

Capacity of Memoryless Channels and Block-Fading Channels With Designable Cardinality-Constrained Channel State Feedback

Vincent K. N. Lau, *Senior Member, IEEE*, Youjian Liu, *Member, IEEE*, and Tai-Ann Chen, *Member, IEEE*

Abstract—A coding theorem is proved for memoryless channels when the channel state feedback of finite cardinality can be designed. Channel state information is estimated at the receiver and a function of the estimated channel state is causally fed back to the transmitter. The feedback link is assumed to be noiseless with a finite feedback alphabet, or equivalently, finite feedback rate. It is shown that the capacity can be achieved with a memoryless deterministic feedback and with a memoryless device which select transmitted symbols from a codeword of expanded alphabet according to current feedback. To characterize the capacity, we investigate the optimization of transmission and channel state feedback strategies. The optimization is performed for both channel capacity and error exponents. We show that the design of the optimal feedback scheme is identical to the design of scalar quantizer with modified distortion measures. We illustrate the optimization using Rayleigh block-fading channels. It is shown that the optimal transmission strategy has a general form of temporal water-filling in important cases. Furthermore, while feedback enhances the forward channel capacity more effectively in low-signal-to noise ratio (SNR) region compared with that of high-SNR region, the enhancement in error exponent is significant in both high- and low-SNR regions. This indicates that significant gain due to finite-rate channel state feedback is expected in practical systems in both SNR regions.

Index Terms—Channel capacity, fading channels, partial channel state information at the transmitter (CSIT), partial feedback.

I. INTRODUCTION

COMMUNICATION theory with channel state feedback¹ has been investigated over many decades. State feedback is generally classified into causal feedback and noncausal feedback. Causal feedback means that at time t_0 , the transmitter has feedback of channel state information from $t = 0$ to $t = t_0^-$. In the case of noncausal feedback, the transmitter has knowledge in advance on the realization of the channel state sequence from the start to the end of transmission. Clearly, causal feedback is more suitable to communication systems where channel states are measured sequentially in the temporal domain. On

the other hand, noncausal feedback is suitable to communication systems where channel states are measured block by block. An example is the orthogonal frequency-division multiplexing (OFDM) system where the channel state over the frequency domain is available noncausally to the transmitter [1]. In this paper, our focus is on causal feedback systems.

Communications with perfect channel state information at the transmitter (perfect CSIT) has been investigated by Shannon in [2]. Causal state feedback was assumed and the capacity is expressed as that of an equivalent memoryless channel without side information at either the transmitter or the receiver. Communication with perfect channel state information at the receiver (perfect CSIR) has been investigated, for example, in [3]. Communication with perfect CSIT and perfect CSIR has been treated in [4]–[9]. On the other hand, the case of partial state feedback has also been investigated in [10]–[14]. For example, noncausal partial state feedback was treated in [12], [13] and the corresponding causal versions has been investigated in [10], [11], [14]. For instance, the optimal transmission strategy (with average transmit power constraint and perfect channel state feedback) is well known to be temporal water-filling. However, in most literature, the forward channel capacity was evaluated based on some assumed statistical relationships $(p(u, v))$ of CSIT U and CSIR V . The general problem of optimal transmission and state feedback strategy design when the state feedback alphabets has a finite cardinality constraint is still an open problem. This problem is proposed in [14] and a simple example is given. Recent related work on outage capacity can be found in [15] and that on asymptotic cases can be found in [16]. The focus of this paper is to investigate the optimal transmission and feedback strategies and characterize the performance limit with finite feedback-alphabet cardinality as a constraint for both the forward channel capacity and error exponents.² We also give numeric results for single-input single-output (SISO) block-fading channels. Results on multiple-input multiple-output (MIMO) case is presented in [17].

We consider a general channel with channel state but, without loss of generality, we use block-fading channel terminology to describe it. The channel states remain static within a fading block but are independent and identically distributed (i.i.d.) across different fading blocks. A codeword spans multiple fading blocks as illustrated in Fig. 1(a) and hence, in the limit of large number of blocks, nonzero ergodic capacity is achieved. CSIR is estimated and a function of it is fed back to

Manuscript received August 26, 2002; revised March 1, 2004.

V. K. N. Lau is with the Department of Electrical and Electronic Engineering, University of Hong Kong, Hong Kong (e-mail: knlau@iee.org).

Y. Liu is with the Department of Electrical and Computer Engineering, University of Colorado at Boulder, Boulder, CO 80309 USA (e-mail: eugeneliu@iee.org).

T.-A. Chen is with Lucent Technologies, Whippany, NJ 07981-0903 USA (e-mail: tachen@iee.org).

Communicated by G. Caire, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2004.833334

¹In this paper, feedback refers to the feedback of channel state rather than the traditional feedback of channel output.

²The optimization based on error exponents is usually ignored in most previous literature.

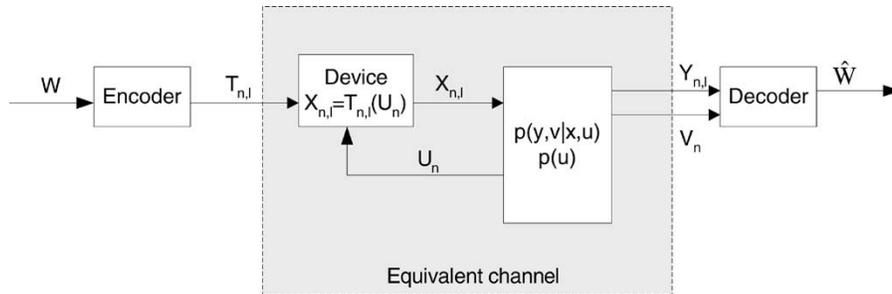


Fig. 2. Equivalent channel model of the communication system with CSIT and CSIR.

with probability density function (pdf) for applications to fading channels in Section III-C.

For the remainder of the paper, we shall use the following conventions in notations. A capital letter such as X denotes a random variable while a lower case letter such as x denotes its realization. A bold-font letter such as \mathbf{X} denotes a vector or a matrix of random variables. Let \mathbf{X}_1^N denote a sequence $\{\mathbf{X}_n, n = 1, 2, \dots, N\}$ and $\mathcal{E}[X]$ denote expectation of X . Index l denotes the l th symbol within a fading block. Index n denotes the n th fading block within a coding block. Following conventions, we will use $p(x)$ to represent pmf or pdf $p_X(x)$ when there is no confusion.

A. Forward Channel Model

We consider a communication system with a transmitter, a forward channel with a channel state, and a receiver as in Fig. 1(b). Let a length- N sequence \mathbf{X}_1^N ($\mathbf{X}_n \in \mathcal{X}^L$) be the discrete transmitted symbols and a length N sequence \mathbf{Y}_1^N ($\mathbf{Y}_n \in \mathcal{Y}^L$) be the discrete received symbols. \mathbf{X}_n is the transmitted signal of the n th fading block and \mathbf{Y}_n is the received signal of the n th fading block. The memoryless channel is characterized by the transition probability $p(\mathbf{y}_1^N | \mathbf{x}_1^N, \mathbf{s}_1^N)$, where \mathbf{s}_1^N represents channel states. We assume a general channel model. However, without loss of generality, we will use the terminologies of block-fading channels to describe the channel. Therefore, the channel states of L symbols of n th fading block take the same value $S_n \in \mathcal{S}$, which is i.i.d. with pmf $p_S(s_n)$. For N fading blocks, the forward channel is characterized by the channel transition probability and the channel state sequence probability given by

$$p(\mathbf{y}_1^N | \mathbf{x}_1^N, \mathbf{s}_1^N) = \prod_{n=1}^N \prod_{l=1}^L p_{Y|X,S}(y_{n,l} | x_{n,l}, s_n)$$

and

$$p(\mathbf{s}_1^N) = \prod_{n=1}^N p_S(s_n).$$

This is illustrated in Fig. 1(b).

