

# Limited-Rate Channel State Feedback for Multicarrier Block Fading Channels

Manish Agarwal, Dongning Guo, and Michael L. Honig

Department of Electrical Engineering and Computer Science, Northwestern University

January 3, 2009

## Abstract

The capacity of a fading channel can be substantially increased by feeding back channel state information from the receiver to the transmitter. With limited-rate feedback what state information to feed back and how to encode it are important open questions. This paper studies power loading in a multicarrier system using no more than one bit of feedback per sub-channel. The sub-channels can be correlated and full channel state information is assumed at the receiver. First, a simple model with  $N$  parallel two-state (good/bad) memoryless sub-channels is considered, where the channel state feedback is used to select a fixed number of sub-channels to activate. The optimal feedback scheme is the solution to a vector quantization problem, and the associated performance for large  $N$  is characterized by a rate distortion function. As  $N$  increases, we show that the loss in forward rate from the asymptotic (rate-distortion) value decreases as  $(\log N)/N$  and  $\sqrt{(\log N)/N}$  with optimal variable- and fixed-rate feedback codes, respectively. We subsequently extend these results to parallel Rayleigh block fading sub-channels, where the feedback designates a set of sub-channels, which are activated with equal power. Rate-distortion feedback codes are proposed for designating subsets of (good) sub-channels with Signal-to-Noise Ratios (SNRs) that exceed a threshold. The associated performance is compared with that of a simpler lossless source coding scheme, which designates groups of good sub-channels, where both the group size and threshold are optimized. The rate-distortion codes can provide a significant increase in forward rate at low SNRs.

## Index Terms

Block fading, channel state feedback, limited feedback, multicarrier transmission, power control, rate distortion theory, Rayleigh fading, vector quantization.

This work was supported by the NSF under grant CCF-0644344, DARPA under grant W911NF-07-1-0028 and U.S. Army Research Office under grant W911NF-06-1-0339. The material in this paper was presented in part at the IEEE International Conference on Communications (ICC), Beijing, China, May, 2008.

## I. INTRODUCTION

Multicarrier transmission techniques, including orthogonal frequency-division multiplexing (OFDM), provide a convenient way to exploit frequency diversity in multipath fading channels. Given the total transmit power, a substantial increase in the channel capacity can be achieved if the power allocation across the sub-channels is adapted to channel variations [1]. For example, consider the sum capacity of  $N$  independent block Rayleigh fading sub-channels with given total power or Signal-to-Noise Ratio (SNR). If the power is equally spread over all  $N$  sub-channels, the capacity is upper bounded by the total SNR regardless of  $N$ , whereas if the power is allocated according to (optimal) water-filling, the capacity increases as  $O(\log N)$  as  $N$  increases [2], [3].

The state or quality of the sub-channels is typically measured at the receiver and sent to the transmitter through a feedback channel. We refer to this as *channel state feedback (CSF)*. Obviously, optimal power allocation requires a prohibitive (infinite) amount of CSF in case of continuous channel state. Even if the channel state can be discretized, the number of sub-channels may exceed the total number of feedback bits. Hence, what state information to feed back and how to encode the feedback are important questions.

This work studies the use of limited CSF for maximizing the achievable rate of multicarrier block fading channels. It is assumed that the sub-channel states are known or can be measured accurately at the receiver. The channel state is encoded using fewer than one bit per sub-channel and then sent to the transmitter through a noiseless feedback channel. The transmitter chooses a subset of sub-channels to activate based on the feedback.

The problem of encoding the feedback is essentially a vector quantization (VQ) problem, where the channel state is mapped to a given number of bits for later reconstruction. Unlike the usual quantization problem, the reconstruction here is to produce a power loading vector for the sub-channels, where the distortion metric is the gap between the rate achieved using the feedback and the capacity achieved with known channel state at the transmitter.

Multicarrier power allocation with limited-rate feedback has been previously considered in [2], [4]–[6]. In particular, [4] applies the Lloyd algorithm to produce a codebook of power loading vectors, which maximizes an objective such as achievable rate. Unfortunately, the size of the codebook in [4], and hence the search complexity, grows exponentially with the amount of feedback. Other heuristic schemes with one bit feedback per sub-channel have been proposed in [2], [3], [5].

This paper investigates the trade-off between the forward data rate and the amount of CSF for block-fading multicarrier channels assuming *no more than one bit* of feedback per sub-channel. Furthermore,

in contrast with the lossless feedback source coding schemes analyzed in [2], here we consider the more general class of lossy (rate-distortion) source codes.

We first consider a model with two fading states only. Each sub-channel randomly assumes either a *good* or *bad* state during a coherence block. For the case of independent two-state sub-channels studied in Section II, the role of the feedback is to direct the transmitter to select as many good sub-channels as possible to activate subject to the power constraint. The fundamental trade-off between the feedback rate and the sum capacity can be characterized using rate distortion theory in the limit of infinite number of sub-channels. For given finite number of sub-channels, we also quantify the gap between rates achievable by random coding and the rate distortion bound. Specifically, with variable-rate feedback codes the gap decreases as  $(\log N)/N$ , whereas with fixed-rate codes the gap decreases as  $\sqrt{(\log N)/N}$ .

We also compare the rate-distortion approach with a simple lossless source coding scheme, which reports as many good sub-channels as the feedback rate allows. Numerical plots show that good codes in the rate distortion sense typically achieve much higher forward rate. The result is then extended to the case of correlated two-state sub-channels in Section III, where the sub-channel states are assumed to form a Markov chain. Upper and lower bounds on the forward rate are derived as a function of the feedback rate.

With the insights gained from the two-state channel model, we then study the problem of limited CSF for Rayleigh fading sub-channels. The fading coefficient, or state of each sub-channel is a Circularly Symmetric Complex Gaussian (CSCG) random variable during each coherence block. The case of independent sub-channels is studied in Section IV whereas the case of correlated sub-channels is discussed in Section V. The state of each sub-channel is first reduced to a binary variable by comparing its gain with a threshold. Similar feedback codes as considered for the two-state channels is used to instruct the transmitter which sub-channels to activate, assuming the power is distributed evenly over the activated sub-channels. The threshold is selected to maximize the forward rate given a fixed feedback rate. It turns out that the trade-off admits a similar characterization as that for two-state sub-channels. Although reduction of Rayleigh states to binary states induces loss, the scheme with optimized threshold and a moderate amount of feedback performs close to optimal water-filling with channel coefficients known at the transmitter. In particular, given a total power constraint, the scheme can achieve a forward rate, which has the same order of increase with the number of sub-channels as that of water-filling [2].

Two heuristic lossless source schemes for the reduced (two-state) version of the Rayleigh channel are also considered for comparison in Section IV. In particular, in one of the schemes, taken from [7], the sub-channels are divided evenly into groups and the feedback indicates the set of groups in which all

sub-channel gains exceed the threshold. A binary state vector, indicating which groups to activate, is then compressed using lossless source coding and fed back to the transmitter. The group size and threshold can be adjusted to maximize the forward achievable rate, subject to the feedback rate constraint. Such grouping, or clustering, of sub-channels to reduce feedback overhead has also been studied in [8] in a multiuser setting. Clustering sub-channels to reduce the training overhead and peak-to-average power ratio was previously studied in [9]. We characterize the growth in achievable rate with the number of sub-channels (for large  $N$ ) as a function of the amount of feedback (which can also scale with  $N$ ). Numerical examples show that the analytical results are quite accurate for finite-size systems of interest. In general, these heuristic schemes achieve a smaller forward rate than for the rate-distortion schemes, given a fixed feedback rate.

## II. INDEPENDENT TWO-STATE SUB-CHANNELS

Consider a bank of  $N$  independent and statistically identical block fading sub-channels. During each coherence block, each sub-channel randomly takes one of two states, namely “good” and “bad,” which is known to the receiver. The input is constrained such that up to a fraction  $p$  of the sub-channels can be activated by the transmitter. Suppose on average the amount of CSF is limited to  $R_f$  bits per sub-channel per coherence block. The problem is to design a feedback scheme to maximize the forward data rate, i.e., to activate as many good sub-channels as possible.

### A. The Fundamental Trade-off via Rate Distortion Theory

Let the state of sub-channel  $i$  be denoted by a Bernoulli random variable<sup>1</sup>  $S_i$ , with the probability of being a good state denoted as  $P\{S_i = 1\} = q$ . Further, let the power loading variable  $\hat{S}_i = 1$  if the  $i^{\text{th}}$  sub-channel is chosen to be activated and  $\hat{S}_i = 0$  otherwise. Constrained by the feedback and transmission power, a feedback scheme specifies a mapping from the set of binary channel state vectors, whose Hamming weight is no greater than  $pN$ .

It is easy to see that the feedback scheme is no different than vector quantization, where the channel state vector  $\mathbf{S} = [S_1, \dots, S_N]$  is mapped to  $NR_f$  bits for recovery at the transmitter. Constrained by the feedback rate, the reconstruction may be prone to errors, and the quantization scheme should be designed to achieve as few errors (or, as small a *distortion*) in reconstruction as possible.

<sup>1</sup>The following convention will be adopted throughout the paper: A boldface letter represents a vector. An uppercase letter represents a random vector or variable (e.g.,  $\mathbf{S}$ ,  $S_i$ ), and the corresponding lower case letter represents a specific realization (e.g.,  $\mathbf{s}$ ,  $s_i$ ). In addition,  $\log(\cdot)$  denotes natural logarithm.

The fundamental trade-off of the forward and feedback rates as  $N \rightarrow \infty$  can be addressed using rate distortion theory. The source is a sequence of independent and identically distributed (i.i.d.) Bernoulli( $q$ ) random variables,  $S_1, S_2, \dots$ . The distortion measure can be described as  $d_N(\mathbf{s}, \hat{\mathbf{s}}) = \frac{1}{N} \sum_{i=1}^N d(s_i, \hat{s}_i)$  with

$$d(s, \hat{s}) = 1_{\{s > \hat{s}\}} = \begin{cases} 1, & \text{if } s = 1 \text{ and } \hat{s} = 0, \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

The metric accounts for *missed opportunities*, i.e., good sub-channels which are not activated, but does not penalize activation of bad sub-channels, which we refer to as *misfires*. Further, the power loading vector has to satisfy a normalized weight constraint:

$$w(\hat{\mathbf{s}}) = \frac{1}{N} \sum_{i=1}^N \hat{s}_i \leq p. \quad (2)$$

This additional challenge of incorporating the weight constraint on the reconstruction distinguishes the problem from the classical rate distortion problem concerning i.i.d. source and single-letter distortion measure. Though not obvious, the rate distortion problem admits the following simple single-letter characterization.

*Theorem 1:* For an i.i.d. Bernoulli( $q$ ) source, given the weight constraint on every binary reconstruction,  $w(\hat{\mathbf{s}}) \leq p$ , and the distortion measure,  $d(s, \hat{s}) = 1_{\{s > \hat{s}\}}$ , the rate distortion function is

$$R(D) = \min_{P_{\hat{s}|s}: \substack{\mathbb{E} d(S, \hat{S}) \leq D \\ \mathbb{P}\{\hat{S}=1\} \leq p}} I(S; \hat{S}) \quad (3)$$

where  $S \sim \text{Bernoulli}(q)$ .

*Proof:* The achievability part of the theorem is based on Shannon's random coding technique (see e.g., [10]). Fix  $P_{\hat{S}|S}$  and some  $0 < \delta_1 < p$ , which satisfy  $\mathbb{E} d(S, \hat{S}) \leq D$  and  $\mathbb{P}\{\hat{S} = 1\} \leq p - \delta_1$ . The code book of  $2^{NR_f}$  codewords can be produced randomly with the marginal distribution  $P_{\hat{S}}$ . Further, an exponentially small fraction of codewords which violate the weight constraint (2) are purged. It can be shown that for sufficiently large code length  $N$ , the random codebook achieves the distortion  $D$  as long as the rate  $R > I(S; \hat{S}_i) + \delta_2$ . The achievability part is thus proved because  $\delta_1$  and  $\delta_2$  can be chosen to be arbitrarily small.

Showing the converse requires incorporating the weight constraint (2) into the standard technique of [10]. Let  $\hat{\mathbf{S}}$  represent the reconstruction of the random source vector  $\mathbf{S}$ . Consider any code of length  $N$  with rate  $R$  which satisfies the distortion and average weight constraints  $\mathbb{E}[w(\hat{\mathbf{S}})] \leq p$ , which is a weaker

than required in the theorem.<sup>2</sup> Then, due to the data processing theorem and the independence of  $S_i$ ,

$$NR \geq I(\mathbf{S}; \hat{\mathbf{S}}) \quad (4)$$

$$\geq \sum_{i=1}^N I(S_i; \hat{S}_i) \quad (5)$$

$$\geq \min_{P_{\hat{\mathbf{S}}|\mathbf{S}}: \begin{matrix} \mathbb{E}\{\frac{1}{N} \sum_{i=1}^N \hat{S}_i\} \leq p \\ \mathbb{E}\{\frac{1}{N} \sum_{i=1}^N d(S_i, \hat{S}_i)\} \leq D \end{matrix}} \sum_{i=1}^N I(S_i; \hat{S}_i). \quad (6)$$

The key task here is to break down the constraints on the distribution of the vector  $\hat{\mathbf{S}}$  in (6) into constraints on the individual random variables. Note that  $P_{\hat{\mathbf{S}}}$  is linear in  $P_{\hat{S}_i|\mathbf{S}}$  because the source distribution  $P_{\mathbf{S}}$  is fixed. An important fact is that  $I(S_i; \hat{S}_i)$  is convex in the distribution  $P_{\hat{S}_i|S_i}$ . Because of the symmetry in the indexes  $i$ , any optimal distribution  $P_{\hat{\mathbf{S}}|\mathbf{S}}$  that achieves the minimum of (6) must be symmetric over all indexes  $i$ . Otherwise replacing all of them by their average yields smaller mutual information. Therefore, due to the symmetry and the additive nature of the constraints, (6) implies that the rate  $R$  is lower bounded by  $R(D)$  given in (3). ■

The minimization over the conditional distribution  $P_{\hat{S}_i|\mathbf{S}}$  in (3) is equivalently over the crossover probabilities:

$$\epsilon_0 = P_{\hat{S}_i|\mathbf{S}}(0|1) \quad \text{and} \quad \epsilon_1 = P_{\hat{S}_i|\mathbf{S}}(1|0), \quad (7)$$

where  $\epsilon_0$  represents the probability of missing a good sub-channel. The mutual information  $I(\mathbf{S}; \hat{\mathbf{S}})$  can be expressed as the following function of  $(\epsilon_0, \epsilon_1)$ :

$$i(\epsilon_0, \epsilon_1) \triangleq H_2(p) - qH_2(\epsilon_0) - (1 - q)H_2(\epsilon_1), \quad (8)$$

where  $H_2(\cdot)$  stands for the binary entropy function. Unless stated otherwise, the units of all information metrics are bits. Note that the weight constraint (2) should be tight at the minimum because there is no penalty on misfires. Thus the optimal crossover probabilities satisfy  $q(1 - \epsilon_0) + (1 - q)\epsilon_1 = p$ .

