

Adaptive Modulation Based on Finite-Rate Feedback in Broadcast Channels

Lan Tang, Pengcheng Zhu, Yan Wang, and Xiaohu You, *Senior Member, IEEE*

Abstract—In a single-input single-output (SISO) downlink with multiple users, to obtain the multiuser diversity gain, the full channel state information (CSI) of all users is required for selecting the desired user and transmission mode. However, feedback channels are often capable of carrying only a limited number of bits. With such rate limited feedback links, we intend to maximize the average throughput of each user by designing transmit power levels, signal constellations and feedback thresholds jointly. Both un-coded adaptive M-ary quadrature amplitude modulation (MQAM) and adaptive trellis-coded modulation (TCM) are investigated. A nested iterative algorithm is proposed to obtain a finite number of transmission modes and thresholds. Optimization in the heterogeneous network is also considered. Simulation results demonstrate a small number of feedback bits make the throughput approach to that with perfect CSI. As the number of users increases, the multiuser diversity gain is achieved and the average feedback load of each user is reduced.

Index Terms—Broadcast channels, finite-rate feedback, adaptive modulation.

I. INTRODUCTION

THE DUALITY of broadcast channels and multiple-access channels reveals, in SISO systems, the sum capacity of broadcast channels can be achieved by a simple time division multiple access (TDMA) strategy: at each fading state, transmit to the user with the strongest channel [1]. Compared with non-faded downlink channels, a system with multiple users with independently varying channels achieves $\log K$ signal-to-noise ratio (SNR) gain (K is the number of users), which is referred to as the multiuser diversity gain. To extract such multiuser diversity benefits, the system requires that each user track its own carrier-to-noise ratio (CNR) and feed back the instantaneous channel quality to the transmitter. However, when the number of simultaneously active users is large, too much overhead information will be transmitted before useful data and a significant delay is incurred. Specifically, in a bandwidth limited system, the feedback channel is only able to convey a finite number of bits per block, which makes the multiuser diversity impractical.

In multiuser diversity systems, it is observed that the scheduled user is bound to have a “good enough” channel, which gives the opportunity to greatly reduce the number of users feeding back CSI. By designing an (or multiple) appropriate

threshold(s), only the users with CNR above the threshold need feed information back to the scheduler [2],[3]. Another candidate scheme to reduce the feedback burden is via a proper quantization. The impact of the CNR quantization on the system throughput was investigated in [4] without considering the optimal design of quantization levels. Further schemes, i.e., 1-bit feedback, were recently proposed in [5], [6]. It was shown that the ergodic sum-capacity grows as $\log \log K$ even if the feedback rate is reduced to 1-bit per user per block. The 1-bit feedback schemes in [5] and [6] only provide user selection information for the transmitter. However, in adaptive links, the transmitter also adapts the encoding mode, modulation constellation, and transmit power level according to feedback information. It motivates our research about how much feedback the throughput gain is really worth in adaptive multiuser systems.

Adaptive transmission schemes have been applied to point-to-point systems widely [7]–[11]. By assuming perfect CSI at the transmitter, the data rate and transmit power were adjusted to maximize the spectral efficiency [7], [8]. Adaptive transmission with discrete code rates and power levels was investigated when partial CSI is available at the transmitter [9], [10]. Because the transmitter cannot continuously adapt encoding rates and modulation constellations, some practical adaptation design issues were concerned in [11].

In this paper, we investigate the adaptation design in multiuser systems based on finite-rate feedback. Only users whose normalized CNRs are above the lowest boundary feed back their indexes which indicate the regions that CNRs lie in, and the other users remain silent. Subject to the predefined BER and transmit power constraints, our goal is to find an optimal set of MQAM constellations, thresholds, and power adaptation schemes in order to maximize the average spectral efficiency of broadcast channels under rate-constrained feedback links. A nested iterative approach is developed to obtain optimal thresholds and transmission modes in both uncoded modulation case and trellis-coded adaptive MQAM case. Meanwhile, a set of capacity-achieving codes, thresholds and transmit power levels are obtained by using the same way. Simulation results show that, the system suffers only a small throughput penalty with the finite feedback rate (at most one or two bits per user per block). Furthermore, the average feedback load decreases as the number of users increases.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Model

Consider a single cell broadcast channel with K single-antenna receivers and a single-antenna transmitter. It is assumed that all users experience block-Rayleigh fading inde-

Manuscript received June 24, 2007; revised October 31, 2007 and February 21, 2008; accepted April 13, 2008. The associate editor coordinating the review of this letter and approving it for publication was H. Dai. This work was supported by the Chinese 863 Program under grant 2006AA01Z268 and Huawei Tech. Co. Ltd.