B. Channel State Feedback Model

We assume the transmitter has some partial knowledge about the channel state. In particular, we consider the scenario of noiseless state feedback from receiver as illustrated in Fig. 1. A state sequence generator produces a channel state S_n for fading block n . The receiver receives $V_n \in \mathcal{V}$ and the transmitter receives $U_n \in \mathcal{U}$ through a noiseless feedback channel. Because U_n is produced based on V_1^n , it implies that $\mathbf{S} \rightarrow \mathbf{V} \rightarrow \mathbf{U}$

forms a Markov chain. The relation between \mathbf{S} and \mathbf{V} is i.i.d., i.e., $p(\mathbf{S}, \mathbf{V}) = \prod_n p(s_n, v_n)$. Therefore,

$$p(\mathbf{s}, \mathbf{v}, \mathbf{u}) = \prod_n p(s_n) p(v_n | s_n) p(\mathbf{u} | \mathbf{v}).$$

We assume the CSIT is causal in the sense that the transmitter at fading block n has knowledge of the state feedback $\{U_1, \dots, U_n\}$ only. The transmitted symbols, $\{X_{n,1}, \dots, X_{n,L}\}$, are therefore functions of $\{U_1, \dots, U_n\}$. The state feedback channel is assumed to be noiseless but with a finite cardinality constraint $Q = |\mathcal{U}|$ on the CSIT alphabets \mathcal{U} .³ In general, the state feedback U may not be uniformly distributed and the entropy $H(U) \leq \log_2(Q)$. We define the feedback rate as $R_{fb} = \log_2(Q)$ (bits per fading block). In this paper, we do not consider rate loss due to pilot symbols for the channel estimation and ignore feedback delay due to channel estimation and feedback transmission. Asymptotically, these overheads are negligible if each fading block is long enough. When we generalize our system model to the continuous case latter, the state feedback will remain as a discrete random variable with finite cardinality.

C. Relevant Channel Capacity Results

In this subsection, we quote relevant capacity results under the assumption that the relation among the channel state \mathbf{S} , the CSIR \mathbf{V} , and the CSIT \mathbf{U} is given and is i.i.d., i.e.,

$$p(\mathbf{s}, \mathbf{u}, \mathbf{v}) = \prod_n p_{S,U,V}(s_n, u_n, v_n).$$

(The random variables S_n , V_n , and U_n do not have to be a Markov chain.) Section III studies the capacity when we have further freedom to design the relation between \mathbf{V} and \mathbf{U} .

The memoryless channel capacity when there is perfect channel state information available at transmitter only has been studied by Shannon [2]. The result is extended by Caire and Shamai in [14] and proved directly in [10] for the case of imperfect CSIR in addition to imperfect CSIT. The extension is done by hiding the channel state S , viewing CSIT U as the channel state, and viewing the CSIR V as part of the channel output. This is illustrated in Fig. 2.

Although block-fading channel has memory within one fading block, the following argument given in [22] makes it

³In practical systems such as UMTS-HSDPA [21], we have a control field carrying a certain number of state feedback bits on a per-fading block basis. This limitation on the maximum number of state feedback bits per fading block translates into a cardinality constraint on state feedback alphabets.

possible to apply the results of memoryless channel capacity directly. If we view the L symbols received in one fading block as outputs of L parallel channels, then we have a memoryless vector channel. Considering one vector channel use as L channel uses and applying the capacity result in [14] directly, the capacity of the block fading channel is

$$\begin{aligned} C_{p(u,s,v)} &= \frac{1}{L} \max_{p(\mathbf{t})} I(\mathbf{T}; \mathbf{Y}, V) \\ &= \frac{1}{L} \max_{p(\mathbf{t})} I(\mathbf{T}; \mathbf{Y}|V) \end{aligned} \quad (1)$$

where $\mathbf{T} \in \mathcal{X}^{L \times |\mathcal{U}|}$ is a random matrix and $\mathbf{Y} \in \mathcal{Y}^L$ is the received signals in a fading block. It is the capacity of an equivalent discrete memoryless channel without CSIT or CSIR and with input \mathbf{T} and output \mathbf{Y} and V . The subscript of $C_{p(u,s,v)}$ indicates that $p(u, s, v)$ is given.

The capacity-achieving coding scheme is as follows. A codeword spanning N fading blocks is a sequence of matrices $\{\mathbf{T}_1^N\}$ constructed according to $p(\mathbf{t})$. Let $T_{n,l}(k)$ denote the element of l th row and k th column of \mathbf{T}_n . Without loss of generality, assume the feedback alphabet $\mathcal{U} = \{1, 2, \dots, Q\}$. At the l th symbol time of fading block n , $X_{n,l} = T_{n,l}(U_n)$ is transmitted and $Y_{n,l}$ and V_n are received. Let $Y_{n,l}$'s be the elements of \mathbf{Y}_n and $X_{n,l}$'s be the elements of \mathbf{X}_n . This coding scheme is depicted in Fig. 2. The channel transition probability for the calculation of (1) is

$$\begin{aligned} p(\mathbf{y}_n, v_n | \mathbf{t}_n) &= \sum_{u_n} p(u_n) p(\mathbf{y}_n, v_n | \mathbf{t}_n, u_n) \\ &= \sum_{u_n} p(u_n) \sum_{s_n} p(s_n, v_n | u_n) p(\mathbf{y}_n | \mathbf{t}_n, u_n, s_n) \\ &= \sum_{u_n} \sum_{s_n} p_{S,U,V}(s_n, u_n, v_n) \prod_{l=1}^L p_{Y|X,S}(y_{n,l} | t_{n,l}(u_n), s_n). \end{aligned}$$

For the case of perfect CSIR, i.e., $V = S$, the capacity in (1) takes a scalar form as

$$C_{p(u,s=v)} = \max_{p(t)} I(T; Y|S) \quad (2)$$

where $T \in \mathcal{X}^{|\mathcal{U}|}$ and $Y \in \mathcal{Y}$ with channel transition probability

$$p(y, s | t) = \sum_u p(s, u) p_{Y|X,S}(y | t(u), s).$$

Capacity $C_{p(u,s=v)}$ is the limiting case of the capacity $C_{p(u,s,v)}$ in (1) when L increases to infinity so that perfect CSIR can be obtained by the negligible overhead of sending pilot symbols.

In the case of imperfect CSIR, i.e., V may not equal to S , the capacity (1) is lower-bounded by

$$C_{p(s,u,v), L=1} = \max_{p(t)} I(T; Y|V) \quad (3)$$

where $T \in \mathcal{X}^{|\mathcal{U}|}$ and $Y \in \mathcal{Y}$ with channel transition probability

$$p(y, v | t) = \sum_u \sum_s p_{S,U,V}(s, u, v) p_{Y|X,S}(y | t(u), s).$$

This lower bound is achieved under the low-complexity constraint that $T_{n,l}(\cdot)$'s are independent within a fading block and the receiver does not attempt to improve channel estimation from other received signals within one fading block, i.e., the receiver regards the channel as i.i.d. even within a fading block, although it happens to receive the same v for every symbol of a fading block. The capacity (1) also equals to this lower bound when $L = 1$.

D. Error Exponent of the Forward Channel With Partial CSIT

While channel capacity deals with performance when we have infinite block length, error exponent is also an important performance measure for communication systems with finite block length [20], [7]. In this section, we shall derive the error exponent of the system with partial feedback.

Considering the equivalent discrete memoryless channel, the codeword error probability $P(e)$ can be upper-bounded as [20]

$$\begin{aligned} P(e) &\leq 2^{NLR\alpha} \sum_{\mathbf{y}_1^N, v_1^N} \left(\sum_{\mathbf{t}_1^N} p(\mathbf{t}_1^N) p(\mathbf{y}_1^N, v_1^N | \mathbf{t}_1^N)^{\frac{1}{1+\alpha}} \right)^{1+\alpha} \\ &= 2^{NLR\alpha} \left(\sum_v p(v) \sum_{\mathbf{y}} \left(\sum_{\mathbf{t}} p(\mathbf{t}) p(\mathbf{y} | \mathbf{t}, v)^{\frac{1}{1+\alpha}} \right)^{1+\alpha} \right)^N \end{aligned}$$

for $0 \leq \alpha \leq 1$. Therefore,

$$P(e) \leq 2^{-NLE(R)}$$

where $E(R)$ is the error exponent given by

$$E(R) = \max_{p(\mathbf{t}), 0 \leq \alpha \leq 1} [E_0(\alpha, p(\mathbf{t})) - \alpha R]$$

and

$$\begin{aligned} E_0(\alpha, p(\mathbf{t})) &= -\frac{1}{L} \log_2 \left(\sum_v p(v) \sum_{\mathbf{y}} \left(\sum_{\mathbf{t}} p(\mathbf{t}) p(\mathbf{y} | \mathbf{t}, v)^{\frac{1}{1+\alpha}} \right)^{1+\alpha} \right). \end{aligned}$$

The Gallager function E_0 can be simplified for several cases. If $u = h(v)$ is a deterministic function, it can be simplified to (4) at the top of the following page. Equation (4) is valid even if u is not a deterministic function of v because an upper bound of $P(e)$ can be derived by averaging the upper bound of $P(e | u_1^N, v_1^N)$. If in addition, CSIR is perfect, i.e., $V = S$ and $u = h(s)$, it can be simplified to (5) at the top of the following page. For continuous alphabets under mild regularity, the error exponents formula can be obtained by replacing the summations with integrations [20].