Let the capacity of a good sub-channel be  $C_1$  and the capacity of a bad sub-channel be  $C_0 < C_1$ . The average number of active good sub-channels is  $Nq(1 - \epsilon_0)$ . The trade-off between the capacity and the feedback rate is characterized by as follows.

*Proposition 1:* Given  $p, q$ , and the feedback rate  $R_f$  bits per sub-channel per coherence block, the

<sup>2</sup>Theorem 1 continues to hold even if the instantaneous input constraint (2) is replaced by an average constraint, namely  $\mathbb{E}[w(\hat{\mathbf{S}})] \leq p$ .

maximum achievable forward data rate per sub-channel is

$$C = q(1 - \epsilon_0^*)(C_1 - C_0) + pC_0, \quad (9)$$

where the optimal proportion of missed good sub-channels  $\epsilon_0^*$  is the solution to the following optimization problem:

$$\text{minimize: } \epsilon_0 \quad (10a)$$

$$\text{subject to: } H_2(p) - qH_2(\epsilon_0) - (1 - q)H_2(\epsilon_1) \leq R_f, \quad (10b)$$

$$q(1 - \epsilon_0) + (1 - q)\epsilon_1 = p, \quad (10c)$$

$$0 \leq \epsilon_0, \epsilon_1 \leq 1. \quad (10d)$$

The optimization problem (10) can be easily solved numerically. Clearly, the maximum forward data rate increases as the feedback rate increases, but the return vanishes beyond a certain point. The minimum feedback rate necessary for achieving the capacity can be determined by tentatively removing the feedback constraint (10b). If  $p \geq q$ , one can activate all good sub-channels so that  $\epsilon_0 = 0$  with  $\epsilon_1 = \frac{p-q}{1-q}$ , whereas if  $p < q$ , then  $\epsilon_0$  can be as small as  $1 - \frac{p}{q}$  by choosing  $\epsilon_1 = 0$ . Substituting these values into (8), the forward rate saturates at the maximum feedback rate,

$$\bar{R} = \begin{cases} H_2(p) - qH_2\left(\frac{p}{q}\right), & \text{if } p \leq q, \\ H_2(1 - p) - (1 - q)H_2\left(\frac{1-p}{1-q}\right), & \text{if } p > q. \end{cases} \quad (11)$$

For any  $R_f \leq \bar{R}$ , the constraint (10b) is tight and  $\epsilon_0^*$  can be calculated by solving the simultaneous equations (10b) and (10c), which can be easily reduced to a fixed-point equation.

### B. Performance Bounds for Finite Number of Sub-channels

Proposition 1 characterizes the asymptotic trade-off as the number of sub-channels  $N$  goes to infinity. For a practical situation with finite  $N$ , the result needs refinement. Note that the solution to Proposition 1 provides an upper bound on the forward data rate for finite  $N$ , because the converse shown in the proof of Theorem 1 holds for all  $N$ . In the following, we consider random feedback codes and derive a lower bound for the achievable forward data rate with given feedback constraint.

1) *Fixed-Length Constant-Composition Feedback Code*: Note that the solution to Proposition 1 upper bounds the forward data rate with the average input power constraint  $E[w(\hat{\mathbf{S}})] = p$  and *average* feedback rate of  $Ni(\epsilon_0, \epsilon_1)$  bits per coherence block. Here we impose two additional constraints without loss of

generality: 1) The binary reconstruction vectors have constant composition, that is,  $w(\hat{s}) = p$  for all the vectors  $\hat{s}$  in the feedback codebook; and 2) The feedback is at most  $Ni(\epsilon_0, \epsilon_1)$  bits every coherence block. The second restriction implies that there are at most  $2^{i(\epsilon_0, \epsilon_1)N}$  codewords.

The following proposition gives a lower bound on the achievable forward rate given these additional constraints.

*Proposition 2:* Let the number of feedback bits per coherence block be  $Ni(\epsilon_0, \epsilon_1)$ , where  $\epsilon_0, \epsilon_1 > 0, \neq 0.5$  are the solution to (10). Then  $\exists N_o < \infty$  such that for  $N \geq N_o$ , the ergodic capacity achieved with the fixed-length constant-composition feedback code is lower bounded as

$$C_{fixed} \geq \left(1 - 2\sqrt{\frac{\log(Nq)}{Nq}}\right) C \quad (12)$$

where  $C$  is given by (9).

The proof is given in Appendix I. The proposition implies that, for large enough  $N$ , the difference between the upper bound (9) and achievable forward rate per sub-channel approaches zero at the rate  $O(\sqrt{(\log N)/N})$ . This further implies that the sum rate across the  $N$  sub-channels incurs a loss, which increases as  $O(\sqrt{N \log N})$  compared to  $NC$ . The proof basically follows the random coding technique of Goblick [11] for analyzing the convergence rate of the rate distortion function for general sources and fixed-length block codes. The contribution in this work is to incorporate the additional constant-composition constraint and to simplify the analysis by exploiting the binary structure of the source and the reconstruction.<sup>3</sup>

Although the result in Proposition 2 is stated for large  $N$ , more refined lower bound is derived in Appendix I which holds for any finite  $N$ .

2) *Variable-Length Feedback Codes:* We note that fixed-length codes can *cover* a subset of most probable channel state vectors, but are unable to adapt to deviations from typical channel conditions. In the following, we analyze the performance of variable-length feedback codes. A variable amount of feedback is allowed during each coherence block as long as the average number of feedback bits is  $Ni(\epsilon_0, \epsilon_1)$ . The instantaneous power constraint is replaced by an average power constraint.

*Proposition 3:* Let the average number of feedback bits per coherence block be  $Ni(\epsilon_0, \epsilon_1)$ , where  $\epsilon_0, \epsilon_1 > 0$  are the solution to Proposition 1 and  $\epsilon_0, \epsilon_1 \neq \frac{1}{2}$ . Then  $\exists N_1 < \infty$  such that for every  $N \geq N_1$ , the ergodic forward rate achieved with a variable-length feedback code, under an average input power

<sup>3</sup>We avoid the use of complicated *partition functions* in [11] by using a Chernoff bound to evaluate the tail probability distributions.



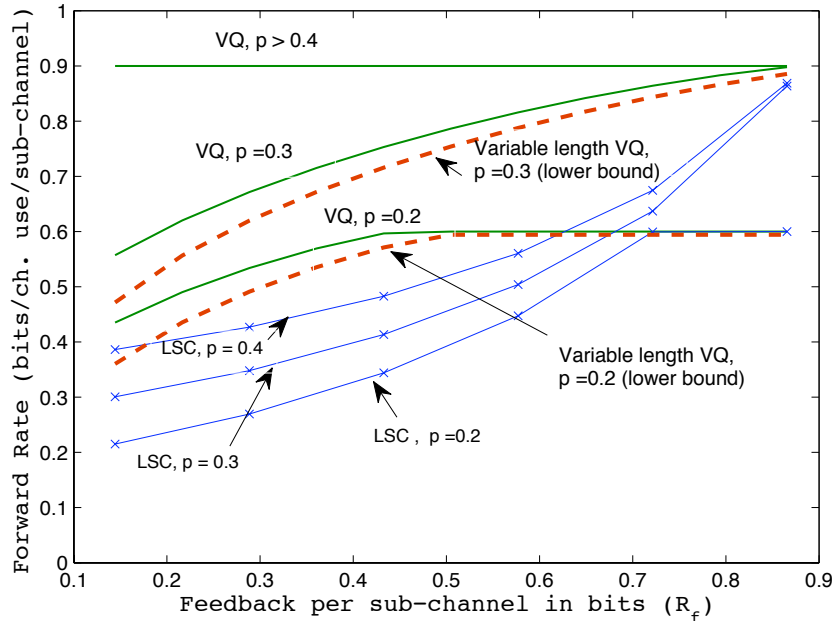


Fig. 1. Forward rate versus feedback rate for different input constraints corresponding to VQ and sub-optimal feedback schemes using lossless source coding with channel state reduction (curves labeled with “LSC”). The sub-channels are assumed to be independent. Other parameters are  $q = 0.3$ ,  $C_1 = 3$  and  $C_0 = 0$ . Also shown is the lower bound on forward rate corresponding to  $N = 500$  and a variable-length feedback code (dotted lines).

constraint  $E[w(\hat{\mathbf{S}})] = p$ , is lower bounded as

$$C_{var} \geq \left(1 - \frac{6}{q[H_2'(\epsilon_0) + H_2'(\epsilon_1)]} \frac{\log_2 N}{N}\right) C \quad (13)$$

where  $C$  is given in (9).

The proof is provided in Appendix II. We follow the technique given by Pinkston [12], albeit with slight modifications. The main difference lies in incorporating the average input power constraint and avoiding the use of partition functions by exploiting the binary structure of the state and power loading vectors.

The proposition says that in this scenario the forward rate converges to the upper bound (9) as  $O((\log N)/N)$ . This is a substantial improvement over the  $O(\sqrt{(\log N)/N})$  convergence rate achieved by the fixed-length constant composition feedback codes (cf. (12)). Although the result in Proposition 3 is stated for large  $N$ , a more refined lower bound is derived in Appendix II, which holds for any finite  $N$ . Fig. 1 plots a few instances of this lower bound against the upper bound (9) as the feedback rate varies. Clearly, the lower bound is fairly close to the optimal forward rate (9) and becomes tighter as the

feedback rate increases.

### C. Practical CSF Codes

Until now, we have shown that for moderate to large  $N$ , the rate distortion trade-off can be approached using random codes. Such a code, however, is not practical. As aforementioned, the Lloyd algorithm can be used to design a near-optimal vector quantizer for small number of sub-channels (see [4], [13]–[15] and references therein). Such a task becomes infeasible with tens or hundreds of sub-channels, as is the case in many applications.

One practical solution in the case of a large number of sub-channels is to use a graphical code similar to a low-density parity-check (LDPC) code. Encoding and decoding of the source (channel state vector) are respectively analogous to iterative decoding and encoding of a graphical error-control code. The complexity of such a code is in general linear in the number of sub-channels. For a discussion of graphical codes for source coding, the reader is referred to [16]–[18]. It is more challenging to design and implement variable-length codes.

### D. A Sub-optimal Scheme: Lossless Source Coding with Channel State Reduction

For comparison, we also consider a feedback scheme using simple channel state reduction and lossless source coding in lieu of vector quantization (henceforth referred to as the “LSC” scheme for convenience). If the feedback rate is greater than the entropy rate of the channel state vector, i.e.,  $R_f > H_2(q)$ , then any lossless codes such as the Huffman code basically suffice. If the feedback rate is less than the entropy rate, we consider a simple scheme which reports a fraction  $f$  of good sub-channels, where  $f$  is chosen such that the entropy rate  $H_2(fq)$  is basically  $R_f$ . On average the transmitter is informed of  $fqN$  good sub-channels. The forward rate achieved with this option is

$$C_f = \begin{cases} p C_1, & \text{if } p \leq fq \\ fq C_1 + (p - fq) \frac{q(1-f)C_1 + (1-q)C_0}{q(1-f) + (1-q)}, & \text{otherwise.} \end{cases} \quad (14)$$

The expression follows directly from the observation that if fewer than fraction  $p$  of the sub-channels are reported as good (that is,  $fq < p$ ), the remaining  $p - fq$  fraction of the sub-channels are chosen at random so that the probability of transmitting on a good sub-channel is given by  $(q - fq)/(1 - fq)$ .

