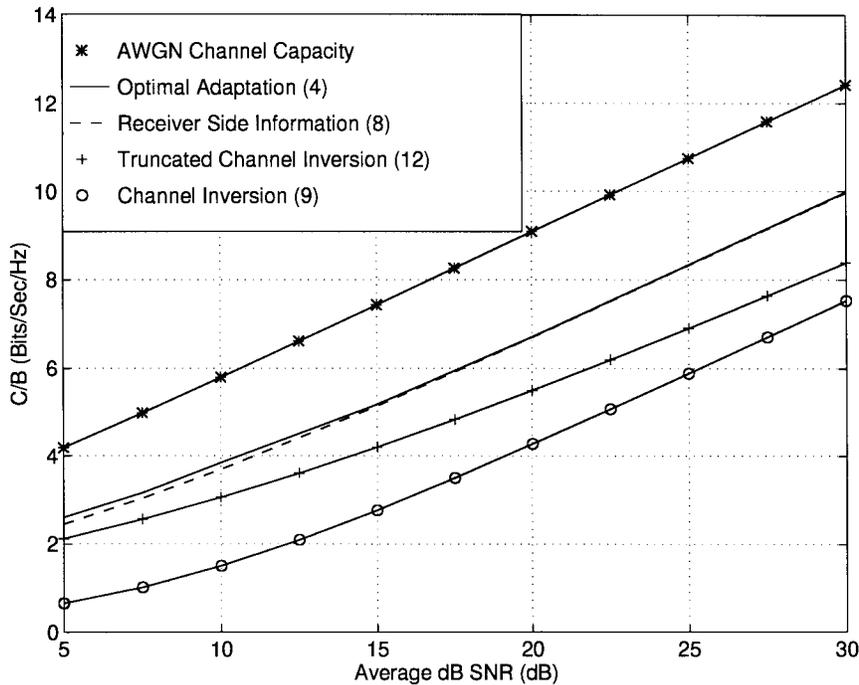
Fig. 3. Capacity in log-normal fading ($\sigma = 8$ dB).

C. Channel Inversion

We now consider a suboptimal transmitter adaptation scheme where the transmitter uses the channel side information to maintain a constant received power, i.e., it inverts the channel fading. The channel then appears to the encoder and decoder as a time-invariant AWGN channel. The power adaptation for channel inversion is given by $S(\gamma)/S = \sigma/\gamma$, where σ equals the constant received SNR which can be maintained under the transmit power constraint (3). The constant σ thus satisfies $\int (\sigma/\gamma)p(\gamma) = 1$, so $\sigma = 1/E[1/\gamma]$.

The fading channel capacity with channel inversion is just the capacity of an AWGN channel with SNR σ:

$$C(S) = B \log\left[1 + \sigma\right] = B \log\left[1 + \frac{1}{\boldsymbol{E}[1/\gamma]}\right]. \qquad (9)$$

Channel inversion is common in spread-spectrum systems with near–far interference imbalances [11]. It is also very simple to implement, since the encoder and decoder are designed for an AWGN channel, independent of the fading statistics. However, it can exhibit a large capacity penalty in extreme fading environments. For example, in Rayleigh fading $\boldsymbol{E}[1/\gamma]$ is infinite, and thus the capacity with channel inversion is zero.

We also consider a truncated inversion policy that only compensates for fading above a certain cutoff fade depth γ_0

$$\frac{S(\gamma)}{S} = \begin{cases} \dfrac{\sigma}{\gamma}, & \gamma \geq \gamma_0 \\ 0, & \gamma < \gamma_0. \end{cases} \qquad (10)$$

Since the channel is only used when $\gamma \geq \gamma_0$, the power constraint (3) yields $\sigma = 1/\boldsymbol{E}_{\gamma_0}[1/\gamma]$, where

$$\boldsymbol{E}_{\gamma_0}[1/\gamma] \triangleq \int_{\gamma_0}^{\infty} \frac{1}{\gamma}\, p(\gamma)\, d\gamma. \qquad (11)$$

For decoding this truncated policy, the receiver must know when $\gamma < \gamma_0$. The capacity in this case, obtained by maximizing over all possible γ_0, is

$$C(S) = \max_{\gamma_0} B \log\left[1 + \frac{1}{\boldsymbol{E}_{\gamma_0}[1/\gamma]}\right]p(\gamma \geq \gamma_0). \qquad (12)$$