III. OPTIMAL STATE FEEDBACK STRATEGY WITH FEEDBACK CARDINALITY CONSTRAINT

The main purpose of this paper is to study the performance limits and the design of transmission and feedback strategies under the constraints that the state feedback U is discrete with finite cardinality $Q = |\mathcal{U}|$. Section III-A provides a coding theorem and shows how to achieve the capacity. In

$$E_0(\alpha, p(\mathbf{x}|u)) = -\frac{1}{L} \log_2 \left(\sum_v p(v) \sum_u p(u|v) \sum_{\mathbf{y}} \left(\sum_{\mathbf{x}} p(\mathbf{x}|u) p(\mathbf{y}|\mathbf{x}, v)^{\frac{1}{1+\alpha}} \right)^{1+\alpha} \right). \quad (4)$$

$$E_0(\alpha, p(x|u)) = -\frac{1}{L} \log_2 \left(\sum_s p(s) \sum_u p(u|s) \left[\sum_{\mathbf{y}} \left(\sum_x p(x|u) p(\mathbf{y}|x, s)^{\frac{1}{1+\alpha}} \right)^{1+\alpha} \right]^L \right). \quad (5)$$

Section III-B, the Lloyd algorithm is employed to solve the optimization problem. Section III-C presents an example for Rayleigh block-fading channels.

A. The Channel Capacity

We investigate the channel capacity when we have the freedom to design the transmission and feedback strategies.

The feedback design means that we design the channel state feedback \mathbf{U} based on the channel state estimate \mathbf{V} . This implies that $\mathbf{S} \rightarrow \mathbf{V} \rightarrow \mathbf{U}$ forms a Markov chain. As described in Section II-B, the relation among \mathbf{S} , \mathbf{V} , and \mathbf{U} is

$$p(\mathbf{s}, \mathbf{v}, \mathbf{u}) = \prod_n p(s_n) p(v_n | s_n) p(\mathbf{u} | \mathbf{v}).$$

Note that it is not clear whether an i.i.d. relation between U_n and V_n is optimal when we have the freedom to design the feedback strategy. However, in Coding Theorem 1, we show that an identical deterministic relation $p(\mathbf{u} | \mathbf{v}) = \prod_n \delta(u_n - h(v_n))$ achieves the capacity.

The transmission strategy design means that we design the relation $p(\mathbf{X}_1^N | W, U_1^N)$ between a message W and the signal \mathbf{X}_1^N to convey the message over N fading blocks based on the causal state feedback U_1^N .

Note that when the channel state feedback can be designed, it is not clear whether the transmitted signal \mathbf{X}_1^N has to be a deterministic function of U_1^N , or whether \mathbf{X}_n is related to the past state feedback U_1^{n-1} . The coding theorem shows that a simple memoryless device identical to Shannon's device [2] discussed in Section II suffices to achieve the capacity.

The following mutual information identity is useful in the proof of the lemma and the coding theorem. The unconditional mutual information $I(A; B)$ and the conditional mutual information $I(A; B|C)$ is related by the identity

$$I(A; B) = I(A; B|C) + I(A; C) - I(A; C|B). \quad (6)$$

This identity is obtained by expanding $I(B, C; A)$ using the chain rule of mutual information, i.e.,

$$\begin{aligned} I(B, C; A) &= I(B; A) + I(C; A|B) \\ &= I(C; A) + I(B; A|C). \end{aligned}$$

To prove the coding theorem, we first prove a simple lemma. The further implication of the lemma is discussed in the remark following the theorem.

Lemma 1: For every statistical relation $p(u|v)$, we can always find a deterministic relation $u = h(v)$ such that

$$I(\mathbf{T}; \mathbf{Y} | V) = I(\mathbf{T}; \mathbf{Y} | U, V) - I(U; \mathbf{T} | \mathbf{Y} V) \quad (7)$$

$$\leq \sum_v p(v) I(\mathbf{T}; \mathbf{Y} | u = h(v), v). \quad (8)$$

Proof: Employing the identity (6), we have

$$\begin{aligned} I(\mathbf{T}; \mathbf{Y} | V) &= I(\mathbf{T}; \mathbf{Y} | U, V) + I(U; \mathbf{T} | V) - I(U; \mathbf{T} | \mathbf{Y} V) \\ &= I(\mathbf{T}; \mathbf{Y} | U, V) - I(U; \mathbf{T} | \mathbf{Y} V) \end{aligned} \quad (9)$$

where we have used the fact $I(U; \mathbf{T} | V) = 0$ because U and \mathbf{T} are independent given V .

The first term of (9) can be upper-bounded as

$$\begin{aligned} I(\mathbf{T}; \mathbf{Y} | U, V) &= \sum_v p(v) \sum_u p(u|v) I(\mathbf{T}; \mathbf{Y} | u, v) \\ &\leq \sum_v p(v) \max_u I(\mathbf{T}; \mathbf{Y} | u, v). \end{aligned}$$

If $u = h(v)$, $I(U; \mathbf{T} | \mathbf{Y} V) = 0$. Therefore,

$$u = h(v) \triangleq \arg \max_u I(\mathbf{T}; \mathbf{Y} | u, v)$$

simultaneously maximize the first term of (9) and minimize the second term of (9).⁴ \square

The following coding theorem gives the channel capacity when we have the freedom to design feedback and transmission strategies. The achievability part shows how to achieve the capacity with a deterministic memoryless feedback function and a deterministic memoryless device at the transmitter to select the transmitted symbol from a codeword with expanded alphabet. Let $\mathcal{E}_{v \in \mathcal{R}}[g(v)]$ denote $\sum_{v \in \mathcal{R}} p(v) g(v)$.

Theorem 1: Consider a discrete memoryless channel having input $\mathbf{X} \in \mathcal{X}^L$ and output $\mathbf{Y} \in \mathcal{Y}^L$, $V \in \mathcal{V}$. The channel transition probability is $p(\mathbf{y}, v | \mathbf{x}) = p(v) p(\mathbf{y} | \mathbf{x}, v)$, i.e., the input \mathbf{X} and the output V are independent. For the n th channel use, the transmitter receives a delay-less channel state feedback $U_n \in \mathcal{U}$ from the transmitter. The cardinality constraint is $|\mathcal{U}| = Q$. The feedback U_1^n is related to the channel output V_1^n through a causal relation $p(u_1^n | v_1^n)$. The channel output \mathbf{Y}_n is conditionally independent of the sequence $\mathbf{U} = \{U_1, U_2, \dots\}$ given V_n and \mathbf{X}_n , i.e., $p(\mathbf{y}_n | \mathbf{x}_n, v_n, \mathbf{u}) = p_{\mathbf{Y} | \mathbf{X}, V}(\mathbf{y}_n | \mathbf{x}_n, v_n)$. When the relation $p(u_1^n | v_1^n)$ can be designed and we count a vector channel use as L channel uses, the capacity of this channel is

$$C = \frac{1}{L} \max_{\mathcal{R}_1^Q} \sum_{q=1}^Q \max_{p_{\mathbf{X} | U}(\mathbf{x} | u_q)} \mathcal{E}_{v \in \mathcal{R}_q} [I(\mathbf{X}; \mathbf{Y} | u_q, v)] \quad (10)$$

where \mathcal{R}_1^Q is a partition of space \mathcal{V} . The partition is related to function $u = h(v)$ through $\mathcal{R}_q = \{v : h(v) = u_q\}$ where $u_q \in \mathcal{U}$. The maximizations are over all the possible partitions, or equivalently all the possible deterministic memoryless functions $u = h(v)$ and distributions $\{p_{\mathbf{X} | U}(\mathbf{x} | u_q), q = 1, \dots, Q\}$.

⁴The upper bound can also be proved using the convexity of the mutual information as a function of the channel transition probability $p(y|v, \mathbf{t}) = \sum_u p(u|v) p(y|u, v, \mathbf{t})$ [5].

Proof: Achievability. Let $h^*(v)$ be the function corresponding to the optimal partition in (10). Let the feedback be $U_n = h^*(V_n)$. It reduces the channel to a memoryless channel with the transmitter side information. This case is studied in [2], [14] and the capacity-achieving coding method is quoted in Section II. A memoryless device

$$\{\mathbf{X}_n = \mathbf{T}_n(U_n), \mathbf{T}_n \in \mathcal{X}^{L \times |U|}\}$$

produces the transmitted symbol \mathbf{X}_n according to current feedback U_n . The equivalent channel $p(\mathbf{y}, v | \mathbf{t})$ is a discrete memoryless channel (DMC) without side information. The capacity can be achieved according to the DMC coding theorem [5]. The code book is constructed according to $p(\mathbf{t}(u_q)) = p_{\mathbf{X}|U}^*(\mathbf{t}(u_q) | u_q)$, where $p_{\mathbf{X}|U}^*(\cdot | \cdot)$ is the optimal distribution in (10).