Note that it might be more efficient for the receiver to inform the transmitter to avoid a subset of bad sub-channels than to report a subset of good sub-channels, depending on the parameters. Suppose a

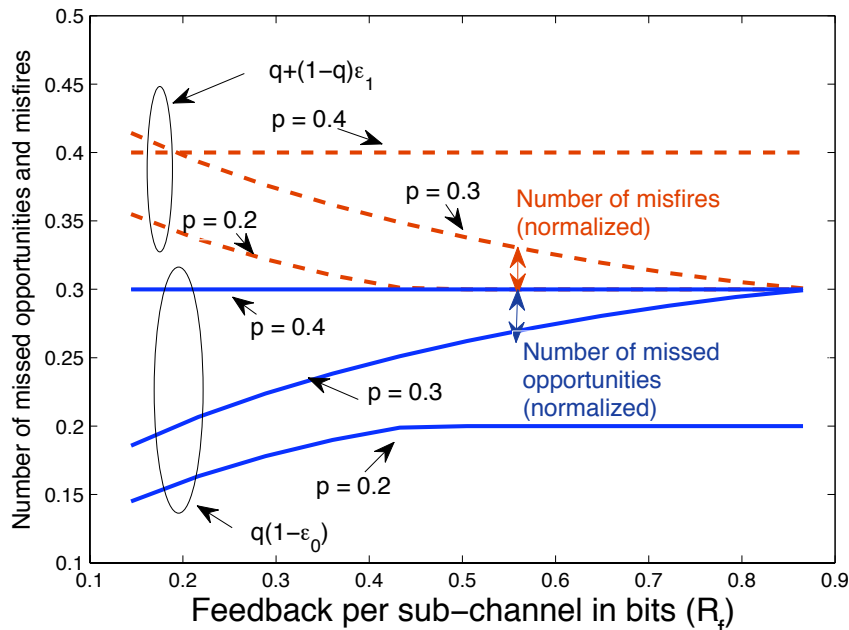


Fig. 2. The number of misfires and missed opportunities (normalized by the total number of sub-channels  $N$ ) versus the feedback rate for different input constraints. The sub-channels are assumed to be independent. Other parameters are  $q = 0.3$ ,  $C_1 = 3$  and  $C_0 = 0$ .

fraction  $\bar{f}$  of the bad sub-channels are reported to the transmitter where  $H_2(\bar{f}(1-q)) = R_f$ . The forward rate achieved with this option is given by

$$\bar{C}_f = \begin{cases} p \frac{qC_1 + (1-\bar{f})(1-q)C_0}{q + (1-\bar{f})(1-q)}, & \text{if } p < q + (1-\bar{f})(1-q) \\ qC_1 + (p-q)C_0, & \text{otherwise.} \end{cases} \quad (15)$$

The maximum forward data rate achievable by the LSC scheme is therefore  $\max\{C_f, \bar{C}_f\}$ .

### E. Numerical Results

We study the asymptotic performance of the optimal VQ scheme (given by (9)) and the sub-optimal LSC scheme (given by (14) and (15)). Fig. 1 plots the forward rate per sub-channel versus the feedback rate  $R_f$  for different values of the input power constraint  $p$ . The LSC scheme is clearly inferior compared to the asymptotic VQ scheme with infinite as well as the variable-length VQ scheme at  $N = 500$  sub-channels. At small values of feedback, the VQ scheme gives substantial gains (up to 100%). The asymptotic result is quite representative of the performance with a relatively large number of sub-channels ( $N = 500$ ). As

expected, the forward rate increases with the feedback amount and saturates at  $R_f = H_2(q)$ , at which point all good sub-channels can be reported at no loss.

The gain achieved with VQ can be better understood by studying the corresponding numbers of missed opportunities and misfires shown in Fig. 2. Intuitively, the larger the values of  $\epsilon_0$  and  $\epsilon_1$ , the more “distortion” or “errors” we allow in the rate-distortion feedback code and hence it will require smaller amount of feedback. This is clearly reflected in (8). Consider the case of  $p = 0.4$  ( $> q$ ) in Fig. 2,  $\epsilon_0 = 0$  and the fraction of misfires  $(1 - q)\epsilon_1 = 0.1$ , which implies that we allow enough misfires so that the required feedback rate is kept small. In contrast, with  $p = 0.4$  we never report a bad sub-channel as good for the LSC scheme, and thereby incur extra feedback overhead. The reverse holds for  $p = 0.2$  ( $< q$ ), where the optimal scheme allows enough missed opportunities ( $\epsilon_0$  is large and  $\epsilon_1$  is small) so that the required feedback rate is again small. Since in this case  $\epsilon_1 \approx 0$ , bad channels are not reported as good.

### III. CORRELATED TWO-STATE SUB-CHANNELS

In multicarrier systems, the states of the sub-channels are often correlated. Consider the same system as in Section II except that the binary (good/bad) channel states of the  $N$  sub-channels,  $S_1, S_2, \dots, S_N$  form a stationary Markov chain. The optimal feedback scheme is nonetheless a vector quantization problem, with its asymptotic performance characterized by the following rate distortion result.

*Theorem 2:* Given a stationary binary Markov source  $\{S_i\}$ , a weight constraint on every binary reconstruction  $w(\hat{s}) \leq p$ , and the single-letter distortion measure  $d(s, \hat{s}) = 1_{s > \hat{s}}$ , the rate distortion function is given by

$$R(D) = \limsup_{N \rightarrow \infty} \min_{P_{\hat{s}|s}: \sum \mathbb{P}\{\hat{S}_i=1\} \leq pN, \sum \mathbb{E}[d(S_i, \hat{S}_i)] \leq DN} \frac{1}{N} I(\mathbf{S}; \hat{\mathbf{S}}) \quad (16)$$

It is straightforward to prove Theorem 2 using the same techniques as developed in [19], with the additional weight constraint for the reconstruction. Hence the proof is omitted. Random codes achieve the rate distortion function. However, in practice, graphical codes can be designed to approach the optimal trade-off. In addition, Theorem 2 continues to hold if the instantaneous input constraint  $w(\hat{s}) \leq p$  is replaced by an average input constraint  $\mathbb{E}[w(\hat{\mathbf{S}})] \leq p$ .

Note that (16) involves minimization over the conditional distribution of the entire power loading vector, and hence is not a single-letter characterization of the rate distortion function. Calculating the rate distortion function for correlated sources is a hard problem in general. The solution is known only in a few special cases pertaining to source alphabets, correlation models and distortion measures [19]. Even for a symmetric binary Markov chain and Hamming distortion, the rate distortion function is

exactly known only for very small distortion values [20]. In the CSF problem, the reconstruction  $\hat{\mathbf{S}}$  is a binary hidden Markov process. There is no known close-form expression for the entropy rate of such processes, although there exist approximations and numerical results in some cases (see, e.g., [21], [22] and references therein).

Since the exact solution to the optimization problem (16) is difficult, we will next find upper and lower bounds on the rate for a given distortion. Due to the stationarity of the source, the optimal conditional probability  $P_{\hat{\mathbf{S}}|\mathbf{S}}$  is also stationary. Consider any stationary process  $\{(S_i, \hat{S}_i) | i = 0, \pm 1, \dots\}$  which satisfies the constraints in (16), i.e.,  $\mathbb{P}\{\hat{S}_i = 1\} \leq p$  and  $\mathbb{E}[d(S_i, \hat{S}_i)] \leq D$ . Then

$$\lim_{N \rightarrow \infty} \frac{1}{N} I(\mathbf{S}; \hat{\mathbf{S}}) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^{N-1} I(S_i; \hat{\mathbf{S}} | \mathbf{S}_0^{i-1}) \quad (17)$$

$$= H(S_1 | S_0) - \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^{N-1} H(S_i | \hat{\mathbf{S}}, \mathbf{S}_0^{i-1}) \quad (18)$$

$$\geq H(S_1 | S_0) - H(S_1 | \hat{S}_1, S_0) \quad (19)$$

$$= I(S_1; \hat{S}_1 | S_0) \quad (20)$$

where (19) is because of stationarity and because conditioning decreases the entropy. The rate distortion function can thus be lower bounded as

$$R(D) \geq \min_{P_{\hat{S}_1|S_1, S_0}} I(S_1; \hat{S}_1 | S_0) \quad (21)$$

with  $P_{\hat{S}_1|S_1, S_0}$  satisfying the constraints in (16). The bounding mutual information depends only on the following four probabilities for the given source:  $q_{s_0 s_1} = P_{\hat{S}_1, S_0, S_1}(0, s_0, s_1)$  with  $s_0, s_1 = 0$  or  $1$ . Denote the lower bound by  $i_l(q_{00}, q_{01}, q_{10}, q_{11}) = I(S_1; \hat{S}_1 | S_0)$ . This bound can be expressed as a function of the crossover probabilities denoted by  $\delta_{01} = \mathbb{P}\{S_i = 1 | S_{i\pm 1} = 0\}$  and  $\delta_{10} = \mathbb{P}\{S_i = 0 | S_{i\pm 1} = 1\}$  for all  $i$ . (The probability of a sub-channel being good is then  $q = \delta_{01} / (\delta_{01} + \delta_{10})$ ) An explicit expression for the lower bound is derived in Appendix III.

In terms of these joint probabilities, the fraction of sub-channels that are good, and are correctly reported as good is given by  $(q - q_{01} - q_{11})$  and the total fraction of sub-channels reported as good is given by  $1 - (q_{00} + q_{01} + q_{10} + q_{11})$ . Consequently, an upper bound on the forward achievable rate can

be obtained as the solution to the following optimization problem:

$$\text{maximize: } (C_1 - C_0)(q - q_{01} - q_{11}) + pC_0 \quad (22a)$$

$$\text{subject to: } 1 - (q_{00} + q_{01} + q_{10} + q_{11}) \leq p \quad (22b)$$

$$i_l(q_{00}, q_{01}, q_{10}, q_{11}) \leq R_f \quad (22c)$$

$$0 \leq q_{00}, q_{01}, q_{10}, q_{11} \leq 1 \quad (22d)$$

which can be easily solved numerically.

In order to find an upper bound on the rate distortion function, we restrict the minimization over  $P_{\hat{\mathcal{S}}|\mathcal{S}}$  in (16) to be a minimization over a finite-dimensional distribution. For example, suppose that conditioned on  $S_i$  and  $S_{i+1}$ , the random variable  $\hat{S}_i$  is independent of all the remaining random variables in  $(\mathcal{S}, \hat{\mathcal{S}})$ . By stationarity and the Markovian property, the joint distribution of  $(\mathcal{S}, \hat{\mathcal{S}})$  is determined by the conditional distribution  $P_{\hat{S}_0|S_0, S_1}$ . Then it can be shown that, conditioned on  $S_{i-1}, S_{i+1}, \hat{S}_{i-1}$  and  $\hat{S}_i$ , the variable  $S_i$  is also independent of all the remaining random variables in  $\mathcal{S}$  and  $\hat{\mathcal{S}}$ . Consequently,

$$H(S_i|\hat{\mathcal{S}}, \mathbf{S}_0^{i-1}) \geq H(S_i|\hat{\mathcal{S}}, \mathbf{S}_0^{i-1}, \mathbf{S}_{i+1}^N) \quad (23)$$

$$= H(S_i|S_{i-1}, S_{i+1}, \hat{S}_{i-1}, \hat{S}_i). \quad (24)$$

Substituting in (18), an upper bound for the rate distortion function is obtained as the solution to the following optimization problem:

$$R(D) \leq \min_{P_{\hat{S}_0|S_0, S_1}} I(S_1; S_2, \hat{S}_0, \hat{S}_1|S_0) \quad (25)$$

where  $P_{\hat{S}_0|S_0, S_1}$  is also subject to the constraints in (16). A simpler but looser bound is obtained if we assume that  $\hat{S}_i$  is independent of everything else conditioned on  $S_i$ , for which the mutual information in (25) becomes  $I(S_1; S_2, \hat{S}_1|S_0)$ , and the minimization is over  $P_{\hat{S}_0|S_0}$ . Again, let the crossover probabilities be given by (7). The upper bound is a function of these two crossover probabilities and is denoted by  $i_u(\epsilon_0, \epsilon_1) = I(S_1; S_2, \hat{S}_1|S_0)$ . An explicit expression is derived in Appendix III. Consequently, a lower bound on the forward achievable rate can be obtained by solving an optimization problem similar to that in Proposition 1 with the constraint (10b) replaced by  $i_u(\epsilon_0, \epsilon_1) \leq R_f$ .

It is easily seen that if no reconstruction errors are allowed, then both the upper and lower bounds reduce to the entropy rate of the channel state process. In other word, if  $H(\hat{\mathcal{S}}|\mathcal{S}) = 0$ , or equivalently,  $\epsilon_0 = \epsilon_1 = 0$ ,  $q_{10} = q_{00} = 1$  and  $q_{01} = q_{11} = 0$ , then  $i_l(1, 0, 1, 0) = i_u(0, 0) = H(S_1|S_0)$ . Next we

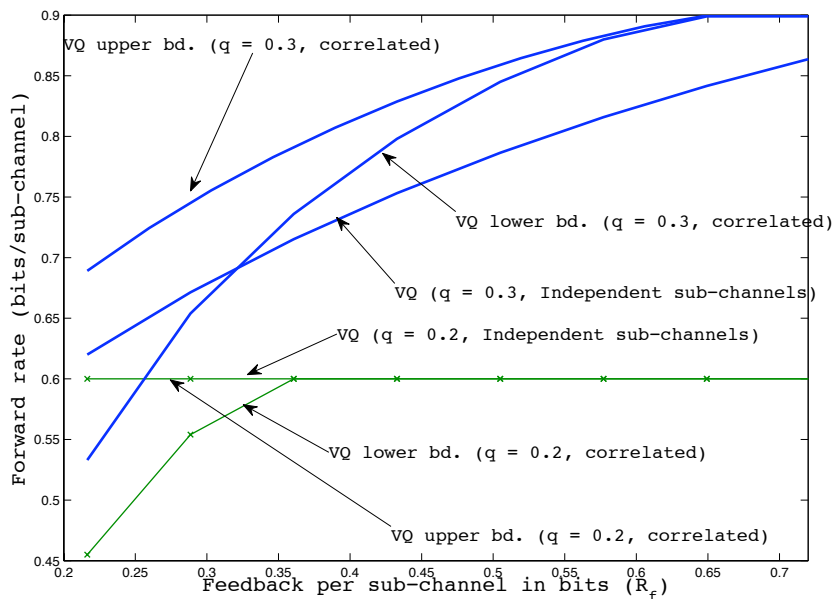


Fig. 3. Upper and lower bounds on the forward rate with VQ versus the feedback rate for correlated sub-channels. Parameters are  $\delta_{10} = 0.3, p = 0.3, C_1 = 3$  and  $C_0 = 0$ .

provide an example in which the upper bound is not tight. Choosing  $p = 1$  implies that the power loading vector  $\hat{S}$  can be chosen all ones and that achieves the capacity with zero feedback rate. However, equivalently choosing  $\epsilon_0 = 0$  and  $\epsilon_1 = 1$  gives the upper bound on required feedback rate  $i_u(0, 1) > 0$ .