IV. NUMERICAL RESULTS

Figs. 3–5 show plots of (4), (8), (9), and (12) as a function of average received SNR for log-normal fading (standard deviation $\sigma = 8$ dB), Rayleigh fading, and Nakagami fading (with Nakagami parameter $m = 2$). The capacity in AWGN for the same average power is also shown for comparison. Several observations are worth noting. First, for this range of SNR values, the capacity of the AWGN channel is larger, so fading reduces channel capacity. This will not always be the case at very low SNR's. The severity of the fading is indicated by the Nakagami parameter m, where $m = 1$ for Rayleigh fading and $m = \infty$ for an AWGN channel without fading. Thus comparing Figs. 4 and 5 we see that, as the severity of the fading decreases (m goes from one to two), the capacity difference between the various adaptive policies also decreases, and their respective capacities approach that of the AWGN channel.

The difference between the capacity curves (4) and (8) are negligible in all cases. Recalling that (2) and (8) are the same, this implies that when the transmission rate is adapted relative to the channel, adapting the power as well yields a negligible capacity gain. It also indicates that for i.i.d. fading, transmitter adaptation yields a negligible capacity gain relative to using only receiver side information. We also see that in severe fading conditions (Rayleigh and log-normal fading), truncated channel inversion exhibits a 1–5-dB-rate penalty and channel inversion without truncation yields a very large capacity loss. However, under mild fading conditions (Nakagami with $m = 2$) the capacity of all the adaptation techniques are within 3 dB of each other and within 4 dB of the AWGN channel capacity. These differences will further decrease as the fading diminishes ($m \to \infty$).

We can view these results as a tradeoff between capacity and complexity. The adaptive policy with transmitter side information requires more complexity in the transmitter (and it typically also requires a feedback path between the receiver and transmitter to obtain the side information). However, the decoder in the receiver is relatively simple. The nonadaptive policy has a relatively simple transmission scheme, but its code design must use the channel correlation statistics (often unknown), and the decoder complexity is
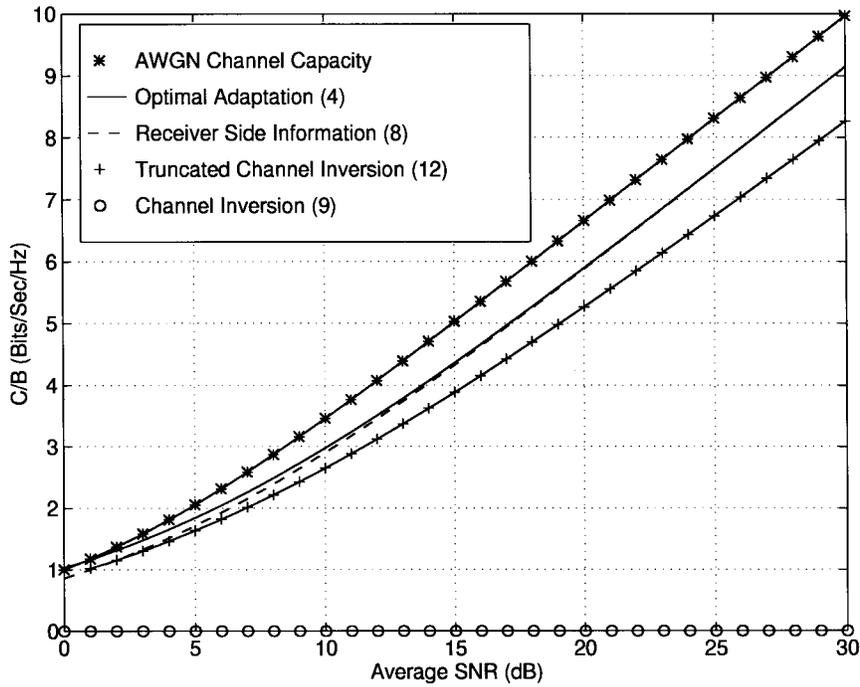
Fig. 4. Capacity in Rayleigh fading ($m = 1$).
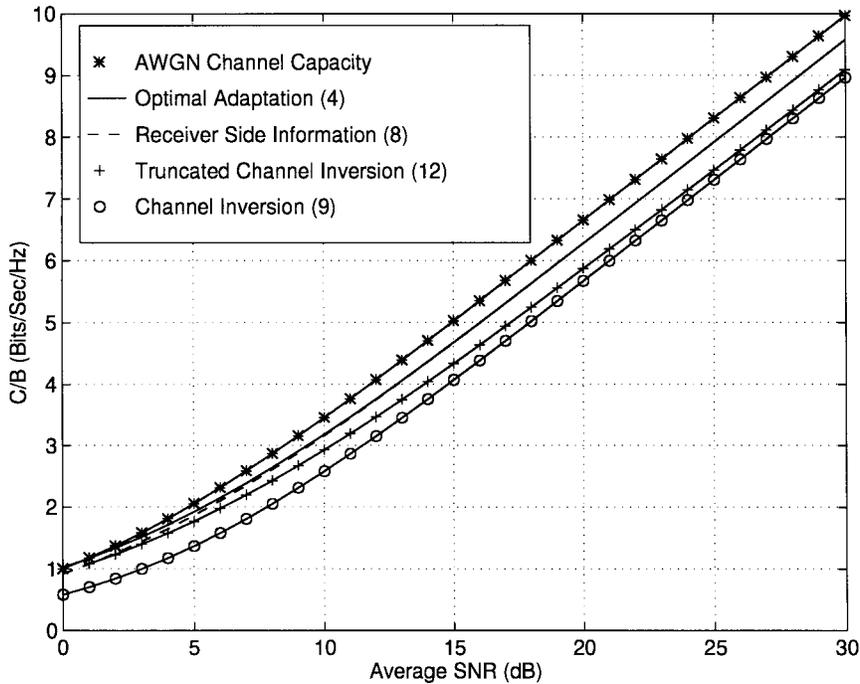
Fig. 5. Capacity in Nakagami fading ($m = 2$).