The authors are with the National Mobile Communication Research Laboratory, Southeast University, Nanjing, P. R. China (e-mail: tanglan1980429@163.com, tanglan1980@yahoo.com.cn).

Digital Object Identifier 10.1109/T-WC.2008.070690

pendently. Letting r_k denote the instantaneous carrier-to-noise ratio (CNR) of the k -th user, the average CNR and normalized CNR of the k -th user are defined as $\bar{r}_k \triangleq E(r_k)$, and $x_k \triangleq r_k/\bar{r}_k$, respectively. We assume all users have the same average CNR, i.e., $\bar{r}_k = \bar{r}$, and the extension to heterogeneous networks is postponed until section V. Due to the Rayleigh fading assumption, x_k is exponentially distributed, and its probability density function (PDF) $f_{x_k}(t) = e^{-t}$, $t \geq 0$.

Each user is aware of its own fading ($\bar{r}$, x_k) only and uses a low-rate, error-free, zero-delay feedback link to convey at most B bits per block to the transmitter. To comply with the capacity constraint of the feedback channel, user k quantizes its normalized CNR x_k . The range of x_k is partitioned into $N+1$ ($N = 2^B$) consecutive regions $[0, \gamma_1)$, $[\gamma_1, \gamma_2)$, $\dots$, $[\gamma_N, \infty)$, where $\{\gamma_n\}_{n=1}^N$ are thresholds (all users employ the same partition). For notation brevity, we define $\gamma_0 = 0$ and $\gamma_{N+1} = \infty$. Similar to the selective feedback scheme presented in [2], if $x_k \geq \gamma_1$, user k feeds back an index $n \in \{1, \dots, N\}$ to the transmitter indicating that $x_k \in [\gamma_n, \gamma_{n+1})$; otherwise user k remains silent.

In each scheduling round, the transmitter randomly picks one user from the users with the largest feedback index. If no user feeds back information, the transmission is disabled. Obviously, under the proposed feedback and scheduling strategy, the fairness is guaranteed, i.e., all users have the same probability to be scheduled.

According to the feedback index of the scheduled user, the transmitter adaptively changes the transmission mode. Specifically, N transmission modes $\{(b_{k,n}, P_{k,n})\}_{n=1}^N$ are designed for user k in advance, where $b_{k,n}$ and $P_{k,n}$ are the information rate and transmit power corresponding to index n , respectively. If user k with feedback index n is scheduled, information symbols are generated according to transmission mode $(b_{k,n}, P_{k,n})$ and transmitted to user k . The receive SNR of user k , termed *post-adaptation* $\rho_{k,n}$ is then given by $P_{k,n} \bar{r} x_k$.

B. Problem formulation

In this paper, our goal is to jointly design the thresholds and transmission modes to maximize the average transmission rate of user k , subject to power and BER constraints. The average transmission rate of user k can be written as

$$\begin{aligned} R_k &= \sum_{n=1}^N b_{k,n} \Pr\{\text{user } k \text{ is scheduled} \mid x_k \in [\gamma_n, \gamma_{n+1})\} \\ &\quad \Pr\{x_k \in [\gamma_n, \gamma_{n+1})\} \\ &= \sum_{n=1}^N \frac{b_{k,n}}{K} \{(1 - e^{-\gamma_{n+1}})^K - (1 - e^{-\gamma_n})^K\} \\ &= \sum_{n=1}^N b_{k,n} \alpha_n \end{aligned} \quad (1)$$

where the second equality follows from [4], and $\alpha_n \triangleq \frac{1}{K} \{(1 - e^{-\gamma_{n+1}})^K - (1 - e^{-\gamma_n})^K\}$. Similarly, the average transmit power of user k is $\sum_{n=1}^N P_{k,n} \alpha_n$. To simplify the adaptive transceiver design, we adopt the following unifying BER approximation for all QAM constellations [13]