Fig. 3 plots the upper and lower bounds on achievable forward rate per sub-channel versus  $R_f$  for different values of  $q$  with  $\delta_{10} = 0.3$  and  $p = 0.3$ . Consider the plot for  $q = 0.3$ . There is a substantial gap between the two bounds for small feedback rates. However, as the feedback rate increases, the gap closes and the bounds provide an accurate measure of the performance of the VQ scheme with correlated sub-channels. Also shown is the performance of the VQ scheme with independent sub-channels. Clearly, correlation improves the forward rate by decreasing the feedback requirement.

Later in Section V, the preceding methodology will be utilized to derive bounds on the performance of VQ schemes for correlated Rayleigh fading sub-channels.

#### IV. INDEPENDENT RAYLEIGH FADING SUB-CHANNELS

The design of a limited-rate CSF strategy for Rayleigh fading sub-channels is again a VQ problem. Unfortunately, the exact distortion measure, which corresponds to the capacity-maximizing power loading

vectors is difficult to work with [4], [13]–[15]. In order to simplify the problem, we focus on threshold-based schemes, which converts the sequence of Rayleigh fading sub-channels to a sequence of “good” (sub-channel gain above the threshold) and “bad” (sub-channel gain below the threshold) sub-channels. This enables the use of the limited feedback schemes developed for two-state sub-channels. It will be shown that, as the number of sub-channels  $N \rightarrow \infty$ , the rate achieved with such a scheme grows at the same rate as that of water-filling with full channel state information at the transmitter.

The limited feedback problem here differs from the case of two-state sub-channels studied in Sections II and III in two key aspects. First, the threshold which determines the fraction of sub-channels that are considered good needs to be optimized. Second, given the total power, the fraction of sub-channels to activate also influence the amount of powers in each active sub-channel.

### A. System Model

Consider a multicarrier channel with  $N$  independent and statistically identical, where the channel output for the  $i$ -th sub-channel is written as

$$Y_i = H_i X_i + Z_i \quad (26)$$

where  $H_i$  and  $Z_i$  are zero-mean circularly symmetric complex Gaussian (CSCG) random variables. Without loss of generality, we assume that the channel and noise variance is one, that is,  $E[|H_i|^2] = E[|Z_i|^2] = 1$ . Also, the noise is assumed to be independent across the sub-channels. The  $N \times 1$  input vector  $\mathbf{X} = [X_1, X_2, \dots, X_N]^\dagger$  satisfies the average total signal-to-noise ratio (SNR) constraint  $E[\mathbf{X}^\dagger \mathbf{X}] \leq P$ . The channel vector  $\mathbf{H} = [H_1, H_2, \dots, H_N]^\dagger$  is assumed to be known perfectly at the receiver. We assume a block fading model so that  $\mathbf{H}$  remains constant for  $T$  channel uses and then changes to an independent value. The time dependence is suppressed to simplify notation.

### B. Optimal Threshold Based VQ

The gain for the  $i$ -th sub-channel,  $|H_i|^2$ , is exponentially distributed with its mean equal to one. Given a threshold  $t \geq 0$ , define the  $N \times 1$  binary state vector  $\mathbf{S}$  so that the  $i$ -th entry  $S_i = 1$  if  $|H_i|^2 \geq t$ , and  $S_i = 0$  otherwise. The probability of a sub-channel being “good” is denoted as  $q = \text{P}\{S_i = 1\} = \text{P}\{|H_i|^2 > t\} = e^{-t}$ .

Suppose that, on average, the transmitter transmits over or, *activates* a fraction  $p$  of the sub-channels. The power is distributed uniformly over the active sub-channels so that each transmission occurs with SNR equal to  $P/(Np)$ . Therefore, the expected capacity of a good and bad sub-channel, respectively, is



given by

$$C_1 = \frac{1}{q} \int_t^\infty e^{-\tau} \log \left( 1 + \frac{P\tau}{Np} \right) d\tau \quad (27)$$

and

$$C_0 = \frac{1}{1-q} \int_0^t e^{-\tau} \log \left( 1 + \frac{P\tau}{Np} \right) d\tau. \quad (28)$$

Assume that on average  $R_f$  bits per sub-channel per coherence block are available for error-free CSF. Also, define  $B = NR_f$  as the average amount of feedback summed across all sub-channels. Similar to the case of two-state sub-channels, the power loading with limited CSF can be seen as a mapping from the space of channels state vector  $\mathbf{S}$  to the space of power loading vectors  $\hat{\mathbf{S}}$ , where  $\hat{S}_i = 1$  if the  $i$ -th sub-channel is activated and  $\hat{S}_i = 0$  otherwise.

We note a key difference between the VQ problem at hand and the usual stationarity assumption in rate distortion theory: The optimal choice of the threshold  $t$  here may vary with the total number of sub-channels, hence so does the statistics of the binary source denoted by probability  $q$ . Nonetheless, the following asymptotic bound on achievable distortion (equivalently, forward rates) can be established.

*Proposition 4:* Given  $N$  parallel Rayleigh fading sub-channels and an average of  $R_f$  bits of feedback per sub-channel per coherence block, the following statements hold.

- a) The forward rate per sub-channel achieved with a threshold-based feedback scheme is upper bounded by  $C$ , the maximized objective in the following optimization problem:

$$\text{maximize: } C = q(1 - \epsilon_0)(C_1 - C_0) + pC_0 \quad (29a)$$

$$\text{subject to: } H_2(p) - qH_2(\epsilon_0) - (1 - q)H_2(\epsilon_1) \leq R_f \quad (29b)$$

$$q(1 - \epsilon_0) + (1 - q)\epsilon_1 = p \quad (29c)$$

$$0 \leq \epsilon_0, \epsilon_1 \leq 1 \quad (29d)$$

with  $q = e^{-t}$ , and where the maximization is over  $p, t$  and  $\epsilon_0$ .

- b) There exist fixed-length constant composition feedback codes which achieve the forward rate per sub-channel given by (12) with sufficiently large  $N$ , where  $C$  and  $q$  are obtained from solving the optimization problem in part (a).
- c) There exist variable-length feedback codes which achieve the forward rate per sub-channel given by (13) with sufficiently large  $N$ , where  $C, q, \epsilon_0$  and  $\epsilon_1$  are obtained by solving the optimization problem in part (a).

Part (a) in Proposition 4 follows directly from the Fano's inequality and is similar to that of Proposi-

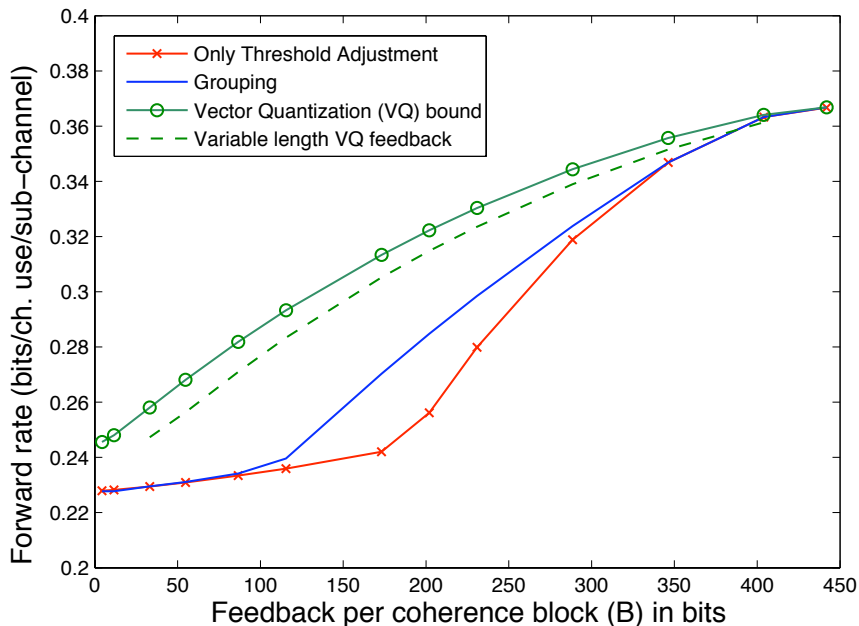


Fig. 4. Forward rate versus feedback rate for different feedback schemes for  $N = 500$  sub-channels at SNR  $P = 20$  dB. For comparison, the water-filling capacity with full channel state information at receiver is 0.385 bits per sub-channel use.

tion 1. Note that the upper bound  $C$  holds for any finite  $N$ . The lower bounds in parts (b) and (c) can be proved similarly as Propositions 2 and 3, respectively, corresponding to two-state sub-channels. The proofs are omitted. Although the lower bounds are stated for large  $N$ , more accurate expressions that apply to any finite  $N$  are derived in Appendices I and II.

Interestingly, (12) and (13) imply that the rate at which the lower bounds approach  $C$  depends on the average number of good sub-channels  $Nq$ , which can be much smaller than the total number of sub-channels  $N$ . The upper and lower bounds on forward rate versus total feedback  $B = NR_f$  per coherence block are shown in Fig. 4 for SNR  $P = 20$  dB.<sup>4,5</sup> Only the lower bound corresponding to a variable-length variable composition feedback code is shown. Unless specified otherwise, here and in the subsequent numerical results we let  $N = 500$ . The plots show that the upper and lower bounds are

<sup>4</sup>The forward rate is measured in bits per sub-channel per channel use whereas  $B$  is the total number of feedback bits per coherence block. Since typically the coherence block is several hundred channel uses, the results in Fig. 4 correspond to the practical regime in which the feedback rate is much smaller than the forward data rate.

<sup>5</sup>Numerical results for a SNR of 27 dB and  $N = 500$  (curves not shown here) show that from the case of no feedback to the case of full feedback (water-filling power allocation) the change in capacity is merely 16%. This is consistent with the understanding that adaptive power allocation does not help much at high SNRs. Of course, the gains from the various power allocation schemes presented here increase with decreasing SNR.

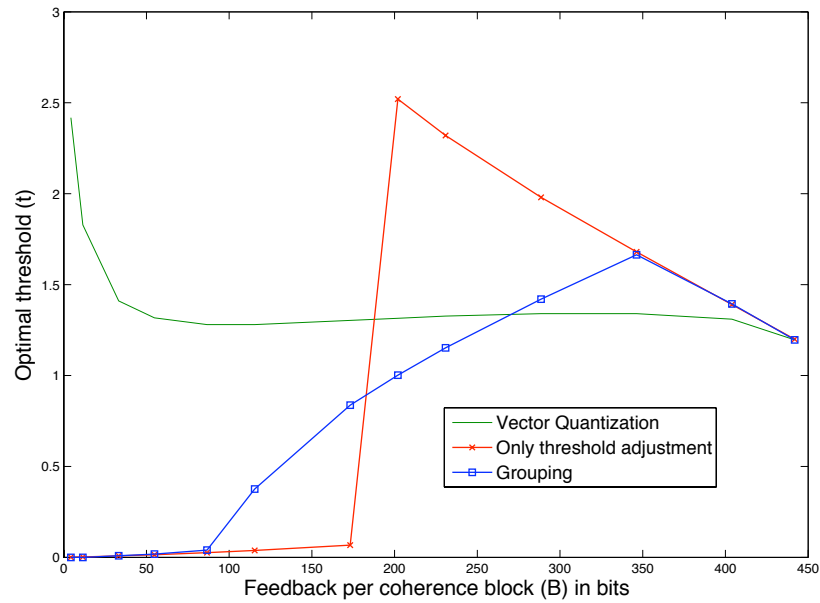


Fig. 5. Optimal threshold versus feedback for different feedback schemes with SNR = 20 dB and N = 500.

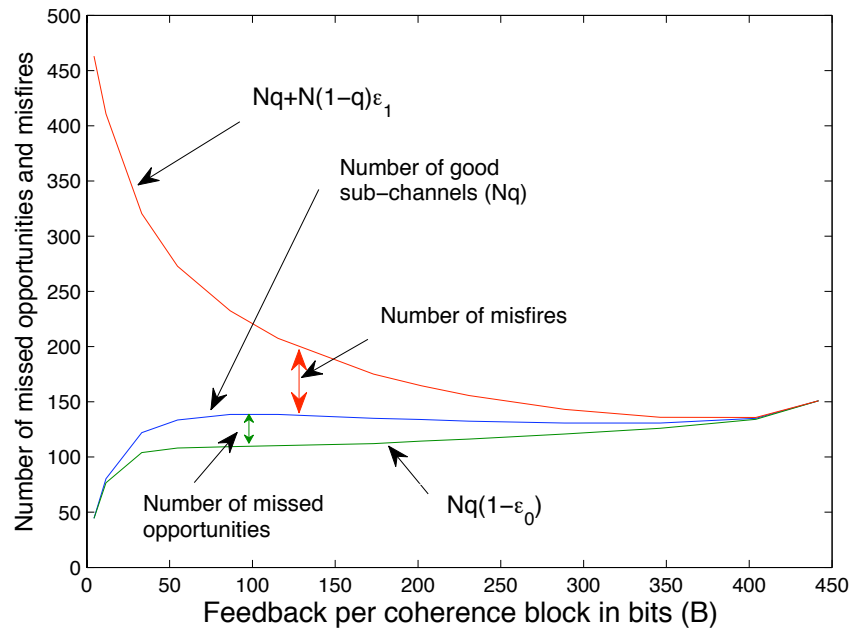


Fig. 6. Optimal number of sub-channels that exceed the threshold together with the optimal number of missed opportunities and misfires corresponding to the VQ feedback scheme for independent Rayleigh fading sub-channels. Other parameters are SNR = 20 dB and N = 500.

quite close (within 10%). The corresponding optimal threshold is shown in Fig. 5. The number of *good* sub-channels, number of missed opportunities and the number of misfires as a function of amount of feedback are shown in Fig. 6. Interestingly, for  $B \geq 70$  bits, the threshold does not change significantly but the number of missed opportunities and misfires adapt to accommodate additional feedback. Also shown in Fig. 4 are curves corresponding to other simpler sub-optimal feedback schemes to be described in subsequent sections.