proportional to the channel decorrelation time. The channel inversion and truncated inversion policies use codes designed for AWGN channels, and are therefore the least complex to implement, but in severe fading conditions they exhibit large capacity losses relative to the other techniques.

In general, Shannon capacity analysis does not give any indication how to design adaptive or nonadaptive techniques for real systems. Achievable rates for adaptive trellis-coded quadrature amplitude modulation (QAM) have been investigated in [4], where a simple four-state trellis code combined with adaptive six-constellation QAM

was shown to achieve rates within 7 dB of the capacity (4) in Figs. 3 and 4. Using more complex codes and a richer constellation set comes within a few decibels of the Shannon capacity limit.

V. CONCLUSIONS

We have determined the capacity of a fading AWGN channel with an average power constraint under different channel side information conditions. When side information about the current channel state is available to both the transmitter and receiver, the optimal adaptive transmission scheme uses water-pouring in time for power adaptation,

and a variable-rate multiplexed coding scheme. In channels with correlated fading this adaptive transmission scheme yields both higher capacity and a lower complexity than nonadaptive transmission using receiver side information. However, it does not exhibit a significant capacity increase or any complexity reduction in i.i.d. fading as compared to nonadaptive transmission. Channel inversion has the lowest encoding and decoding complexity, but it also suffers a large capacity penalty in severe fading. The capacity of all of these techniques converges to that of an AWGN channel as the severity of the fading diminishes.

APPENDIX

We now prove that the capacity of the time-varying channel in Section II is given by (4). We first prove the coding theorem, followed by the converse proof.

Coding Theorem: Let $C(S)$ be given by (4). Then for any $R < C(S)$ there exists a sequence of $(2^{nR_n}, n)$ block codes with average power S, rate $R_n \to R$, and probability of error $\epsilon_n \to 0$ as $n \to \infty$.

Proof: Fix any $\epsilon > 0$, and let $R = C(S) - 3\epsilon$. Define

$$\gamma_j = j/m + \rho_0, \qquad j = 0, \cdots, mM = N$$

to be a finite set of SNR values, where ρ_0 is the cutoff associated with the optimal power control policy for average power S (defined as γ_0 in (5) from Section III-A). The received SNR of the fading channel takes values in $0 \le \gamma < \infty$, and the γ_j values discretize the subset of this range $\rho_0 \le \gamma \le M + \rho_0$ for a step size of $1/m$. We say that the fading channel is in state s_j, $j = 0, \cdots, mM$, if $\gamma_j \le \gamma < \gamma_{j+1}$, where $\gamma_{mM+1} = \infty$. We also define a power control policy associated with state s_j by

$$\frac{\sigma_j}{S} = \frac{S(\gamma_j)}{S} = \frac{1}{\rho_0} - \frac{1}{\gamma_j}. \tag{13}$$

Over a given time interval $[0, n]$, let N_j denote the number of transmissions during which the channel is in state s_j. By the stationarity and ergodicity of the channel variation

$$\frac{N_j}{n} \to p(\gamma_j \le \gamma < \gamma_{j+1}), \qquad \text{as } n \to \infty. \tag{14}$$