$$\text{BER} \approx 0.2 \exp(-g_{k,n} \rho_{k,n}) \quad (2)$$

where $g_{k,n} = 1.5/(2^{b_{k,n}} - 1)$ for square M -QAM and $g_{k,n} = 1.2/(2^{b_{k,n}} - 0.8)$ for rectangular M -QAM. As [7] and [12], we focus on data communications with "zero information outage", which means each scheduled user must satisfy the instantaneous BER requirement. Note that there is no decoding error and no outage when no transmission occurs. To maintain a target BER for each user, we define $\text{BER}_{k,n}$ as the worst BER of user k over the region n , i.e.,

$$\begin{aligned} \text{BER}_{k,n} &= \max_{x_k \in [\gamma_n, \gamma_{n+1})} \text{BER} \\ &= 0.2 \exp(-g_{k,n} P_{k,n} \bar{r} \gamma_n). \end{aligned} \quad (3)$$

Let P_{bg} and BER_{tg} stand for the average power budget and target BER of each user. Our ultimate objective is to find out optimal thresholds $\{\gamma_n\}_{n=1}^N$ and transmission modes $\{(b_{k,n}, P_{k,n})\}_{n=1}^N$ to

$$\text{maximize } R_k = \sum_{n=1}^N \alpha_n b_{k,n} \quad (4a)$$

$$\begin{aligned} \text{subject to } 1. \quad & \sum_{n=1}^N \alpha_n P_{k,n} \leq P_{\text{bg}} \\ 2. \quad & \text{BER}_{k,n} = 0.2 \exp(-g_{k,n} P_{k,n} \bar{r} \gamma_n) \leq \text{BER}_{\text{tg}}, \quad \forall n \end{aligned} \quad (4b)$$

$$\begin{aligned} 3. \quad & b_{k,n} > 0 \text{ (continuous rate)} \\ & b_{k,n} \in \{1, 2, \dots\} \text{ (discrete rate)}, \forall n. \end{aligned} \quad (4d)$$

Note that the user index k in (4a)-(4d) is a dummy variable, because the maximization problem is identical for all k . Hence, we can omit the user index k in (4a)-(4d), and rewrite the transmission modes and average transmission rate as $\{(b_n, P_n)\}_{n=1}^N$ and R . Consequently, the average throughput of the system is KR .

III. DESIGN OF TRANSMISSION MODES AND FEEDBACK THRESHOLDS IN ADAPTIVE MODULATION SYSTEMS

Equ. (4a)-(4d) provide a problem formulation whose closed-form solution is so hard to obtain that we resort to the suboptimal iterative algorithm. Firstly, the continuous-rate system design is implemented. Generally, it is difficult to vary the information rate on a continuous rate. However, the derived result provides intuition and guideline for a practical discrete rate system. Then by utilizing the results from the continuous-rate case, we carry out the discrete-rate system design.

A. the continuous-rate case

When BER_n equals BER_{tg} , a system reaches the maximum transmission rate. Thus, by inverting the BER formulation (4c), one could find an approximation for b_n

$$b_n = \log_2(1 + \frac{1.5 P_n \bar{r} \gamma_n}{-\ln(5 \text{BER}_{\text{tg}})}). \quad (5)$$

Replacing (5) in (4a), for an arbitrary power budget P_T , the problem can be reformulated as

$$\begin{aligned} \text{minimize } & - \sum_{n=1}^N \alpha_n \log_2(1 + \frac{1.5 P_n \bar{r} \gamma_n}{-\ln(5 \text{BER}_{\text{tg}})}) \\ \text{subject to } & \sum_{n=1}^N \alpha_n P_n = P_T. \end{aligned} \quad (6)$$

Let $D(P_T)$ be the minimum value of (6) with a power budget P_T . According to [10], we have Lemma 1, which is a basis of the optimization algorithm.

Lemma 1: $D(P_T)$ is monotonically decreasing and differentiable with P_T . Furthermore, $D(P_T)$ is convex.

So, we transform (6) to an unconstrained optimization problem by constructing the Lagrangian

$$J(\lambda) = - \sum_{n=1}^N \alpha_n \log_2(1 + \frac{1.5P_n \bar{\gamma} \gamma_n}{-\ln(5\text{BER}_{\text{ig}})}) + \lambda \sum_{n=1}^N \alpha_n P_n. \quad (7)$$

By minimizing $J(\lambda)$, we can find an optimal operation point $(D(P_T(\lambda)), P_T(\lambda))$ on the D- P_T plane, such that $\partial D(P_T)/\partial P_T|_{P_T=P_T(\lambda)} = -\lambda$. Due to the power constraint in (4b), we need to find optimal λ_0 to make $\partial D(P_T)/\partial P_T|_{P_T=P_{\text{bg}}} = -\lambda_0$. As in [10], the bisection search algorithm is adopted to locate the optimal λ_0 . In the following, we firstly introduce the sub-algorithm to minimize (7) for a given λ . Based on the sub-algorithm, a nested iterative algorithm is then presented.