### C. Lossless Coding of Feedback with Threshold Adjustment

In this section, we consider an alternative feedback scheme using lossless source coding of reduced channel states. As in Section IV-B, a threshold  $t$  is used to control the fraction of sub-channels qualifying as good (or bad) in order to meet the limited feedback constraint. All good (or bad) sub-channels are then reported to the transmitter using entropy coding. The feedback per coherence block required by this scheme is essentially  $NH_2(q)$  bits with  $q = e^{-t}$ . Note that the feedback constraint  $H_2(q) \leq R_f$  can be met by choosing either  $q \leq 1/2$  or  $q > 1/2$ . The optimal choice corresponds to the one that maximizes the forward rate

$$C(t) = N \int_t^\infty e^{-\tau} \log \left( 1 + \frac{P\tau}{Nq} \right) d\tau. \quad (30)$$

Fig. 4 plots the forward rate achieved with this scheme versus the feedback per coherence block at 20 dB. For small to moderate feedback, this scheme performs worse than the optimal VQ scheme described in the previous section. For higher feedback rates, the performance of the two schemes converge. The optimal threshold versus  $B$  for 20 dB is shown in Fig. 5. For small amounts of feedback the optimal threshold  $t$  is close to zero. Furthermore, Figs. 4 and 5 show that once the feedback crosses a certain threshold ( $B \approx 170$  bits here), the optimal threshold decreases and the capacity increases with the amount of feedback since more good sub-channels can be reported with additional feedback. As  $B$  increases further, the threshold decreases monotonically to an asymptotic value, and the capacity reaches its maximum value at around  $B = B_{max} \approx 440$  bits per coherence block. More feedback beyond this value cannot be utilized by the threshold-based scheme.<sup>6</sup>

The asymptotic forward rate versus feedback performance of this scheme (assuming that the amount of feedback and number of sub-channels go to infinity) is discussed in [2] and hence the details are omitted

<sup>6</sup>The additional bits could be used to increase the number of quantization levels for the power on active sub-channels. The corresponding increase in rate, relative to the one-bit quantization assumed here, is typically quite small [2].

here. A more refined analysis characterizing the asymptotic growth rate of the maximum achievable forward rate and  $B_{max}$  as a function of  $N$  is given in Section IV-E.

#### D. Group-Based Power Loading

Another feedback scheme for comparison is based on sub-channel groups as well as threshold-based state reduction, which is an enhancement of the scheme discussed in Section IV-C. Such a scheme was originally proposed in [23] for reducing feedback in downlink orthogonal frequency division multiple access (OFDMA) systems. The idea is to divide the sub-channels into  $G$  nonoverlapping groups, each containing  $m = N/G$  consecutive sub-channels. Given a threshold  $t$ , the receiver informs the transmitter to use only those group in which all sub-channel gains exceed  $t$ . The probability of this event is  $e^{-mt}$ , so that for large  $N$  the average amount of feedback required per coherence block for this scheme can be compressed to the entropy rate  $GH_2(e^{-mt})$ , which should not exceed the feedback constraint  $B$ .

1) *Asymptotic Rate Versus Feedback*: Assuming that the transmitter codes across coherence blocks in frequency and time, the achievable rate is given by the average mutual information (ergodic capacity),<sup>7</sup>

$$C(m, t) = Nq \int_t^\infty e^{t-\tau} \log \left( 1 + \frac{P\tau}{Nq} \right) d\tau. \quad (31)$$

Note that the rate (31) does not depend on the coherence block length  $T$ . We wish to choose the feedback parameters  $m$  and  $t$  to maximize  $C(m, t)$  subject to the feedback constraint  $GH_2(q) \leq B$ . Although it appears to be difficult to obtain an analytical characterization of the solution for arbitrary  $N$ , the solution for large  $N$  and  $B$  can be characterized as follows.

*Proposition 5*: For fixed signal-to-noise ratio  $P$ , as  $N \rightarrow \infty$  and that  $B \rightarrow \infty$  with  $N$ , the capacity (31) optimized over  $t$  and  $m$  is given by<sup>8</sup>

$$C^* = \begin{cases} \sqrt{\frac{PB}{u^*}} \log(1 + u^*) + o(1) & \text{if } B < B_1 \\ \frac{B}{\log N} \log \left( 1 + \frac{S \log N}{B} \log \left( \frac{N \log N}{B} \right) \right) + O(1) & \text{if } B_1 \leq B < B_{max} \\ P(\log N - (1 + \eta_2) \log \log N) + O(1) & \text{if } B \geq B_{max} \end{cases} \quad (32)$$

where  $B_1 = \frac{P}{u^*}(\log N)^{2-\eta_1}$ ,  $B_{max} = P(\log N)^{2+\eta_2}$ , and  $u^* \approx 3.92$  is the positive solution to  $\log(1+u) = 2u/(1+u)$ . In addition,  $\eta_1 \in (0, 2)$  and  $\eta_2 \in (0, 1)$  are functions of  $N$  such that  $\eta_1 \rightarrow 0$  and  $\eta_2 \rightarrow 1$  as  $N \rightarrow \infty$ .

<sup>7</sup>A somewhat more conservative rate is obtained by selecting the code rate assuming that all active sub-channel gains  $|h_{gi}|^2 = t$  [23]. This does not change the asymptotic results in Section IV-D.

<sup>8</sup>As  $N \rightarrow \infty$  and  $B \rightarrow \infty$  with  $N$ ,  $o(1)$  is vanishingly small and  $O(1)$  is bounded by a finite constant.

A sketch of the proof is given in Appendix IV. The appendix also provides optimal threshold and group sizes as functions of  $B$  and  $N$ , and expressions for  $\eta_1$  and  $\eta_2$ .

The capacity expressions in Proposition 5 are good approximations when  $N$  is a few hundred and  $B$  is a few tens of bits. In fact, given fixed large enough  $N$ , the results can be understood as follows: In the range of relative small to moderate amount of feedback  $B$  (specifically,  $P/u^* \ll B < (P/u^*)(\log N)^{2-\eta_1}$ ) the capacity is proportional to  $\sqrt{B}$ . If  $B$  is greater than  $(P/u^*)(\log N)^{2-\eta_1}$ , the capacity increases, albeit comparatively slower, with  $B$ . Finally, the forward rate does not increase when  $B$  exceeds  $P(\log N)^{2+\eta_2}$ . The corresponding maximum achievable rate is given by (32), which is roughly  $P \log N$ . However, the negative second-order (log log) term in (32) can be substantial, as will be seen in the subsequent numerical examples. A specific numerical example will be provided in Section IV-D.3.

2) *Optimal Threshold and Group Size:* Expressions for optimal values of the threshold  $t$  and group size  $m$  as a function of  $N$  and  $B$  are derived in Appendix IV. Here we outline the main characteristics of the optimized parameters.

As expected, when the feedback is in the small to moderate range,  $B < (P/u^*)(\log N)^{2-\eta_1}$ , the optimal group size  $m^* > 1$ . In this range, the threshold increases with feedback and is proportional to  $\sqrt{B}$ . The exact expressions for  $m$  and  $t$  are given by (76) and (77). For  $B \geq (P/u^*)(\log N)^{2-\eta_1}$ , the optimal group size  $m^* = 1$ . Hence, for this range of feedback, the group-based scheme reduces to the previous scheme described in Section IV-C. Interestingly, here the threshold is a decreasing function of  $B$  and is given by (81). As the feedback increases, decreasing the threshold beyond a certain value decreases the capacity. It is shown in Appendix IV that the optimal threshold corresponds to feedback  $B_{max} = P(\log N)^{2+\eta_2}$  and is slightly smaller than  $\log N$ . The exact expression is given by (33).

3) *Numerical Examples:* The preceding asymptotic results are illustrated in Fig. 7. The optimized group size, threshold, and corresponding capacity are plotted as function of the amount of feedback. These results are obtained by optimizing the original capacity expression (31) subject to the feedback constraint.<sup>9</sup> Fig. 7 also shows the asymptotic analytical results. (The values plotted here are refined versions of the expressions presented in Proposition 5 and are derived in Appendix IV.) The plot shows that the asymptotic values are close to the values obtained from numerical optimization. As predicted by Proposition 5, the plot shows that as  $B$  increases from zero, the group size decreases, and the threshold increases. However, once the group size crosses one (when  $B$  is about 40 bits/coherence block), the threshold decreases with  $B$  and the capacity increases relatively slowly. Finally, for large amounts of

<sup>9</sup>Of course, in practice  $m$  can assume only positive integer values, as opposed to the real values obtained from the optimization, which are shown in Fig. 7.

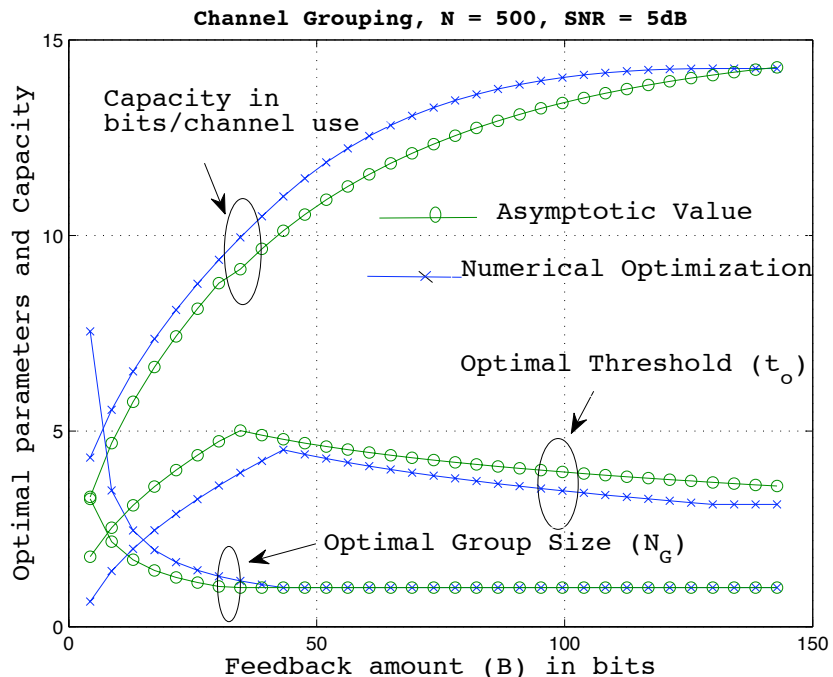


Fig. 7. Comparison of numerically optimized values and asymptotic values versus feedback per coherence block for  $N = 500$  and  $P = 5$  dB SNR. For comparison, the water-filling capacity is 14.93 bits per channel use.

feedback (say, greater than 135 bits/coherence block for this example) the capacity and threshold saturate, and increasing the feedback further does not improve performance. Referring to (79)-(83), these values correspond to  $\eta_1 \approx \eta_2 \approx 0.25$ .

4) *Performance comparison:* The optimized forward rate (31) versus  $B$  for the group-based scheme is shown in Fig. 4. At 20 dB SNR, grouping sub-channels can provide about 15% gain over threshold adjustment alone for small to moderate feedback rates. As the feedback increases, the two schemes become the same (since the group size converges to  $m = 1$  at around  $B \approx 340$  bits). Recall that the advantage of the group-based scheme over threshold adjustment alone is limited to the feedback range of  $B \leq (P/u^*)(\log N)^{2-\eta_1}$ . The optimal threshold for the group-based scheme is shown in Fig. 5. Observe that the optimal thresholds with and without grouping converge for  $B \geq 340$ . The thresholds behave in strikingly different manners for the two schemes when feedback is smaller.

It is also seen in Fig. 4 that the VQ scheme performs substantially better than both LSC schemes with and without grouping for small to moderate amounts of feedback. Namely, VQ saves between 100 and 150 feedback bits per coherence block over a wide range of target forward rates.

### E. Growth in achievable forward rate

In this section we highlight several common features of all three schemes discussed in Section IV. With sufficient amount of feedback, all three schemes correspond to the optimal “on-off” power allocation in which the power is uniformly spread over active channels [2], [24]. The optimal threshold is given by

$$t^* = \lceil \log N - (1 + \eta_2) \log \log N - \log P \rceil + o(1) \quad (33)$$

and the corresponding capacity is given by (32) for  $B \geq B_{max}$  (see Appendix IV).

From (83) in Appendix IV, as  $N \rightarrow \infty$ ,  $\eta_2 \rightarrow 1$ , so that (32) states that  $O(\log^3 N)$  feedback can achieve the optimal  $O(\log N)$  growth in achievable rate. This result has been previously presented in [2], which considers the same threshold-based feedback scheme considered here without grouping ( $m = 1$ ). Also, the numerical examples given in the previous section show that for reasonable values of  $N$ , the amount of feedback needed to achieve the  $O(P \log N)$  forward rate may be closer to  $P \log^2 N$  than to  $P \log^3 N$ .

At the other extreme of small feedback  $B \rightarrow 0$ , the three schemes also perform similarly and the forward rate converges to the SNR  $P$  (see Fig. 4). This limit is the ergodic capacity of a Rayleigh fading channel without feedback when the bandwidth becomes large, i.e.,  $N \rightarrow \infty$ . For the VQ scheme, as  $B \rightarrow 0$ , the optimal parameters converge to either  $t \rightarrow 0, \epsilon_0 \rightarrow 0$ , or  $t \rightarrow \infty, \epsilon_1 \rightarrow 1$ , both of which imply the same transmission strategy. Clearly, for threshold adjustment without grouping, the optimal threshold  $t^* \rightarrow 0$  as feedback becomes small. Finally, although Proposition 5 does not cover the case of finite  $B$ , it is easy to show that for the group-based scheme, as the amount of feedback  $B \rightarrow 0$ , we have  $m \rightarrow N$ ,  $t \rightarrow 0$  and the achievable rate  $C \rightarrow P$ .