Consider a time-invariant AWGN channel with SNR γ_j and transmit power σ_j. For a given n, let

$$n_j = \lfloor np(\gamma_j \le \gamma < \gamma_{j+1}) \rfloor = np(\gamma_j \le \gamma < \gamma_{j+1})$$

for n sufficiently large. From Shannon [12], for

$$R_j = B \log (1 + \gamma_j \sigma_j/S) = B \log (\gamma_j/\rho_0)$$

we can develop a sequence of $(2^{n_j R_j})$ codes

$$\{x_{w_j}[k]\}_{k=1}^{n_j}, \qquad w_j = 1, \cdots, 2^{n_j R_j}$$

with average power σ_j and error probability $\epsilon_{n,j} \to 0$ as $n_j \to \infty$.

The message index $w \in [1, \cdots, 2^{nR_n}]$ is transmitted over the $N + 1$ channels in Fig. 2 as follows. We first map w to the indices $\{w_j\}_{j=0}^{N}$ by dividing the nR_n bits which determine the message index into sets of $n_j R_j$ bits. We then use the multiplexing strategy described in Section III-A to transmit the codeword $x_{w_j}[\cdot]$ whenever the channel is in state s_j. On the interval $[0, n]$ we use the jth channel N_j times. We can thus achieve a transmission rate of

$$R_n = \sum_{j=0}^{mM} R_j \frac{N_j}{n} = \sum_{j=0}^{mM} B \log \left(\frac{\gamma_j}{\rho_0}\right) \frac{N_j}{n}. \tag{15}$$

The average transmit power for the multiplexed code is

$$S_n = \sum_{j=0}^{mM} S \left(\frac{1}{\rho_0} - \frac{1}{\gamma_j}\right) \frac{N_j}{n}. \tag{16}$$

From (14) and (15), it is easily seen that

$$\lim_{n \to \infty} R_n = \sum_{j=0}^{mM} B \log \left(\frac{\gamma_j}{\rho_0}\right) p(\gamma_j \le \gamma < \gamma_{j+1}). \tag{17}$$

So, for ϵ fixed, we can find n sufficiently large such that

$$R_n \ge \sum_{j=0}^{mM} B \log \left(\frac{\gamma_j}{\rho_0}\right) p(\gamma_j \le \gamma < \gamma_{j+1}) - \epsilon. \tag{18}$$

Moreover, the power control policy σ_j satisfies the average power constraint for asymptotically large n

$$\lim_{n \to \infty} \sum_{j=0}^{mM} S \left(\frac{1}{\rho_0} - \frac{1}{\gamma_j}\right) \frac{N_j}{n} \overset{\text{a}}{=} \sum_{j=0}^{mM} S \left(\frac{1}{\rho_0} - \frac{1}{\gamma_j}\right) \int_{\gamma_j}^{\gamma_{j+1}} p(\gamma) \, d\gamma$$

$$\overset{\text{b}}{\le} \sum_{j=0}^{mM} \int_{\gamma_j}^{\gamma_{j+1}} S \left(\frac{1}{\rho_0} - \frac{1}{\gamma}\right) p(\gamma) \, d\gamma$$

$$= S \int_{\rho_0}^{\infty} \left(\frac{1}{\rho_0} - \frac{1}{\gamma}\right) p(\gamma) \, d\gamma \overset{\text{c}}{\le} S \tag{19}$$

where a follows from (14), b follows from the fact that $\gamma_j \le \gamma$ for $\gamma \in [\gamma_j, \gamma_{j+1})$, and c follows from (3).

Since the SNR of the channel during transmission of the code x_j is greater than or equal to γ_j, the error probability of the multiplexed coding scheme is bounded above by

$$\epsilon_n \le \sum_{j=0}^{mM} \epsilon_{n,j} \to 0, \qquad \text{as } n \to \infty \tag{20}$$

since $n \to \infty$ implies $n_j \to \infty$ for all j channels of interest. Thus it remains to show that for fixed ϵ there exists m and M such that

$$\sum_{j=0}^{mM} B \log \left(\frac{\gamma_j}{\rho_0}\right) p(\gamma_j \le \gamma < \gamma_{j+1}) \ge C(S) - 2\epsilon. \tag{21}$$