Converse. The converse is proved using Fano's lemma [5]. Let the message be $W \in \mathcal{W}$ with an uniform distribution and $|\mathcal{W}| = 2^{LN}$. The message is related to the channel input \mathbf{X}_1^N and the corresponding channel output \mathbf{Y}_1^N . Let \hat{W} be the detected message at the receiver and define the error probability as $P_e^{(LN)} = \Pr\{\hat{W} \neq W\}$. The rate R satisfies

$$\begin{aligned} LNR &= H(W) \\ &\stackrel{(a)}{=} H(W | \mathbf{Y}_1^N, V_1^N) + I(W; \mathbf{Y}_1^N, V_1^N) \\ &\stackrel{(b)}{=} H(W | \mathbf{Y}_1^N, V_1^N) + I(W; \mathbf{Y}_1^N | V_1^N) \end{aligned} \quad (11)$$

where (a) follows from the definition of the mutual information, (b) follows from the independence between W and V_1^N . By Fano's lemma [5], the first term of (11) is upper-bounded as

$$H(W | \mathbf{Y}_1^N, V_1^N) \leq 1 + P_e^{(LN)} LNR. \quad (12)$$

In general, \mathbf{X}_n and W are related through $p(\mathbf{x}_n | w, u_1^n)$ due to the causality of the feedback. A key step in the proof of the converse is to introduce an intermediate random variable \mathbf{T}_n with conditional distribution $p(\mathbf{t}_n | w, u_1^{n-1})$ and assume that \mathbf{X}_n is independent of W and U_1^{n-1} given \mathbf{T}_n and U_n , i.e., $p(\mathbf{x}_n | w, u_1^n, \mathbf{t}_n) = p(\mathbf{x}_n | u_n, \mathbf{t}_n)$. The assumption has no loss of generality because we can always let $\mathbf{T}_n = \{W, U_1^{n-1}\}$ to restore the original relation.

The second term of (11) can be upper-bounded as

$$\begin{aligned} I(W; \mathbf{Y}_1^N | V_1^N) &\stackrel{(a)}{=} I(W; \mathbf{Y}_1^N | V_1^N, U_1^N) + I(U_1^N; W | V_1^N) \\ &\quad - I(U_1^N; W | \mathbf{Y}_1^N, V_1^N) \\ &\stackrel{(b)}{\leq} I(W; \mathbf{Y}_1^N | V_1^N, U_1^N) \\ &\stackrel{(c)}{\leq} \sum_{i=1}^N I(W; \mathbf{Y}_i | V_1^N, U_1^N, \mathbf{Y}_1^{i-1}) \\ &\stackrel{(d)}{\leq} \sum_{i=1}^N I(\mathbf{T}_i; \mathbf{Y}_i | V_1^N, U_1^N, \mathbf{Y}_1^{i-1}) \\ &= \sum_{i=1}^N H(\mathbf{Y}_i | V_1^N, U_1^N, \mathbf{Y}_1^{i-1}) \\ &\quad - H(\mathbf{Y}_i | V_1^N, U_1^N, \mathbf{Y}_1^{i-1}, \mathbf{T}_i) \\ &\stackrel{(e)}{\leq} \sum_{i=1}^N H(\mathbf{Y}_i | V_i, U_i) \end{aligned} \quad (13)$$

$$\begin{aligned} &- H(\mathbf{Y}_i | V_1^N, U_1^N, \mathbf{Y}_1^{i-1}, \mathbf{T}_i) \\ &\stackrel{(f)}{=} \sum_{i=1}^N H(\mathbf{Y}_i | V_i, U_i) - H(\mathbf{Y}_i | V_i, U_i, \mathbf{T}_i) \\ &= \sum_{i=1}^N I(\mathbf{T}_i, \mathbf{Y}_i | V_i, U_i) \\ &\stackrel{(g)}{\leq} \sum_{i=1}^N \max_{h(\cdot)} \sum_{v_i} p(v_i) I(\mathbf{T}_i, \mathbf{Y}_i | v_i, u_i = h(v_i)) \\ &\stackrel{(h)}{\leq} \sum_{i=1}^N \max_{h(\cdot)} \sum_{v_i} p(v_i) I(\mathbf{X}_i, \mathbf{Y}_i | v_i, u_i = h(v_i)) \\ &\stackrel{(i)}{\leq} LNC \end{aligned} \quad (14)$$

where (a) follows from the mutual information identity (6), (b) follows from the independence between U_1^N and W given V_1^N and the nonnegativity of mutual information $I(U_1^N; W | \mathbf{Y}_1^N, V_1^N)$, (c) follows from the chain rule of the mutual information, (d) follows from the data processing inequality [5] and that $W \rightarrow \mathbf{T}_i \rightarrow \mathbf{Y}_i$ is a Markov chain given $\{V_1^N, U_1^N, \mathbf{Y}_1^{i-1}\}$, (e) follows from the fact that the conditioning reduces the entropy, (f) follows from the fact that \mathbf{Y}_i is independent from the other variables given $\{V_i, U_i, \mathbf{T}_i\}$, (g) follows from Lemma 1, (h) follows from the fact that $\mathbf{T}_i \rightarrow \mathbf{X}_i \rightarrow \mathbf{Y}_i$ is a Markov chain given u_i, v_i , and (i) follows from (10).

Combining (11), (12), and (14), we obtain

$$R \leq \frac{1}{LN} + P_e^{(LN)} R + C.$$

If $P_e^{(LN)} \rightarrow 0$ as $N \rightarrow \infty$, then the rate R must be less than or equal to C . \square

Remark 1: If there is memoryless feedback error, the relation between U_n and V_n is an i.i.d. statistical relation $p_{U|V}(u_n | v_n)$. Comparing (7) of Lemma 1 and the capacity C in (10), we observe that the feedback error results in two penalties. The first penalty is due to averaging, i.e.,

$$\sum_u p(u|v) I(\mathbf{T}; \mathbf{Y} | u, v) \leq \max_u I(\mathbf{T}; \mathbf{Y} | u, v)$$

and the second penalty is due to the fact that the receiver is uncertain about U . Therefore, Lemma 1 will be useful for the design of the index assignment of the feedback when there is feedback error.

Remark 2: Theorem 1 is proved with the assumption that the receiver does not know the feedback U . Of course, since the optimal feedback is a deterministic function of V , the receiver knows the feedback implicitly. The capacity is the same even if we start by assuming the receiver knows the feedback. In this case, the proof of the converse part will start from (13). The rest of the proof is the same.

For the case of perfect CSIR $V = S$, similar to $C_{p(u,s=v)}$ in (2), the capacity C in (10) can be simplified to a scalar form as

$$C_{V=S} = \max_{\mathcal{R}_1^Q} \sum_{q=1}^Q \max_{p_{X|U}(x|u_q)} \mathcal{E}_{s \in \mathcal{R}_q} [I(X; Y | u_q, s)]. \quad (15)$$

For the case of $L = 1$ or the case of the low-complexity receiver, similar to $C_{p(u,s,v),L=1}$ in (3), the capacity C in (10) can be simplified to a scalar form as

$$C_{L=1} = \max_{\mathcal{R}_1^Q} \sum_{q=1}^Q \max_{p_{X|U}(x|u_q)} \mathcal{E}_{v \in \mathcal{R}_q} [I(X; Y|u_q, v)]. \quad (16)$$

In applications, we might have additional constraints on the input distribution of X . For example, a common constraint is the average power of codewords. In this case, we need to choose $p(\mathbf{x}|u)$ to maximize the mutual information under the constraint $\mathcal{E}[|\mathbf{X}|^2] \leq LP_0$. Section III-C will give examples of fading channels with power constraints.

To characterize the benefit of the state feedback, the following two optimization problems need to be solved.

Problem 1: Find the feedback strategy with partition $\{\mathcal{R}_1, \dots, \mathcal{R}_Q\}$ and the transmission strategy $\{p(\mathbf{x}|u_1), \dots, p(\mathbf{x}|u_Q)\}$, to achieve the capacities in (10), (15), and (16).

We are especially interested in the simpler case of perfect CSIR ($V = S$) and the case of low-complexity receiver.

Problem 2: Find a feedback strategy with partition $\{\mathcal{R}_1, \dots, \mathcal{R}_Q\}$ and transmission strategy $\{p(\mathbf{x}|u_1), \dots, p(\mathbf{x}|u_Q)\}$, such that the Gallager function E_0 is maximized. For the general case, see the first equation at the bottom of the page. For the case of perfect CSIR ($V = S$), according to (5), $E_0(\alpha)$ reduces to (17) at the bottom of the page.