## V. CORRELATED RAYLEIGH FADING SUB-CHANNELS

In this section, we remove the assumption that the Rayleigh fading sub-channels are independent. Suppose the sequence of complex sub-channel coefficients is a Gauss-Markov process generated by the following first-order autoregressive model,

$$H_i = \alpha H_{i-1} + \sqrt{1 - \alpha^2} W_i, \quad i = 2, \dots, N, \quad (34)$$

where  $W_i$  are i.i.d. zero-mean CSCG random variables with unit variance, and  $\alpha \in (0, 1)$  represents the correlation between the sub-channels. Each sub-channel gain  $|H_i|^2$  is exponentially distributed. The sequence of sub-channel gains  $\{|H_i|^2\}$  is a Markov process with joint second order probability density



function given by

$$g(x, y) = \frac{1}{1 - \alpha^2} e^{-\frac{x+y}{1-\alpha^2}} I_0 \left( \frac{2\alpha\sqrt{xy}}{1 - \alpha^2} \right) \quad (35)$$

where  $I_0(\cdot)$  is the modified Bessel function of the first kind and zero-order.

Again, given a threshold  $t \geq 0$ , the state vector  $\mathbf{S}$  is defined such that  $S_i = 1$  if  $|H_i|^2 \geq t$  and  $S_i = 0$  otherwise. Also,  $q = \Pr\{S_i = 1\} = e^{-t}$ . The sequence  $\{S_i\}$  is a *hidden* Markov process rather than i.i.d. Nonetheless, for a fixed  $p$  and  $t$ , the rate-distortion trade-off is still given by (16). However, obtaining upper and lower bounds on the required feedback rate (16), analogous to the two state Markov model of Section III, seems to be difficult due to the hidden Markov structure of  $\mathbf{S}$ .

We proceed by assuming that the receiver approximates the sequence  $\{S_i\}$  as a first-order Markov chain. Ignoring the higher order correlation in  $\mathbf{S}$  results in a larger feedback requirement (or equivalently, an upper bound on  $R(D)$  in (16)). Using this upper bound as the feedback rate then gives an *achievable* forward rate (a lower bound on capacity). The transition probabilities for the first-order Markov model are,

$$\delta_{10} = \Pr\{S_i = 0 | S_{i\pm 1} = 1\} = 1 - \frac{1}{q} \int_t^\infty \int_t^\infty g(x, y) dx dy, \quad (36)$$

$$\delta_{01} = \Pr\{S_i = 1 | S_{i\pm 1} = 0\} = \frac{q\delta_{10}}{1 - q}. \quad (37)$$

With the first-order Markov approximation, the problem reduces to the one discussed in Section III. Consequently, an achievable forward rate can be computed by solving the optimization problem (29) over  $p, t, \epsilon_0$  and  $\epsilon_1$  with  $q = e^{-t}$  and the constraint (10b) replaced by  $i_u(\epsilon_0, \epsilon_1) \leq R_f$ . An expression for  $i_u(\epsilon_0, \epsilon_1)$  is given in Appendix III.

Fig. 8 plots this rate versus feedback per coherence block  $B = NR_f$  for  $\alpha = 0.6$  and 20 dB SNR. Also shown in Fig. 8 is the forward rate achieved with the simpler LSC feedback scheme that encodes the state vector  $\mathbf{S}$  and controls the amount of feedback by adjusting the threshold without grouping. Allowing no errors in the reconstructed channel state vector means that the minimum required feedback rate is  $H(\mathbf{S})$ , which is upper bounded by  $H(S_1 | S_0) = qH_2(\delta_{10}) + (1 - q)H_2(\delta_{01})$ . This feedback scheme was considered for correlated sub-channels in [24]. In Fig. 8, VQ performs substantially better at small to moderate feedback rates. The forward rate values were also computed for  $\alpha = 0.9$  (not shown in Fig. 8), but there the comparison is inconclusive, since the lower bound on forward rate achieved with VQ is very close to the forward rate achieved with the LSC scheme.

The behavior of the optimal threshold and crossover probabilities  $\epsilon_0, \epsilon_1$  is the same as for independent

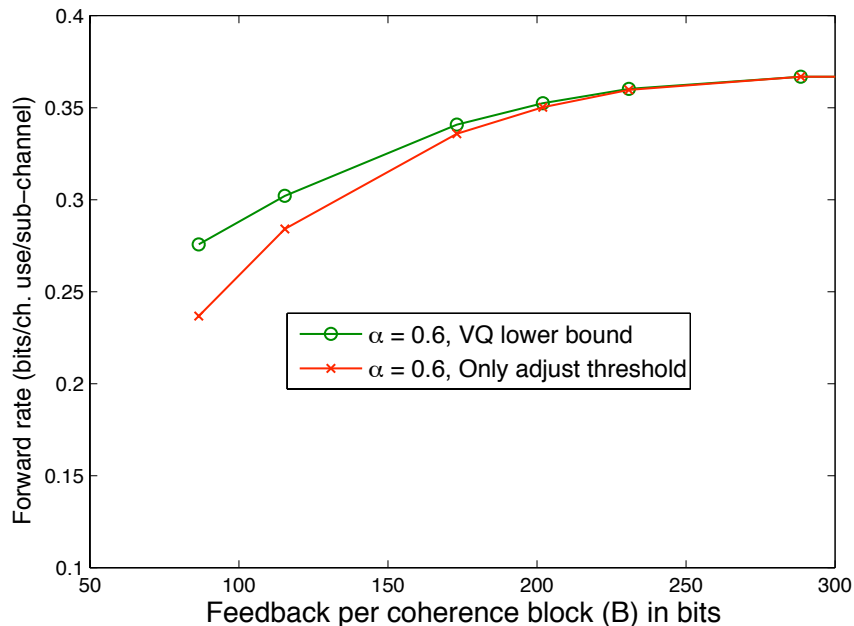


Fig. 8. Achievable forward rate versus the feedback rate for a VQ scheme with correlated fading sub-channels. Also shown is the forward rate achieved with the threshold adjustment feedback scheme without grouping. Other parameters are  $N = 500$ ,  $P = 20$  dB SNR. For comparison, the water-filling capacity with full channel state information at receiver is 0.385 bits per sub-channel use (the same as in Fig. 4).

sub-channels and hence is not shown here. Similar to the independent sub-channels case, we can again apply the group-based scheme to correlated sub-channels. Performance evaluation appears to be difficult, but the performance of the group-based scheme should be inferior to that of the VQ scheme and at least as good than without grouping.

## VI. CONCLUSIONS

We have studied limited feedback of channel states for multicarrier systems with two-state and Rayleigh sub-channels. The asymptotic performance has been characterized, using rate distortion theory, along with bounds on the performance loss with a finite number of sub-channels. For Rayleigh channels, the threshold-based VQ scheme shows a substantial improvement over lossless coding of reduced channel states based on thresholding, especially at low to moderate SNRs. Of course, this benefit comes at the price of higher source coding complexity (e.g., using graphical codes).

Our results have assumed perfect channel knowledge at the receiver and have neglected the feedback overhead. In practice, the combined overhead for channel estimation and feedback compromises the

benefits of feedback. This becomes more important as the channel coherence time decreases. A model, which accounts for feedback overhead in a multicarrier time-division duplex system is presented in [25]. There the overhead is optimized, assuming the lossless feedback code presented in Section IV-D. Training and feedback overhead in the context of beamforming has been studied in [26], [27]. That approach may also be appropriate for the multicarrier scenario considered here.

We have also assumed a noiseless feedback link. For the schemes considered here a noisy feedback link requires additional overhead in the form of channel coding or higher transmit power. Other alternatives include analog CSF (e.g., see [28], [29]), which gives noisy estimates at the transmitter, and sending a pilot signal from receiver to transmitter at the beginning of each coherence block [30] (assuming channel reciprocity applies). Comparative advantages and disadvantages of these schemes remain to be studied.

Finally, the results presented here can conceivably be extended to more elaborate system and channel models, such as continuous fading, instead of block fading (e.g., see [31], [32]), and Multiple-Input Multiple-Output (MIMO) OFDM. Such systems typically operate at low SNRs per antenna (or coefficient) and hence can benefit substantially from adaptive power loading [33], [34]. Limited CSF schemes for downlink and uplink OFDMA are presented in [23], [35], [36] and the references therein. (See also the comprehensive survey of the limited feedback literature in [37].) Extensions of the VQ scheme presented here to those settings is also left for future work.

## APPENDIX I

### PROOF OF PROPOSITION 2

First we introduce some notation, then describe the construction of a fixed-length, constant-composition feedback code, and finally give the performance bounds as a function of  $N$ .

Let  $\mathcal{P}_{\mathbf{s}}(1)$  represent the *empirical probability* of ones in a length  $N$  binary vector channel state  $\mathbf{s}$ . Namely,  $\mathcal{P}_{\mathbf{s}}(1) = q$  implies that  $\sum_{i=1}^N s_i = qN$ . A similar definition holds for  $\mathcal{P}_{\hat{\mathbf{s}}}(1)$ . Furthermore, for any pair of vectors  $(\hat{\mathbf{s}}, \mathbf{s})$ ,  $\mathcal{P}_{\hat{\mathbf{s}}|\mathbf{s}}(0|1)$  represents the empirical probability of zeroes in  $\hat{\mathbf{s}}$  at the positions where  $\mathbf{s}$  has ones. In other words,  $\mathcal{P}_{\hat{\mathbf{s}}|\mathbf{s}}(0|1) = \epsilon_0$  implies that  $\sum_{i=1}^N 1_{\{\hat{s}_i=0, s_i=1\}} = \epsilon_0 \mathcal{P}_{\mathbf{s}}(1)N$ . Let  $\mathcal{T}_{\mathbf{S};q}$  represent the set of all constant composition length  $N$  channel state vectors  $\mathbf{s}$  which have  $qN$  number of ones, that is,  $\mathcal{T}_{\mathbf{S};q} = \{\mathbf{s} : \mathcal{P}_{\mathbf{s}}(1) = q\}$ . Similarly, define the set of constant composition power loading vectors  $\mathcal{T}_{\hat{\mathbf{S}};p} = \{\hat{\mathbf{s}} : \mathcal{P}_{\hat{\mathbf{s}}}(1) = p\}$ .

Let the *feedback codebook* be a subset  $\mathcal{T}_{\hat{\mathbf{S}}|\mathbf{S}} \subset \mathcal{T}_{\hat{\mathbf{S}};p}$ . The amount of feedback required per coherence block is therefore  $\log_2 |\mathcal{T}_{\hat{\mathbf{S}}|\mathbf{S}}|$  bits. A vector  $\hat{\mathbf{s}} \in \mathcal{T}_{\hat{\mathbf{S}}|\mathbf{S}}$  is said to *cover*  $\mathbf{s} \in \mathcal{T}_{\mathbf{S};q}$  if  $\mathcal{P}_{\hat{\mathbf{s}}|\mathbf{s}}(0|1) = \epsilon_0$  and  $\mathcal{P}_{\hat{\mathbf{s}}|\mathbf{s}}(1|0) = \epsilon_1$ . Note that depending on the size of the codebook, all the vectors in  $\mathcal{T}_{\mathbf{S};q}$  might not be

covered. The feedback occurs as follows: Every channel realization vector is matched to the closest (in hamming distance) typical vector  $\mathbf{s} \in \mathcal{T}_{\mathcal{S};q}$ . The receiver then looks up a  $\hat{\mathbf{s}}$  in  $\mathcal{T}_{\hat{\mathcal{S}}|\mathcal{S}}$  which covers  $\mathbf{s}$  feeds back the index of  $\hat{\mathbf{s}}$ . If there is no such  $\hat{\mathbf{s}}$  then a random index is fed back. Therefore, the elements of the codebook  $\mathcal{T}_{\hat{\mathcal{S}}|\mathcal{S}}$  should be chosen carefully to minimize the distortion or, equivalently, maximize the forward rate. Next, along the lines of [11], we will evaluate the average performance assuming that the subset  $\mathcal{T}_{\hat{\mathcal{S}}|\mathcal{S}}$  is chosen at random. Then by usual argument we can claim that there must exist a structured fixed-length constant composition codebook that does at least as well.

Consider the performance of the codebook as a function of  $N$ . The Type covering lemma [38] suggests that we can find a codebook with size  $|\mathcal{T}_{\hat{\mathcal{S}}|\mathcal{S}}| = g(N)2^{i(\epsilon_0, \epsilon_1)N}$ , where  $g(N)$  is a polynomial, such that every vector in  $\mathcal{T}_{\mathcal{S};q}$  is covered. Here, we use random coding arguments to get an estimate of  $g(N)$ . Suppose  $\mathcal{T}_{\hat{\mathcal{S}}|\mathcal{S}}$  consists of  $M$  vectors that are randomly and independently<sup>10</sup> drawn from  $\mathcal{T}_{\hat{\mathcal{S}};p}$ .

Given a vector  $\mathbf{s} \in \mathcal{T}_{\mathcal{S};q}$ , probability that it will be not be *covered* by the  $M$  randomly chosen vectors is given by

$$p_n = [1 - p_c]^M \leq e^{-Mp_c} \quad (38)$$

where

$$p_c = \frac{\binom{qN}{\epsilon_0 q N} \binom{(1-q)N}{\epsilon_1 (1-q)N}}{\binom{N}{pN}}. \quad (39)$$

Further using Robbin's approximation [38], [39] for the factorial

$$\sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n+\frac{1}{12(n+1)}} \leq n! \leq \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n+\frac{1}{12n}}, \quad n \geq 1, \quad (40)$$

it is straightforward to show that

$$p_c \geq N^{-1/2} 2^{-K_2} 2^{-i(\epsilon_0, \epsilon_1)N} \quad (41)$$

where the mutual information  $i(\epsilon_0, \epsilon_1)$  is given by (8) and  $K_2 = \log_2(\sqrt{2\pi}e^{5/12}) - \frac{1}{2} \log_2(p(1-p))$  is a constant. Now choosing  $M = (\log N)N^{1/2}2^{K_2} 2^{i(\epsilon_0, \epsilon_1)N}$  and combining (38) and (41) gives  $p_n \leq 1/N$ . Therefore, with this codebook the feedback rate per sub-channel per coherence block given by

$$\frac{\log_2 |\mathcal{T}_{\hat{\mathcal{S}}|\mathcal{S}}|}{N} = \frac{\log_2 M}{N} = i(\epsilon_0, \epsilon_1) + \frac{(\log_2 N + 2K_2 + 2 \log_2(\log N))}{2N} \quad (42)$$

converges to the mutual information  $i(\epsilon_0, \epsilon_1)$ .