It is easily shown that

$$C(S) = \int_{\rho_0}^{\infty} B \log \left(\frac{\gamma}{\rho_0}\right) p(\gamma) \, d\gamma \le B \log (1 + \overline{\gamma}) - B \log \rho_0 p(\gamma \ge \rho_0) < \infty \tag{22}$$

where the finite bound on $C(S)$ follows from the fact that ρ_0 must be greater than zero to satisfy (6). So for fixed ϵ there exists an M_ϵ such that

$$\int_{M_\epsilon + \rho_0}^{\infty} B \log \left(\frac{\gamma}{\rho_0}\right) p(\gamma) \, d\gamma < \epsilon. \tag{23}$$

Moreover, for M fixed, the monotone convergence theorem [13] implies that

$$\lim_{m \to \infty} \sum_{j=0}^{mM-1} B \log \left(\frac{\gamma_j}{\rho_0}\right) p(\gamma_j \le \gamma < \gamma_{j+1})$$

$$= \lim_{m \to \infty} \sum_{j=0}^{mM-1} \int_{\gamma_j}^{\gamma_{j+1}} B \log \left(\frac{\gamma_j}{\rho_0}\right) p(\gamma) \, d\gamma$$

$$= \int_{\rho_0}^{M + \rho_0} B \log \left(\frac{\gamma}{\rho_0}\right) p(\gamma) \, d\gamma. \tag{24}$$

Thus using the M_ϵ in (23) and combining (23) and (24) we see that for the given ϵ there exists an m sufficiently large such that

$$\sum_{j=0}^{mM_\epsilon} B \log \left(\frac{\gamma_j}{\rho_0}\right) p(\gamma_j \le \gamma < \gamma_{j+1})$$

$$\ge \int_{\rho_0}^{\infty} B \log \left(\frac{\gamma}{\rho_0}\right) p(\gamma) \, d\gamma - 2\epsilon \tag{25}$$

which completes the proof.

Converse: Any sequence of $(2^{nR}, n)$ codes with average power S and probability of error $\epsilon_n \to 0$ as $n \to \infty$ must have $R \le C(S)$.

Proof: Consider any sequence of $(2^{nR}, n)$ codes

$$\{x_w[i]\}_{i=1}^n, \qquad w = 1, \cdots, 2^{nR}$$

with average power S and $\epsilon_n \to 0$ as $n \to \infty$. We assume that the codes are designed with *a priori* knowledge of the channel side information $\gamma^n = \{\gamma[1], \cdots, \gamma[n]\}$, since any code designed under this assumption will have at least as high a rate as if $\gamma[i]$ were only known at time i. Assume that the message index W is uniformly distributed on $\{1, \cdots, 2^{nR}\}$. Then

$$
\begin{aligned}
nR &= H(W|\gamma^n)\\
&= H(W|Y^n, \gamma^n) + I(W; Y^n, \gamma^n)\\
&\overset{a}{\le} H(W|Y^n, \gamma^n) + I(X^n; Y^n|\gamma^n)\\
&\overset{b}{\le} 1 + \epsilon_n nR + I(X^n; Y^n|\gamma^n)
\end{aligned}
\tag{26}
$$

where a follows from the data processing theorem [14] and the side information assumption, and b follows from Fano's inequality.

Let N_γ denote the number of times over the interval $[0, n]$ that the channel has fade level γ. Also let $S_\gamma^n(w)$ denote the average power in x_w associated with fade level γ, so

$$S_\gamma^n(w) = \frac{1}{n} \sum_{i=1}^n |x_w[i]|^2 \mathbf{1}[\gamma[i] = \gamma].\tag{27}$$

The average transmit power over all codewords for a given fade level is denoted by $\overline{S}_\gamma^n = E_w[S_\gamma^n(w)]$, and we define

$$\overline{S}^n \triangleq \{\overline{S}_\gamma^n; 0 \le \gamma \le \infty\}.$$

With this notation, we have

$$
\begin{aligned}
I(X^n; Y^n|\gamma^n) &\overset{a}{=} \sum_{i=1}^n I(X_i; Y_i|\gamma[i])\\
&= \sum_{i=1}^n \int_0^\infty I(X; Y|\gamma)\mathbf{1}[\gamma[i]=\gamma]\,d\gamma\\
&= \int_0^\infty I(X; Y|\gamma)N_\gamma\,d\gamma\\
&= \int_0^\infty E_w[I(X; Y|\gamma, S_\gamma^n(w))]N_\gamma\,d\gamma\\
&\overset{b}{\le} \int_0^\infty I(X; Y|\gamma, \overline{S}_\gamma^n)N_\gamma\,d\gamma\\
&\overset{c}{\le} \int_0^\infty B\log\left(1 + \frac{\gamma\overline{S}_\gamma^n}{S}\right)N_\gamma\,d\gamma
\end{aligned}
\tag{28}
$$

where a follows from the fact that the channel is memoryless when conditioned on $\gamma[i]$, b follows from Jensen's inequality, and c follows from the fact that the maximum mutual information on an AWGN channel with bandwidth B and SNR $\sigma = \gamma\overline{S}_\gamma^n/S$ is $B\log(1+\sigma)$.