B. The Lloyd Algorithm

We shall illustrate that the design of the optimal feedback strategy $\{\mathcal{R}_1, \dots, \mathcal{R}_Q\}$ and the optimal transmission strategy $\{p(\mathbf{x}|u_1), \dots, p(\mathbf{x}|u_Q)\}$ are equivalent to the design of a scalar quantizer with a modified distortion measure.

Forward Channel Capacity: Define the *distortion measure* for forward channel capacity as

$$d_{\text{cap}}(v, p(\mathbf{x}|u_q)) = -I(\mathbf{X}; \mathbf{Y}|V = v, U = u_q). \quad (18)$$

Solving Problem 1 is equivalent to selecting $\{p(\mathbf{x}|u_1), \dots, p(\mathbf{x}|u_Q)\}$ and $\{\mathcal{R}_1, \dots, \mathcal{R}_Q\}$ to minimize the *average distortion*

$$D_{\text{cap}} = \sum_{q=1}^Q \mathcal{E}_{v \in \mathcal{R}_q} [d_{\text{cap}}(v, p(\mathbf{x}|u_q))].$$

Forward Channel Error Exponent: Define the *distortion measure* for the forward channel error exponent as

$$d_{\text{exp}}(v, p(\mathbf{x}|u_q)) = \sum_{\mathbf{y}} \left(\sum_{\mathbf{x}} p(\mathbf{x}|u_q) p(\mathbf{y}|\mathbf{x}, v)^{\frac{1}{1+\alpha}} \right)^{1+\alpha}. \quad (19)$$

Solving Problem 2 is equivalent to selecting $\{p(\mathbf{x}|u_1), \dots, p(\mathbf{x}|u_Q)\}$ and $\{\mathcal{R}_1, \dots, \mathcal{R}_Q\}$ so as to minimize the *average distortion*

$$D_{\text{exp}} = \sum_{q=1}^Q \mathcal{E}_{v \in \mathcal{R}_q} [d_{\text{exp}}(v, p(\mathbf{x}|u_q))]$$

where we have used the monotonicity of $\log_2(\cdot)$. For perfect CSIR ($V = S$), according to (17), the distortion measure (19) can be simplified to

$$d_{\text{exp}}(v, p(\mathbf{x}|u_q)) = \left(\sum_{\mathbf{y}} \left(\sum_{\mathbf{x}} p(\mathbf{x}|u_q) p(\mathbf{y}|\mathbf{x}, v)^{\frac{1}{1+\alpha}} \right)^{1+\alpha} \right)^L. \quad (20)$$

Hence, both optimization problems can be approximately solved by the Lloyd algorithm which is outlined in the following. For each initial configuration, Step 1 and Step 2 are repeated until convergence. Many initial configurations are chosen randomly to search for the optimal solution.

Step 1. Find optimal transmission strategy $\{p(\mathbf{x}|u_q)\}$ given partition $\{\mathcal{R}_1^Q\}$. The optimal transmission strategy $p(\mathbf{x}|u_q)$ is given by the generalized centroid $p(\mathbf{x}|u_q) = \arg \min_{p(\mathbf{x}|u_q)} \mathcal{E}_{v \in \mathcal{R}_q} [d(v, p(\mathbf{x}|u_q))]$.

Step 2. Find optimal partition $\{\mathcal{R}_1^Q\}$ given transmission strategy $\{p(\mathbf{x}|u_q)\}$. The optimal partition is given by the nearest neighbor rule $\mathcal{R}_q = \{v : d(v, p(\mathbf{x}|u_q)) \leq d(v, p(\mathbf{x}|u_j)); \forall j \neq q\}$.

C. Example: Rayleigh-Fading Channels

CSIR Estimation and the Equivalent Channel Model: We illustrate the feedback design by an example of SISO Rayleigh-fading channel. The multiple-antenna case is considered in [17]. All of the following complex Gaussian random variables are circularly symmetric. The received symbol $Y_{n,l}$ is given by $Y_{n,l} = S_n X_{n,l} + Z_{n,l}$ where S_n is the zero-mean, unit variance complex Gaussian fading coefficient during the n th fading block and $Z_{n,l}$ is the complex Gaussian noise with variance σ_z^2 . Furthermore, assume that the CSIR, $V_{n,l}$, is estimated from the preamble transmission once every fading block. The CSIR is

$$E_0(\alpha) = \max_{\{\mathcal{R}_1^Q\}} \frac{-1}{L} \log_2 \left(\sum_{q=1}^Q \min_{p(\mathbf{x}|u_q)} \mathcal{E}_{v \in \mathcal{R}_q} \left[\sum_{\mathbf{y}} \left(\sum_{\mathbf{x}} p(\mathbf{x}|u_q) p(\mathbf{y}|\mathbf{x}, v)^{\frac{1}{1+\alpha}} \right)^{1+\alpha} \right] \right).$$

$$E_0(\alpha) = \max_{\{\mathcal{R}_1^Q\}} \frac{-1}{L} \log_2 \left(\sum_{q=1}^Q \min_{p(\mathbf{x}|u_q)} \mathcal{E}_{s \in \mathcal{R}_q} \left[\left(\sum_{\mathbf{y}} \left(\sum_{\mathbf{x}} p(\mathbf{x}|u_q) p(\mathbf{y}|\mathbf{x}, s)^{\frac{1}{1+\alpha}} \right)^{1+\alpha} \right)^L \right] \right). \quad (17)$$

$V_n = S_n + \epsilon_n$, where ϵ_n represents the complex channel estimation error with variance σ_ϵ^2 . We assume the input distribution satisfies the average transmit power constraint $\mathcal{E}[|X|^2] \leq P_0$. The SNR is defined as P_0/σ_z^2 .

For imperfect CSIR V_n , we consider an equivalent channel which corresponds to a receiver treating V_n as a perfect channel state and suffering from additional noise. This is a practical situation and it simplifies the optimization problem to the case of perfect CSIR. The equivalent channel model is given by

$$Y_{n,l} = (V_{n,l} - \epsilon_n)X_{n,l} + Z_{n,l} = V_{n,l}X_{n,l} + Z'_{n,l}$$

where $Z'_{n,l} = Z_{n,l} - \epsilon_n X_{n,l}$ is the equivalent channel noise, which is not correlated with the signal $X_{n,l}$ for unbiased channel estimator, with variance $\sigma_0^2 = \sigma_z^2 + \sigma_\epsilon^2 \rho$ where ρ is the variance of $X_{n,l}$. The equivalent channel noise is non-Gaussian, and finding the optimizing input distribution $p(x|u_q)$ is non-trivial. Yet, in practice, with the coherent accumulation in the CSIR estimation, the estimation error variance σ_ϵ^2 is much smaller than the channel noise variance of the forward channel σ_z^2 . Since Gaussian noise is the worst case channel noise [20], we lower-bound the capacity (or upper-bound the distortion measure) by assuming the equivalent channel noise to be Gaussian and choosing the input distribution $p(x|u_q)$ in (18) to be zero-mean complex Gaussian distribution with variance (transmitted power) ρ_q . For perfect CSIR, Gaussian input is the optimal input distribution.

Distortion Measure for Forward Channel Capacity: Including the Lagrange multiplier λ for the average transmit power constraint, the corresponding distortion measure is given by

$$\begin{aligned} d_{\text{cap}}(v, p(x|u_q), \lambda) &= -[I(X; Y|v, u_q) - \lambda \mathcal{E}[|X|^2|u_q]/\ln(2)] \\ &\leq -\left[\log_2 \left(1 + \frac{|v|^2 \rho_q}{\sigma_\epsilon^2 \rho_q + \sigma_z^2} \right) - \lambda \rho_q / \ln(2) \right] \\ &\triangleq d_{\text{cap}}(v, \rho_q, \lambda). \end{aligned}$$

The upper bound on the distortion measure is asymptotically tight as $\frac{\sigma_\epsilon^2}{\sigma_z^2} \rightarrow 0$. We will minimize $d_{\text{cap}}(v, \rho_q, \lambda)$ in the following.

Transmission and Feedback Strategies for Forward Channel Capacity: The modified Lloyd algorithm for optimizing the transmission and feedback strategies to achieve the forward channel capacity of the block-fading channel is as follows.

Step 1. Find $\{\rho_q\}$ given partition $\{\mathcal{R}_q\}$: ρ_q is given by the solution of

$$\mathcal{E}_{v \in \mathcal{R}_q} \left[\frac{\partial d_{\text{cap}}(v, \rho_q, \lambda)}{\partial \rho_q} \right] = 0$$

which is simplified as

$$\mathcal{E}_{v \in \mathcal{R}_q} \left[\frac{|v|^2 \sigma_z^2}{(\sigma_z^2 + (\sigma_\epsilon^2 + |v|^2) \rho_q)(\rho_q \sigma_\epsilon^2 + \sigma_z^2)} \right] = \lambda P_q$$

where $P_q = \Pr\{V \in \mathcal{R}_q\}$. Solution to the above equation is quite messy. We shall further approximate the solution based on the following observation.