<sup>10</sup>We have  $\binom{N}{Np}$  choices in each drawing.

Next we bound the average forward rate achieved by this codebook. For this we need to account for the variations in the channel gain vector  $\mathbf{S}$ . Consider the set of channel gain vectors with fraction of ones in the range  $(q(1 - \epsilon), q(1 + \epsilon))$ , that is, say  $\mathcal{T}_{\mathbf{S};q}^\epsilon = \{\mathbf{s} : \mathcal{P}_{\mathbf{s}}(1) \in (q(1 - \epsilon), q(1 + \epsilon))\}$ . Using Chernoff's inequality [40], the following is easily seen

$$\Pr \left\{ \left| \sum_{i=1}^N S_i - qN \right| \geq q\epsilon N \right\} \leq p_u = 2 \exp \left( -\frac{qN\epsilon^2}{4} \right), \quad 0 \leq \epsilon \leq 2(1 - q), \quad (43)$$

which in turn implies that  $\Pr\{\mathcal{T}_{\mathbf{S};q}^\epsilon\} \geq 1 - p_u$ . Using the definition of  $\epsilon$  above, the ergodic capacity is lower bounded as

$$C_{fixed} \geq q(1 - \epsilon_0 - \epsilon)(C_1 - C_0)(1 - p_u)(1 - p_n) + pC_0 \quad (44)$$

$$\geq C - C \left( \epsilon + p_u + \frac{1}{N} \right), \quad (45)$$

where  $C = q(1 - \epsilon_0)(C_1 - C_0) + pC_0$ . The loss factors  $1 - p_n$  and  $1 - p_u$  in (44) account for the fact that channel gain vectors might not be covered or, might not be in the set  $\mathcal{T}_{\mathbf{S};q}^\epsilon$ , respectively. Further, choosing  $\epsilon = \sqrt{\frac{2 \log(Nq)}{Nq}}$  gives a tight lower bound as

$$C_{fixed} \geq C - C \left( \frac{\sqrt{2 \log(Nq)} + 2}{\sqrt{Nq}} + \frac{1}{N} \right). \quad (46)$$

In summary, since (46) bounds the average performance of a randomly generated codebook with  $M$  codewords, there must exist a codebook of size  $M$  which performs at least as good as this lower bound.

Lastly, we consider the convergence rate to the upper bound given by Proposition 1. Note that required feedback rate (42) is more than mutual information  $i(\epsilon_0, \epsilon_1)$ . If we do not wish to allow the feedback to exceed  $i(\epsilon_0, \epsilon_1)$  bits per sub-channel, an additional distortion of  $\delta$  can be introduced so that  $\epsilon_0$  and  $\epsilon_1$  are replaced by  $\epsilon_0 + \delta$  and  $\epsilon_1 + \delta'$ , respectively, where  $\delta' = q\delta/(1 - q)$  (so that the input power constraint (10c) is satisfied). Using Taylor series expansion, for a small enough  $\delta$  we have

$$-2q[H_2'(\epsilon_0) + H_2'(\epsilon_1)]\delta \leq i(\epsilon_0 + \delta, \epsilon_1 + \delta') - i(\epsilon_0, \epsilon_1) \leq -\frac{1}{2}q[H_2'(\epsilon_0) + H_2'(\epsilon_1)]\delta. \quad (47)$$

Substituting  $\epsilon_0 + \delta$  and  $\epsilon_1 + \delta'$  for  $\epsilon_0$  and  $\epsilon_1$  in (42) and using (47), for large enough  $N$  and small enough  $\delta$  we have

$$\frac{\log_2 |\mathcal{T}_{\hat{\mathbf{S}}|\mathbf{S}}|}{N} \leq i(\epsilon_0, \epsilon_1) - \frac{1}{2}q[H_2'(\epsilon_0) + H_2'(\epsilon_1)]\delta + \frac{\log_2 N}{N}. \quad (48)$$

Now choosing  $\delta = K_1 \frac{\log_2 N}{N}$  where  $K_1 = \frac{2}{q[H_2'(\epsilon_0) + H_2'(\epsilon_1)]}$ , gives that the number of feedback bits per sub-channels  $\frac{\log_2 |\mathcal{T}_{\hat{\mathbf{S}}|\mathbf{S}}|}{N} \leq i(\epsilon_0, \epsilon_1)$ . The additional distortion of  $\delta$  will result in a loss in capacity which

can be quantified by replacing  $\epsilon_0$  by  $\epsilon_0 + \delta$  in (44) which gives a lower bound on forward rate as

$$C_{fixed} \geq C - C \left( \frac{\sqrt{2 \log(Nq)} + 2}{\sqrt{Nq}} + \frac{1}{N} + K_1 \frac{\log_2 N}{N} \right) \quad (49)$$

which for large enough  $N$  can be further lower bounded as (12).

## APPENDIX II

### PROOF OF PROPOSITION 3

We again resort to the random coding techniques as in [12], assuming that corresponding to each channel state vector  $\mathbf{s}$ , the power loading vector  $\hat{\mathbf{s}}$  is produced with Bernoulli- $p$  distribution. A randomly generated codeword  $\hat{\mathbf{s}}$  is admitted only if it satisfies the empirical probabilities  $\mathcal{P}_{\hat{\mathbf{s}}|\mathbf{s}}(0|1) = \epsilon_0$  and  $\mathcal{P}_{\hat{\mathbf{s}}|\mathbf{s}}(1|0) = \epsilon_1$ , where  $\epsilon_0$  and  $\epsilon_1$  are fixed and  $q(1 - \epsilon_0) + (1 - q)\epsilon_1 = p$ . This will ensure that averaged over all the state vectors, the number of sub-channels used for transmission are given by  $pN$  and average number of unused good sub-channels are  $q\epsilon_0 N$ . Next, we shall find the expected variable-length encoder rate, averaged over this ensemble of codes, and then, by usual argument, we can assert that there must exist at least one set of  $\{\hat{\mathbf{s}}\}$  that gives the performance as good as the average.

Let  $L$  represent the random variable denoting the fraction of ones in the state vector  $\mathbf{S}$ . Clearly,  $H(L) \leq \log_2(N + 1)$ . If a variable-length feedback codebook with average rate of  $R_f$  bits per sub-channel per coherence block is used, then we can write  $NR_f \leq H(\hat{\mathbf{S}}) + 1$ . The rate can be further upper bounded as  $NR_f \leq H(\hat{\mathbf{S}}, L) + 1 = H(\hat{\mathbf{S}}|L) + H(L) + 1$ . Averaging over the random code book selection we get that,

$$NR_f \leq E_{\hat{\mathbf{S}}}[H(\hat{\mathbf{S}}|L)] + H(L) + 1. \quad (50)$$

Corresponding to a channel state vector with  $L = l$ , define  $q_l$  as the probability that a randomly drawn codeword  $\hat{\mathbf{s}}$  is admissible. Then we have,

$$q_l = \binom{lN}{\epsilon_0 lN} \binom{(1-l)N}{\epsilon_1(1-l)N} p^{n(l)N} (1-p)^{(1-n(l))N}, \quad (51)$$

where  $n(l) = (1 - \epsilon_0)l + \epsilon_1(1 - l)$ . Further, it is argued in [12] that given the geometric distribution  $p(k|l) = q_l(1 - q_l)^{k-1}$  we have

$$E_{\hat{\mathbf{S}}}[H(\hat{\mathbf{S}}|l)] = - \sum_{k=1}^{\infty} p(k|l) \log_2 p(k|l) \quad (52)$$

$$\leq -\log_2 q_l + \log_2 e. \quad (53)$$

Combining (50) and (53) and using the fact that  $H(L) \leq \log_2(N+1)$  we have,

$$NR_f \leq - \sum_{l=0}^N (\log_2 ql) \Pr\{L=l\} + \log_2 e + \log_2(N+1) + 1. \quad (54)$$

Further, applying Robbin's approximation (40) to (51) we have

$$\begin{aligned} -\log_2(ql) &\leq -lNH_2(\epsilon_0) - (1-l)NH_2(\epsilon_1) \\ &\quad + n(l)N \log_2(p) + (1-n(l))N \log_2(1-p) + \log_2 N + K_3, \end{aligned} \quad (55)$$

where  $K_3 = \frac{1}{2} \log_2[e^8(2\pi)^2\epsilon_0\epsilon_1(1-\epsilon_0)(1-\epsilon_1)]$ . Substituting (55) into (54) and using the fact that  $E[n(l)] = p$  and  $E[l] = q$ , we get,

$$R_f \leq i(\epsilon_0, \epsilon_1) + \frac{1}{N}(\log_2 e + K_3 + \log_2 N + \log_2(N+1) + 1). \quad (56)$$

Since this rate exceeds  $i(\epsilon_0, \epsilon_1)$ , similar to Appendix I, we can introduce additional distortion so that  $\epsilon_0$  and  $\epsilon_1$  are replaced by  $\epsilon_0 + \delta$  and  $\epsilon_1 + \frac{q\delta}{1-q}$ . Therefore, using (47), for large enough  $N$  and small enough  $\delta$ , (56) yields

$$R_f \leq i(\epsilon_0, \epsilon_1) - \frac{1}{2}q[H'_2(\epsilon_0) + H'_2(\epsilon_1)]\delta + \frac{3\log_2 N}{N} \quad (57)$$

Finally, choosing  $\delta = \frac{6\log_2 N}{q[H'_2(\epsilon_0) + H'_2(\epsilon_1)]N}$  gives  $R_f \leq i(\epsilon_0, \epsilon_1)$  and capacity as (13).

### APPENDIX III

#### DERIVATION OF $i_l(q_{00}, q_{01}, q_{10}, q_{11})$ AND $i_u(\epsilon_0, \epsilon_1)$

The lower bound can be explicitly computed as follows

$$i_l(q_{00}, q_{01}, q_{10}, q_{11}) = I(S_1; \hat{S}_1|S_0) \quad (58)$$

$$= H(\hat{S}_1|S_0) - H(\hat{S}_1|S_1, S_0). \quad (59)$$

Each entropy term is further computed as

$$H(\hat{S}_1|S_0) = (1-q)H_2\left(\frac{q_{00} + q_{01}}{1-q}\right) + qH_2\left(\frac{q_{10} + q_{11}}{q}\right) \quad (60)$$

and

$$\begin{aligned} H(\hat{S}_1|S_1, S_0) &= (1-q)(1-\delta_{01})H_2\left(\frac{q_{00}}{(1-q)(1-\delta_{01})}\right) + q\delta_{10}H_2\left(\frac{q_{10}}{q\delta_{10}}\right) \\ &\quad + (1-q)\delta_{01}H_2\left(\frac{q_{01}}{(1-q)\delta_{01}}\right) + q(1-\delta_{10})H_2\left(\frac{q_{11}}{q(1-\delta_{10})}\right). \end{aligned} \quad (61)$$

Next we compute the upper bound  $f_2(\epsilon_0, \epsilon_1)$ . Recall that, in order to arrive at the upper bound, we have assumed that conditioned on  $S_i$ ,  $\hat{S}_i$  is independent of all other elements in  $\mathbf{S}$ , thus

$$i_u(\epsilon_0, \epsilon_1) = I(S_1; S_2, \hat{S}_1 | S_0) \quad (62)$$

$$= H(S_2, \hat{S}_1 | S_0) - H(S_2, \hat{S}_1 | S_1, S_0) \quad (63)$$

$$= H(S_2 | S_0) + H(\hat{S}_1 | S_0, S_2) - H(\hat{S}_1 | S_1) - H(S_2 | S_1). \quad (64)$$

Each entropy term can be further computed as

$$H(S_2 | S_0) = qH_2((1 - \delta_{10})^2 + \delta_{10}\delta_{01}) + (1 - q)H_2(\delta_{01}(1 - \delta_{10}) + (1 - \delta_{01})\delta_{01}), \quad (65)$$

$$H(\hat{S}_1 | S_1) = qH_2(\epsilon_0) + (1 - q)H_2(\epsilon_1), \quad (66)$$

$$H(S_2 | S_1) = qH_2(\delta_{10}) + (1 - q)H_2(\delta_{01}) \quad (67)$$

and

$$\begin{aligned} H(\hat{S}_1 | S_0, S_2) &= ((1 - \delta_{01})^2(1 - q) + \delta_{10}^2 q) H_2(w_{00}) \\ &\quad + 2(\delta_{01}(1 - \delta_{01})(1 - q) + \delta_{10}(1 - \delta_{10})q) H_2(w_{01}) \\ &\quad + (\delta_{01}^2(1 - q) + (1 - \delta_{10})^2 q) H_2(w_{11}), \end{aligned} \quad (68)$$

where the probabilities in the argument of binary entropy functions are defined as,

$$w_{s_0 s_2} = P_{\hat{S}_1 | S_0, S_2}(0 | s_0, s_2), \quad s_0, s_2 = 0 \text{ or } 1. \quad (69)$$