Combining (26) and (28) yields

$$nR \le 1 + \epsilon_n nR + \int_0^\infty B\log\left(1 + \frac{\gamma\overline{S}_\gamma^n}{S}\right)n_\gamma\,d\gamma.\tag{29}$$

By assumption, each codeword satisfies the average power constraint, so for all w

$$\int_0^\infty S_\gamma^n(w)(N_\gamma/n) \le S.$$

Thus

$$\int_0^\infty \overline{S}_\gamma^n(n_\gamma/n) \le S$$

also. Moreover, $\overline{S}^n$ takes values on a compact space, so there is a convergent subsequence

$$\overline{S}^{n_i} \to \overline{S}^\infty \triangleq \{\overline{S}_\gamma^\infty; 0 \le \gamma \le \infty\}.$$

Since $\overline{S}^{n_i}$ satisfies the average power constraint

$$\lim_{n_i \to \infty} \int_0^\infty \overline{S}_\gamma^{n_i}\frac{(N_i)_\gamma}{n_i}\,d\gamma = \int_0^\infty \overline{S}_\gamma^\infty p(\gamma)\,d\gamma \le S.\tag{30}$$

Dividing (29) by n, we have

$$R \le \frac{1}{n} + \epsilon_n R + \int_0^\infty B\log\left(1 + \frac{\gamma\overline{S}_\gamma^n}{S}\right)\frac{N_\gamma}{n}\,d\gamma.\tag{31}$$

Taking the limit of the right-hand side of (31) along the subsequence n_i yields

$$R \le \int_0^\infty B\log\left(1 + \frac{\gamma\overline{S}_\gamma^\infty}{S}\right)p(\gamma)\,d\gamma \le C(S)\tag{32}$$

by definition of $C(S)$ and the fact that, from (30), $\overline{S}_\gamma^\infty$ satisfies the average power constraint.

ACKNOWLEDGMENT

REFERENCES

[1] R. G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.
[2] S. Kasturia, J. T. Aslanis, and J. M. Cioffi, "Vector coding for partial response channels," *IEEE Trans. Inform. Theory*, vol. 36, pp. 741–762, July 1990.
[3] S.-G. Chua and A. J. Goldsmith, "Variable-rate variable-power MQAM for fading channels," in *VTC'96 Conf. Rec.*, Apr. 1996, pp. 815–819; also to be published in *IEEE Trans. Commun.*
[4] ——, "Adaptive coded modulation," in *ICC'97 Conf. Rec.*, June 1997; also submitted to *IEEE Trans. Commun.*
[5] M. Mushkin and I. Bar-David, "Capacity and coding for the Gilbert–Elliot channel," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1277–1290, Nov. 1989.
[6] A. J. Goldsmith and P. P. Varaiya, "Capacity, mutual information, and coding for finite-state Markov channels," *IEEE Trans. Inform. Theory*, vol. 42, pp. 868–886, May 1996.
[7] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Channels.* New York: Academic, 1981.
[8] I. Csiszár and P. Narayan, "The capacity of the arbitrarily varying channel," *IEEE Trans. Inform. Theory*, vol. 37, pp. 18–26, Jan. 1991.
[9] J. Wolfowitz, *Coding Theorems of Information Theory*, 2nd ed. New York: Springer-Verlag, 1964.
[10] R. J. McEliece and W. E. Stark, "Channels with block interference," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 44–53, Jan. 1984.
[11] K. S. Gilhousen, I. M. Jacobs, R. Padovani, A. J. Viterbi, L. A. Weaver, Jr., and C. E. Wheatley, III, "On the capacity of a cellular CDMA system," *IEEE Trans. Veh. Technol.*, vol. 40, pp. 303–312, May 1991.
[12] C. E. Shannon and W. Weaver, *A Mathematical Theory of Communication.* Urbana, IL: Univ. of Illinois Press, 1949.
[13] P. Billingsley, *Probability and Measure*, 2nd ed. New York: Wiley, 1986.
[14] T. Cover and J. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.