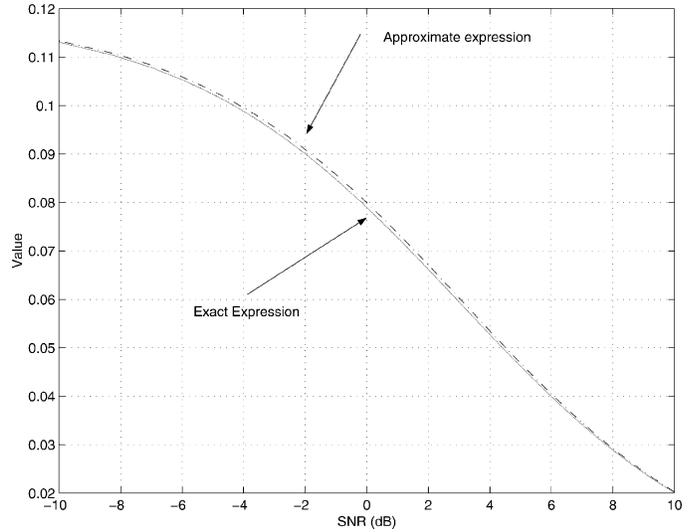


Fig. 3 Approximation of (21). The partition region is $[0.2, 0.7]$.

For sufficiently fine partition $\{\mathcal{R}_q\}$, we have

$$\begin{aligned} \mathcal{E}_{v \in \mathcal{R}_q} \left[\frac{|v|^2 \sigma_z^2}{(\sigma_z^2 + (\sigma_\epsilon^2 + |v|^2) \rho_q)(\rho_q \sigma_\epsilon^2 + \sigma_z^2)} \right] \\ \approx P_q \left[\frac{\overline{|V|^2}/P_q \sigma_z^2}{(\sigma_z^2 + (\sigma_\epsilon^2 + \overline{|V|^2}/P_q) \rho_q)(\rho_q \sigma_\epsilon^2 + \sigma_z^2)} \right] \end{aligned} \quad (21)$$

where $\overline{|V|^2} = \mathcal{E}_{v \in \mathcal{R}_q}[|v|^2]$.

In general, we observe that the preceding approximation is quite good even for moderate size partition. This is illustrated in Fig. 3, where “Value” denotes the value of (21). The error of the approximation is less than 0.1 dB. In any case, the approximated expression in (21) serves as a lower bound on the forward capacity. Moreover, if the iterative optimization ends in Step 2, the partition is optimal given $\{\rho_q\}$.

With the above approximation, the generalized centroid condition in Step 1 for perfect CSIR ($\sigma_\epsilon^2 = 0$) is

$$\rho_q = \left[\frac{1}{\lambda} - \frac{P_q \sigma_z^2}{\overline{|V|^2}} \right]^+$$

where $[x]^+ = \max(x, 0)$. Note that it has the form of water-filling. On the other hand, the generalized centroid for imperfect CSIR is obtained by one of the roots of (21).

Step 2. Find partition $\{\mathcal{R}_q\}$ given $\{\rho_q\}$: The optimal partition is given by the nearest neighbor rule

$$\mathcal{R}_q = \{v : d_{\text{cap}}(v, \rho_q, \lambda) \leq d_{\text{cap}}(v, \rho_j, \lambda); \forall j \neq q\}.$$

The resulting feedback and transmission strategy with feedback cardinality constraint Q is characterized by a set of partitions $\{\mathcal{R}_1, \dots, \mathcal{R}_Q\}$ and the corresponding set of transmission powers $\{\rho_1, \dots, \rho_Q\}$. It can be implemented by power control as illustrated in Fig. 4. At the start of each fading block, a preamble is transmitted and CSIR V is estimated at the receiver. The CSIR V is compared with the Q partitions and the one

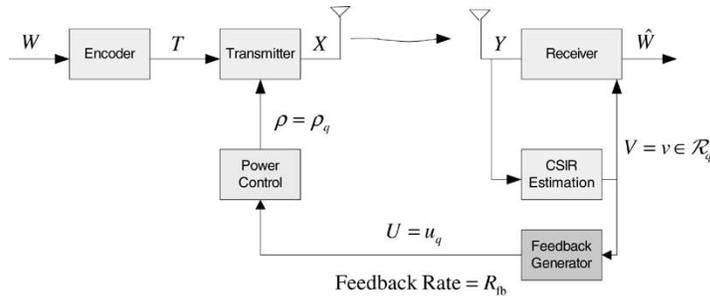


Fig. 4. Block diagram of the transmission and feedback strategies with state feedback cardinality constraint for Rayleigh-fading channels.

that contains V is selected. The corresponding CSIT index $U \in \{u_1, \dots, u_Q\}$ is fed back to the transmitter. If u_q is the feedback, the transmitter shall apply ρ_q as the transmission power for the current fading block. A single Gaussian code book can be used as discussed in [14]. The set of transmit powers $\{\rho_1, \dots, \rho_Q\}$ is set up at the transmitter prior to communication.

The average transmitted power over a coding block is $P_0 = \sum_{q=1}^Q P_q \rho_q$. Given the feedback strategy $\{\mathcal{R}_1, \dots, \mathcal{R}_Q\}$ and the transmission strategy $\{\rho_1, \dots, \rho_Q\}$, the corresponding forward channel capacity is

$$C_{u=h(v), p(x|u)} = \sum_{q=1}^Q \mathcal{E}_{v \in \mathcal{R}_q} \left[\log_2 \left(1 + \frac{|v|^2 \rho_q}{\sigma_\epsilon^2 \rho_q + \sigma_z^2} \right) \right].$$

Distortion Measure for Forward Error Exponent: Assuming Gaussian input $p(x|u_q)$ with variance ρ_q , considering V as perfect CSIR as in (20), and including the Lagrange multiplier λ for the average transmit power constraint, the corresponding distortion measure for error exponent is given by

$$\begin{aligned} d_{\text{exp}}(v, p(x|u_q)) &= \left[\int_y \left(\int_x p(x|u_q) p(y|x, v)^{1/(1+\alpha)} dx \right)^{(1+\alpha)} dy \right]^L \\ &\quad + \lambda \mathcal{E}[|x|^2 | u_q] \\ &= \left(1 + \frac{|v|^2 \rho_q}{(\sigma_\epsilon^2 \rho_q + \sigma_z^2)(1+\alpha)} \right)^{-\alpha L} + \lambda \rho_q \\ &\triangleq d_{\text{exp}}(v, \rho_q, \lambda). \end{aligned}$$

Optimal Feedback Strategy for Forward Error Exponent: Step 1 of the modified Lloyd algorithm for optimizing the transmission strategy with respect to the forward channel error exponent is to find ρ_q satisfying

$$\mathcal{E}_{v \in \mathcal{R}_q} \left[\frac{\partial d_{\text{exp}}(v, \rho_q, \lambda)}{\partial \rho_q} \right] = 0$$

which is as shown in the first equation at the bottom of the page. Using an approach similar to (21), the *generalized centroid* in Step 1 with perfect CSIR is shown in the second equation at the bottom of the page, which is again in the form of water-filling. The case of imperfect CSIR is solved numerically.

Step 2 of the Lloyd algorithm is similar to the part for channel capacity. Hence, when the optimal feedback strategy is characterized by $\{\mathcal{R}_q\}$ and $\{\rho_q\}$, the error exponent is

$$E_0(\alpha) = \frac{-1}{L} \log_2 \left(\sum_{q=1}^Q \mathcal{E}_{v \in \mathcal{R}_q} \left(1 + \frac{|v|^2 \rho_q}{(\sigma_\epsilon^2 \rho_q + \sigma_z^2)(1+\alpha)} \right)^{-\alpha L} \right).$$

IV. NUMERIC RESULTS AND DISCUSSIONS

We illustrate the performance limits, capacity, and error exponents, in terms of the noiseless channel state feedback alphabet cardinality $Q = 2^{R_{\text{fb}}}$ with numerical results in this section. The feedback rate is R_{fb} bits per fading block. We consider both *perfect CSIR* and *imperfect CSIR* in the following subsections.