We note that  $w_{10} = w_{01}$  and the probabilities can be explicitly computed as

$$w_{00} = \frac{(1 - \epsilon_1)(1 - \delta_{01})^2(1 - q) + \epsilon_0 \delta_{10}^2 q}{(1 - \delta_{01})^2(1 - q) + \delta_{10}^2 q}, \quad (70)$$

$$w_{01} = \frac{(1 - \epsilon_1)\delta_{01}(1 - \delta_{01})(1 - q) + \epsilon_0 \delta_{10}(1 - \delta_{10})q}{\delta_{01}(1 - \delta_{01})(1 - q) + \delta_{10}(1 - \delta_{10})q}, \quad (71)$$

$$w_{11} = \frac{(1 - \epsilon_1)\delta_{01}^2(1 - q) + \epsilon_0(1 - \delta_{10})^2 q}{\delta_{01}^2(1 - q) + (1 - \delta_{10})^2 q}. \quad (72)$$



## APPENDIX IV

## PROOF SKETCH OF PROPOSITION 5

Consider the maximization of the capacity  $C(m, t)$  in (31) over the group size  $m$  and threshold  $t$  subject to  $GH(q) \leq B$ . The capacity can be bounded as

$$Nq \log \left( 1 + \frac{Pt}{Nq} \right) \leq C(m, t) \leq Nq \log \left( 1 + \frac{P(t+1)}{Nq} \right). \quad (73)$$

The lower bound is simply by observing that the logarithm term in (31) takes on its minimum value at the boundary  $\tau = t$ , whereas the exponential term integrates to 1. The upper bound can be shown using the fact that  $\int_0^\infty e^{-x} \log(x+a) d\tau < \log(1+a)$  for all  $a > 0$ . Clearly, the maximum value of  $C(m, t)$  subject to  $GH(q) \leq B$  is no greater than the maximum value of the upper bound of  $C(m, t)$  in (73) subject to the same constraint. Next we obtain a solution to the latter optimization problem and show that for large  $B$  and  $N$ , it provides a good approximation to the solution of the original optimization problem. Without loss of generality, substituting  $w = Nq$  the optimization problem can be written as

$$\max_{w, t} \bar{C} = w \log \left( 1 + \frac{P(t+1)}{w} \right), \text{ subject to: } wt \leq B. \quad (74)$$

Assuming that the feedback constraint is tight, i.e.,  $wt = B$ , the optimum  $w$  must satisfy

$$(1 + u + P/w) \log(1 + u + P/w) = (1 + 2u + P/w), \quad (75)$$

where  $u = PB/w^2$ . A closed-form solution to (75) seems difficult, however, insight can be obtained by assuming that  $N, B$  are large. In addition, we assume that the optimal  $w$  increases with  $B$  such that  $u \gg P/w$  or, equivalently, as  $B \rightarrow \infty$ ,  $w/B \rightarrow 0$ . We will see later that this is indeed true. Therefore, observing that the  $P/w$  terms in (75) are small compared to  $u$ , the optimal  $w^* = \sqrt{\frac{PB}{u^*}} + o(1)$ , where  $o(1)$  is vanishingly small as  $B \rightarrow \infty$ , and  $u^*$  is the solution to  $(1 + u^*) \log(1 + u^*) = (1 + 2u^*)$ . Since we assume that  $wt = B$ , solving (74) gives

$$t^* = \sqrt{\frac{u^* B}{P}} + o(1) \quad (76)$$

$$m^* = \sqrt{\frac{P}{u^* B}} \log \frac{N}{\sqrt{P B / u^*}} + o(1) \quad (77)$$

$$\bar{C} = \sqrt{\frac{PB}{u^*}} \log(1 + u^*) + o(1). \quad (78)$$

The parameter values satisfy the original feedback constraint  $GH(q) = B$  and in fact the lower bound in (73) also behaves as (78) (although the value associated with the  $o(1)$  term change). This implies that

the optimal parameters that maximize the capacity  $C(m, t)$  satisfy (76) and (77), and the capacity  $C^*$  is approximated by (78) to within a vanishingly small term.

Note that (77) implies that for  $m^* > 1$  we should have feedback in the range  $B < \frac{P}{u^*}(\log N)^{2-\eta_1}$  for large  $N$  where  $\eta_1 \in (0, 2)$  is the solution to

$$\log N - \log \left[ \frac{P}{u^*} (\log N)^{1-\frac{\eta_1}{2}} \right] = (\log N)^{1-\frac{\eta_1}{2}}. \quad (79)$$

It is easy to see that  $\eta_1 \rightarrow 0$  as  $N \rightarrow \infty$ . Next we solve for the optimal parameters when  $B \geq \frac{P}{u^*}(\log N)^{2-\eta_1}$ . Again we first solve the upper bound maximization problem (74) with  $m = 1$  or, equivalently  $q = e^{-t}$ . Namely

$$\max_t \bar{C} = N e^{-t} \log \left( 1 + \frac{P(t+1)}{N e^{-t}} \right), \text{ subject to: } N t e^{-t} \leq B. \quad (80)$$

Assuming that the feedback constraint is tight, i.e.,  $N t e^{-t} = B$ , we get the optimal threshold and upper bound on capacity

$$t^* = \log \frac{N \log N}{B} + o(1) \quad (81)$$

$$\bar{C} = \frac{B}{\log N} \log \left( 1 + \frac{P \log N}{B} \log \frac{N \log N}{B} \right) + O(1) \quad (82)$$

Again, it can be checked that with appropriate adjustments to the  $o(1)$  and  $O(1)$  terms in (81) and (82), respectively, the threshold (81) satisfies the original feedback constraint  $NH(e^{-t}) = B$  and the lower bound in (73) also behaves as (82). This implies that the threshold, which maximizes the capacity  $C$  in the feedback range  $B \geq \frac{P}{u^*}(\log N)^{2-\eta_1}$  satisfies (81),  $m^* = 1$  and the capacity  $C^*$  is also given by (82).

Furthermore, note that as  $B$  increases, the threshold (81) decreases. However, decreasing the threshold beyond a certain optimal value decreases the capacity upper bound in (80). The optimum value of the threshold that maximizes the upper bound in (80) is given by (33) and the corresponding upper bound is given in (32), corresponding to  $B > B_{max}$ , where  $\eta_2 \in (0, 1)$  is the solution to

$$\log N - \log [P(\log N)^{1+\eta_2}] = (\log N)^{(1+\eta_2)/2}. \quad (83)$$

Clearly,  $\eta_2 \rightarrow 1$  as  $N \rightarrow \infty$ . Substituting  $m = 1$  and (33) into the lower bound in (73) gives that the lower bound and hence the capacity also behave as in (32). Therefore, the optimal threshold that maximizes the capacity is given by (33) with adjusted  $o(1)$  term. Corresponding to (33), the maximum required feedback is given by  $B_{max} = NH(e^{-t}) = P(\log N)^{2+\eta_2} + o(\log^2 N)$ .

## REFERENCES

- [1] D. N. C. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.
- [2] Y. Sun and M. L. Honig, "Asymptotic capacity of multicarrier transmission with frequency-selective fading and limited feedback," *IEEE Trans. Inform. Theory*, vol. 54, no. 7, pp. 2879–2902, July 2008.
- [3] S. Sanayei and A. Nosratinia, "Opportunistic downlink transmission with limited feedback," *IEEE Trans. Inform. Theory*, vol. 53, no. 11, pp. 4363–4372, Nov. 2007.
- [4] D. Love and R. W. Heath Jr., "OFDM power loading using limited feedback," *IEEE Trans. Veh. Technol.*, vol. 54, no. 5, pp. 1773–1780, Sept. 2005.
- [5] Y. Rong, S. A. Vorobyov, and A. B. Gershman, "Adaptive OFDM techniques with one-bit-per-subcarrier channel-state feedback," *IEEE Trans. Wireless Commun.*, vol. 54, no. 11, Nov. 2006.
- [6] Y. Sun and M. Honig, "Minimum feedback rates for multicarrier transmission with correlated frequency-selective fading," *Proc. IEEE GLOBECOM*, vol. 3, pp. 1628–1632, Dec. 2003.
- [7] J. Chen, R. Berry, and M. L. Honig, "Performance of limited feedback schemes for downlink OFDMA with finite coherence time," in *Proc. IEEE Int. Symp. Inform. Theory*. Nice, France, June 2007.
- [8] T. Tang, R. W. Heath, S. Cho, and S. Yun, "Opportunistic feedback in clustered OFDM system," in *International Symposium on Wireless Personal Multimedia Communications*. San Diego, CA, Sept. 2006.
- [9] L. Cimini Jr., B. Daneshrad, and N. Sollenberger, "Clustered OFDM with transmitter diversity and coding," *Proc. IEEE GLOBECOM*, vol. 1, pp. 703–707, Nov 1996.
- [10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley, 2006.
- [11] T. J. Goblick, "Coding for a discrete information source with a distortion measure," Ph.D. dissertation, M.I.T., 1962.
- [12] J. T. Pinkston, "Encoding independent sample information sources," Ph.D. dissertation, M.I.T., 1967.
- [13] V. K. N. Lau, Y. J. Liu, and T.-A. Chen, "Capacity of memoryless channels and block-fading channels with designable cardinality-constrained channel state feedback," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 2038–2049, Sept. 2004.
- [14] 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, no. 1, pp. 62–70, Jan. 2004.
- [15] A. Lau and F. Kschischang, "Feedback quantization strategies for multiuser diversity systems," *IEEE Trans. Inform. Theory*, vol. 53, no. 4, pp. 1386–1400, April 2007.
- [16] E. Martinian and M. Wainwright, "Low density codes achieve the rate-distortion bound," *Proc. Data Compression Conference*, pp. 153–162, March 2006.
- [17] M. Wainwright, "Sparse graph codes for side information and binning," *IEEE Signal Processing Mag.*, vol. 24, no. 5, pp. 47–57, Sept. 2007.
- [18] G. Caire, S. Shamai, and S. Verdú, "Universal data compression with LDPC codes," in *Third International Symposium On Turbo Codes and Related Topics*. Brest, France, Sept. 2003.
- [19] T. Berger, *Rate Distortion Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1971.
- [20] R. Gray, "Information rates of autoregressive processes," *IEEE Trans. Inform. Theory*, vol. 16, no. 4, pp. 412–421, Jul 1970.
- [21] E. Ordentlich and T. Weissman, "New bounds on the entropy rate of hidden Markov processes," in *Proc. IEEE Inform. Theory Workshop*. San Antonio, TX, USA, Oct. 2004, pp. 117–122.
- [22] J. Luo and D. Guo, "On the entropy rate of hidden Markov processes observed through arbitrary memoryless channels," *IEEE Trans. Inform. Theory*, 2009, to appear.

- [23] J. Chen, R. Berry, and M. Honig, "Limited feedback schemes for downlink OFDMA based on sub-channel groups," *IEEE J. Select. Areas Commun.*, vol. 26, no. 8, pp. 1451–1461, October 2008.
- [24] Y. Sun and M. Honig, "Minimum feedback rates for multicarrier transmission with correlated frequency-selective fading," *IEEE Global Telecommunications Conference*, vol. 3, pp. 1628–1632 vol.3, 1-5 Dec. 2003.
- [25] M. Agarwal, D. Guo, and M. L. Honig, "Multi-carrier transmission with limited feedback: Power loading over sub-channel groups," in *Proc. IEEE Int. Conf. Commun.* Beijing, China, 2008, pp. 981–985.
- [26] W. Santipach and M. L. Honig, "Optimization of training and feedback for beamforming over a MIMO channel," in *Proc. IEEE Wireless Commun. and Networking Conf. (WCNC)*. Hong Kong, China, Mar. 2007.
- [27] M. Kobayashi, G. Caire, N. Jindal, and N. Ravindran, "How much training and feedback are needed in MIMO broadcast channels?" in *Proc. IEEE Int. Symp. Inform. Theory*. Toronto, Canada, July 2008.
- [28] T. Marzetta and B. Hochwald, "Fast transfer of channel state information in wireless systems," *IEEE Trans. Signal Processing*, vol. 54, no. 4, pp. 1268–1278, Apr. 2006.
- [29] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO downlink made practical: Achievable rates with simple channel state estimation and feedback schemes," *Submitted to IEEE Trans. on Info. Theory*, Nov. 2007.
- [30] X. Qin and R. A. Berry, "Distributed power allocation and scheduling for parallel channel wireless networks," *Journal Wireless Networks*, no. 5, pp. 601–613, Oct. 2008.
- [31] J. Y. Yun, S. Chung, J. Choi, Y. Jang, and Y. H. Lee, "Predictive transmit beamforming for MIMO-OFDM in time-varying channels with limited feedback," in *International Conference on Wireless Communications and Mobile Computing (IWCMC)*, 2007.
- [32] K. Huang, R. W. Heath, and J. G. Andrews, "Limited feedback beamforming over temporally correlated channel," *IEEE Transactions on Signal Processing*, 2008, submitted.
- [33] N. Khaled, B. Mondal, G. Leus, R. Heath, and F. Petre, "Interpolation-based multi-mode precoding for MIMO-OFDM systems with limited feedback," *IEEE Trans. Wireless Commun.*, vol. 6, no. 3, pp. 1003–1013, March 2007.
- [34] S. M. Hooman and G. Caire, "Channel state feedback schemes for multiuser MIMO-OFDM downlink," *Submitted to IEEE Trans. on Communications*, 2008.
- [35] R. Agarwal, V. Majjigi, Z. Han, R. Vannithamby, and J. Cioffi, "Low complexity resource allocation with opportunistic feedback over downlink OFDMA networks," *IEEE J. Select. Areas Commun.*, vol. 26, no. 8, pp. 1462–1472, Oct. 2008.
- [36] M. Agarwal and M. L. Honig, "Spectrum sharing on a wideband fading channel with limited feedback," in *Proc. CrownCom Conf.*, Orlando, Florida, Aug. 2007.
- [37] D. J. Love, R. Heath, V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE J. Select. Areas Commun.*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008.
- [38] I. Csizar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1981.
- [39] H. Robbins, "A remark of Stirling's formula," *Amer. Math. Monthly*, vol. 62, pp. 26–29, 1955.
- [40] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Annals of Mathematical Statistics*, vol. 23, pp. 493–507, 1952.