A. Perfect CSIR

We first consider the case of *perfect CSIR* (i.e., $\sigma_\epsilon^2 = 0$). Fig. 5(a) illustrates the forward channel capacity versus the $\text{SNR} = P_0/\sigma_z^2$ with feedback rate constraint varying from $R_{\text{fb}} = 0$ to $R_{\text{fb}} = \infty$. When feedback rate $R_{\text{fb}} = 0$, it corresponds to the case without CSIT. When $R_{\text{fb}} = 1$, the optimization converges to on/off power control. When the feedback rate $R_{\text{fb}} = \infty$, it corresponds to the case with perfect CSIT. In the low-SNR region, as illustrated in Fig. 5(b), there is a significant gain of around 2.5 dB in forward channel capacity between perfect CSIT and no CSIT. Furthermore, there is a gain of 2.1 and 2.3 dB for partial feedback with feedback rate $R_{\text{fb}} = 1$ and $R_{\text{fb}} = 2$, respectively. In other words, $R_{\text{fb}} = 2$ realized about 93% of the feedback gain.

$$\mathcal{E}_{v \in \mathcal{R}_q} \left[(\alpha L) \left(1 + \frac{|v|^2 \rho_q}{(\sigma_\epsilon^2 \rho_q + \sigma_z^2)(1+\alpha)} \right)^{-\alpha L - 1} \left(\frac{\sigma_z^2 |v|^2}{(1+\alpha)(\sigma_z^2 + \rho_q \sigma_\epsilon^2)^2} \right) \right] = \lambda P_q.$$

$$\rho_q = \left(\frac{(\alpha+1)\sigma_z^2 P_q}{|V|_q^2} \right)^{\frac{\alpha L}{\alpha L + 1}} \left[\left(\frac{\alpha L}{\lambda} \right)^{1/(\alpha L + 1)} - \left(\frac{(\alpha+1)\sigma_z^2 P_q}{|V|_q^2} \right)^{1/(\alpha L + 1)} \right]^+$$

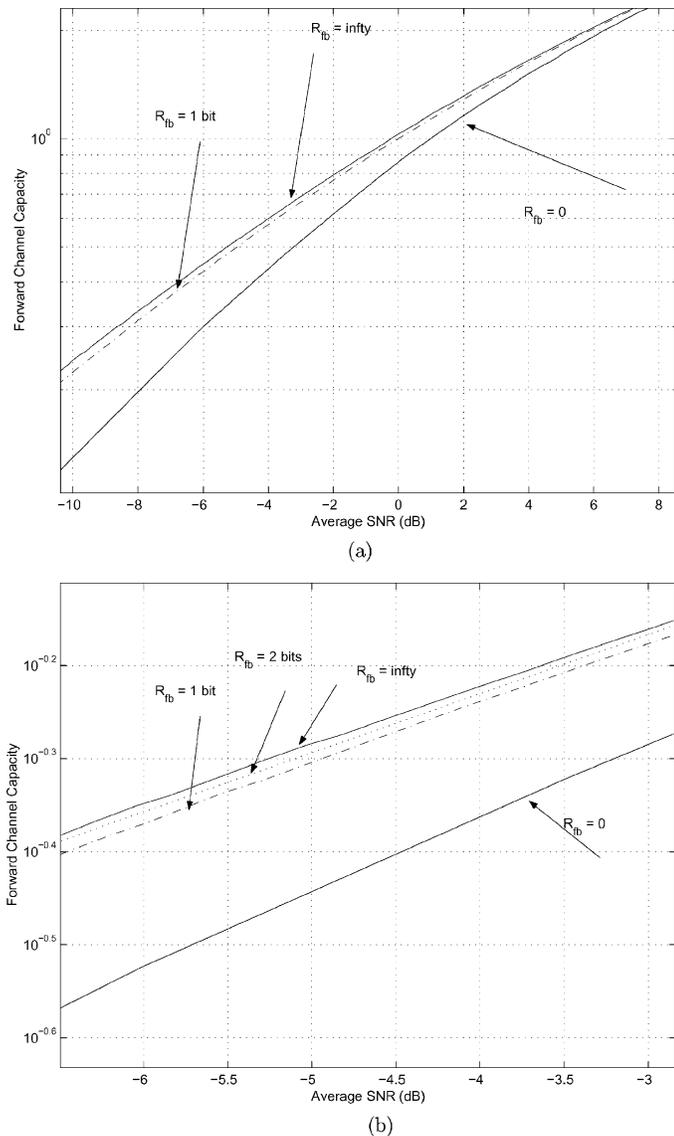


Fig. 5. Forward capacity versus state feedback cardinality for perfect CSIR case. (a) Forward channel capacity versus SNR for various feedback rate $R_{fb} \in \{0, 1, \infty\}$. (b) Feedback capacity at low-SNR range for $R_{fb} \in \{0, 1, 2, \infty\}$.

On the other hand, there is small SNR gain in the forward capacity in the high-SNR region. This is reasonable because when the SNR is large, the penalty of transmitting power less efficiently is small compared to the case when the SNR is small. This is illustrated in Fig. 5(a). Hence, the performance of feedback with respect to forward channel capacity is significant in the low-SNR region only.

The corresponding forward channel error exponents for large L with $R_{fb} = 0, 1, 2, \infty$ are illustrated in Fig. 6(a) and (b) for the high-SNR (10 dB) and the low-SNR (-6 dB) regions, respectively. Unlike the forward channel capacity, feedback performance gain is significant in both the high- and low-SNR regions. This implies that even in high-SNR region, state feedback introduces reasonable gain in practical systems although the channel capacity gain is limited. It is the same as in the case of channel capacity when a few bits of feedback is sufficient to achieve most of the gain.

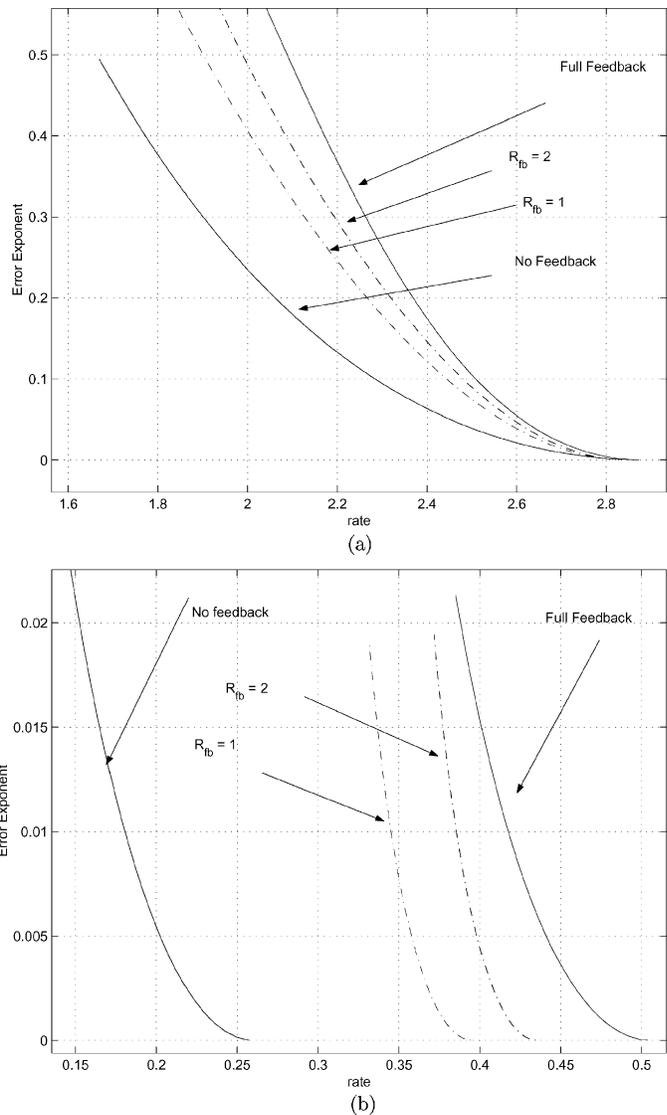


Fig. 6. Error exponents of forward channel with ideal CSIR. a) Error exponents for large L for various feedback channel rate $R_{fb} \in \{0, 1, 2, \infty\}$ at SNR = 10 dB; b) forward error exponent for various feedback channel rate $R_{fb} \in \{0, 1, 2, \infty\}$ at SNR = -6 dB.

B. Imperfect CSIR

In this subsection, we consider the performance with imperfect CSIR. We express the quality of the CSIR by estimation error-to-noise ratio $\xi = \sigma_\epsilon^2 / \sigma_z^2$. Fig. 7(a) illustrates the forward channel capacity with feedback rate constraint varying from $R_{fb} = 0$ to $R_{fb} = \infty$ and CSIR estimation error-to-noise ratio $\xi = \sigma_\epsilon^2 / \sigma_z^2$ being $\xi = 0$ and $\xi = 0.75$. CSIR estimation noise degrades the forward channel capacity for various R_{fb} . Furthermore, the SNR gain of full feedback over the case with no feedback is decreased when CSIR estimation noise is increased. This is because of the increased penalty $\sigma_\epsilon^2 \rho_q$ due to large transmit power.

The performance of forward error exponent with various feedback rates $R_{fb} = \{0, 1, 2, \infty\}$ and CSIR estimation-error-to-noise ratio $\xi = \{0, 0.75\}$ is illustrated in Fig. 7(b). The error exponent curves all shift to the left when the estimation noise is increased. Similar to the forward capacity, the error ex-

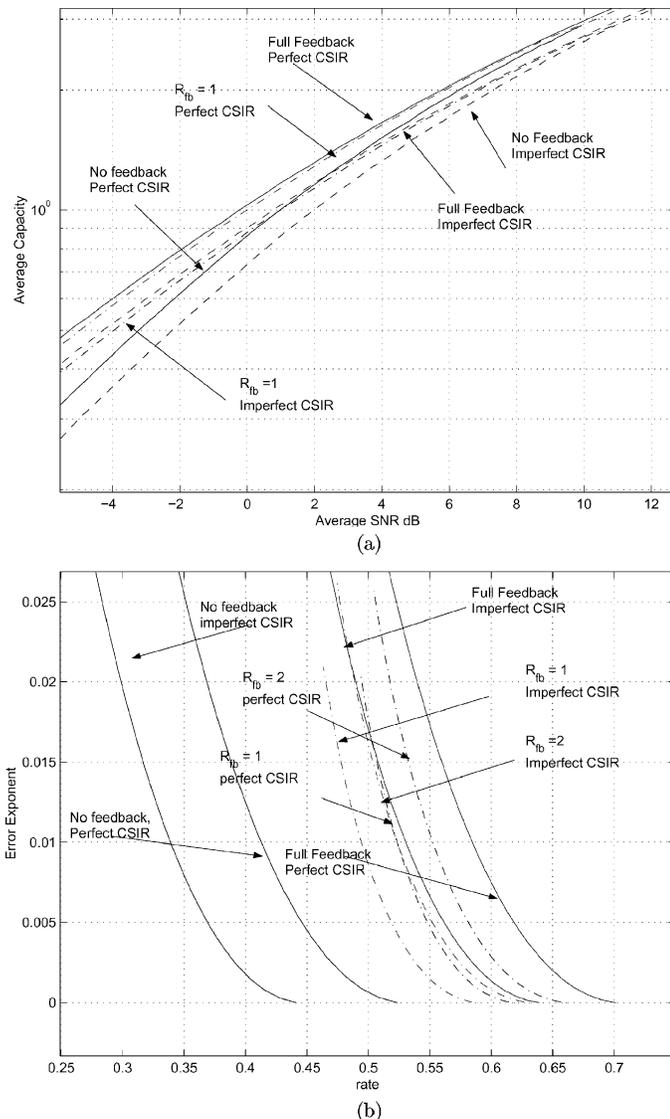


Fig. 7. Feedback gain with imperfect CSIR at various feedback rates $R_{fb} \in \{0, 1, \infty\}$ and various CSIR estimation noise ratios $\xi \in \{0, 0.75\}$. (a) Forward channel capacity versus SNR. (b) Feedback gain with imperfect CSIR at various feedback rates $R_{fb} \in \{0, 1, \infty\}$ and various CSIR estimation noise ratios $\xi \in \{0, 0.75\}$.

ponent gain of full feedback over the case with no feedback decreases as the estimation noise increases.

V. CONCLUSION

In this paper, we find the capacity of a memoryless channel when we have the freedom to design the channel state feedback which has finite cardinality Q . To calculate the capacity, we consider the optimization of transmission and channel state feedback strategies. The Lloyd algorithm is employed to approximately solve the problem. Numeric examples are provided for Rayleigh block-fading channels. Channel state information is estimated at the receiver and fed back to the transmitter. The feedback is assumed to be noiseless and causal. Due to the causal constraint, no block encoding over all the fading realizations can be done in the feedback channel. Hence, the feedback channel is characterized by a feedback rate constraint $R_{fb} =$

$\log_2(Q)$, which is the number of feedback bits per fading block. Both the forward capacity and error exponent are considered. We show that the design of the optimal transmission and feedback scheme is identical to the design of scalar quantizer with a modified *distortion measure*.

For Rayleigh-fading channels, it is shown that the transmission strategy has a general form of temporal water-filling in some cases. Furthermore, while feedback enhances the forward channel capacity more effectively in low-SNR region compared to that of high-SNR region, the performance gain of the error exponent is significant in both SNR regions. Note that it is possible to obtain good CSIR and thus, meaningful channel state feedback in low-SNR region. Theoretically, when the fading block size approaches infinity, the CSIR can be obtained with negligible training overhead.⁵ In practice, for example, a code-division multiple-access (CDMA) cellular system downlink has a separate strong pilot channel for channel estimation and other purposes. From the pilot channel, good CSIR can be obtained even when the data channel has low SNR. There also exist systems, e.g., systems with more transmit antennas than receive antennas, for which the feedback gain in capacity is significant in both low- and high-SNR regions. Therefore, the above suggests that reasonable benefit from channel state feedback is possible in the design of practical systems at all SNR regions.

ACKNOWLEDGMENT

The authors wish to thank the Associate Editor Giuseppe Caire and the anonymous reviewers for their constructive suggestions. This paper has benefited significantly from their outstanding work.

REFERENCES

- [1] S. W. Lei and V. K. N. Lau, "Performance analysis of adaptive interleaving for OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 51, pp. 435–444, May 2002.
- [2] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Devel.*, pp. 289–293, 1958.
- [3] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 359–378, May 1994.
- [4] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2619–2692, Oct 1998.
- [5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 1991.
- [6] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1986–1992, Nov. 1997.
- [7] K. N. Lau, "Channel capacity and error exponent of variable rate adaptive channel coding," *IEEE Trans. Commun.*, vol. 43, pp. 311–316, Apr. 1999.
- [8] —, "Variable rate adaptive channel coding for DS-CDMA," *Bell Labs Tech. J.*, vol. 3, pp. 311–316, Nov. 2000.
- [9] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1468–1489, July 1999.
- [10] M. Salehi, "Capacity and coding for memories with real-time noisy defect information at encoder and decoder," *Proc. Inst. Elec. Eng. Pt. I—Commun., Speech and Vision*, vol. 139, pp. 113–117, Apr. 1992.

⁵This does not conflict with the conclusion that half of the coherence time interval is needed for training in low-SNR region [23]. Comparing [23, eqs. (40) and (42)], this conclusion is drawn with the assumption that the product of the coherence time and the SNR is much less than 1. For any fixed SNR, if the coherence time approaches infinity, the time spent on training is negligible.

- [11] U. Erez and R. Zamir, "Noise prediction for channel coding with side information at the transmitter," in *Proc. IEEE Int. Symp. Information Theory*, Cambridge, MA, Aug 1998, p. 43.
- [12] C. D. Heegard and A. A. El Gamal, "On the capacity of computer memory with defects," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 731–739, Sept. 1983.
- [13] S. Gelfand and M. Pinsker, "Coding for channels with random parameters," *Probl. Control Inform. Theory*, vol. 9, pp. 19–31, 1980.
- [14] G. Caire and S. Shamai, "On the capacity of some channels with channel state information," *IEEE Trans. Inform. Theory*, vol. 45, pp. 2007–2019, Sept. 1999.
- [15] K. Mulkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple-antenna systems," *IEEE Trans. Inform. Theory*, pp. 2562–2579, Oct. 2003.
- [16] W. Santipach, Y. Sun, and M. L. Honig, "Benefits of limited feedback for wireless channels," in *Proc. 41th Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2003, pp. 1436–1445.
- [17] V. K. N. Lau, Y. Liu, and T.-A. Chen, "On the design of MIMO block-fading channels with feedback-link capacity constraint," *IEEE Trans. Commun.*, vol. 52, pp. 62–70, Jan. 2004.
- [18] S. P. Lloyd, "Least square quantization in PCM," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 129–137, Mar. 1982.
- [19] J. Max, "Quantizing for minimum distortion," *IEEE Trans. Inform. Theory*, vol. IT-6, pp. 7–12, Mar. 1960.
- [20] R. Gallager, *Information Theory and Reliable Communication*, 2nd ed. New York: Wiley, 1968.
- [21] "Physical layer aspects of ultra high speed downlink packet access," Tech. Rep. 3GPP TR25.848, 0.4.0 ed., 2000.
- [22] R. McEliece and W. Stark, "Channels with block interference," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 44–53, Jan. 1984.
- [23] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inform. Theory*, vol. 49, pp. 951–963, Apr. 2003.