Letting $A = \frac{1.5\bar{\gamma}}{-\ln(5\text{BER}_{\text{ig}})}$, the power control that minimizes (7) follows the water-filling principle

$$\frac{\partial J}{\partial P_n} = 0 \Rightarrow P_n = (\frac{1}{\lambda \ln 2} - \frac{1}{A\gamma_n})^+. \quad (8)$$

Then, the optimal region boundaries can be obtained by solving the following equalities

$$\frac{\partial J}{\partial \gamma_n} = 0, \quad n = 1, \dots, N-1. \quad (9)$$

For a given λ , we aim at seeking $\{\gamma_n, P_n\}_{n=1}^N$ to make $\frac{\partial J}{\partial \gamma_n} = 0$. By initializing γ_1 , one can successively compute $\gamma_2, \dots, \gamma_N, P_1, \dots, P_N$ using (8) and (9). Thus, $\frac{\partial J}{\partial \gamma_n} = 0$ is a nonlinear equation with respect to γ_1 . We observe that for many given λ , $\frac{\partial J}{\partial \gamma_n} = 0$ appears to have a unique solution. The detailed procedures for the sub-algorithm (named SF-C) and the nested iterative algorithm are summarized in Table 1-algorithm 1.

B. The discrete-rate case

For clarity, the transmission rate in the discrete-rate case is defined as $\tilde{b}_n$, $n = 1, \dots, N$, which is restricted to the set $\mathcal{R} = \{1, 2, 3, \dots\}$. Utilizing the bit loading result $\{b_n\}_{n=1}^N$ from continuous rate design, we determine the $\{\tilde{b}_n\}_{n=1}^N$ of discrete rate design by rounding b_n , $n = 1, \dots, N$ to the nearest integer in set $\mathcal{R}$. Therefore, there are M ($M < 2^N$) possible bit loading schemes (in practice, the number of reasonable bit loading scheme is far less than 2^N . For example, $\tilde{b}_n$ should be no smaller than $\tilde{b}_{n-1}$). Let the m th bit loading scheme and corresponding thresholds, transmit powers and average transmission rate be $\{\tilde{b}_n^{(m)}\}_{n=1}^N$, $\{\gamma_n^{(m)}, P_n^{(m)}\}_{n=1}^N$ and $R^{(m)}$, $m = 1, \dots, M$. In the following, we introduce the algorithm to maximize $R^{(m)}$ given the m th bit loading scheme $\{\tilde{b}_n^{(m)}\}_{n=1}^N$ and power budget.

To satisfy the BER requirement, we propose the power control policy by inverting (4c)

$$P_n^{(m)} = \begin{cases} (2^{\tilde{b}_n^{(m)}} - 1)/A\gamma_n^{(m)} & \tilde{b}_n^{(m)} \text{ is an even number} \\ (2^{\tilde{b}_n^{(m)}} - 0.8)/B\gamma_n^{(m)} & \tilde{b}_n^{(m)} \text{ is an odd number} \end{cases} \quad (10)$$

TABLE I
ALGORITHMS FOR CONTINUOUS RATE CASE AND DISCRETE RATE CASE

Algorithm 1: Sub-algorithm to minimize (7) and nested iterative algorithm

Sub-function in continuous rate case (SF-C)

1. Choose $\gamma_{1,L} > \frac{\lambda \ln 2}{A}$ to make $\frac{\partial J}{\partial \gamma_N} > 0$;
choose a proper $\gamma_{1,H}$ to make $\frac{\partial J}{\partial \gamma_N} < 0$;
2. Determine $\{P_n\}_{n=1}^N$ and $\{\gamma_n\}_{n=2}^N$ successively according to (8) and (9) with $\gamma_1 \leftarrow \frac{\gamma_{1,L} + \gamma_{1,H}}{2}$;
3. Replace $\gamma_{N+1} = +\infty$ in $\frac{\partial J}{\partial \gamma_N}$ to check the equality:

IF $|\frac{\partial J}{\partial \gamma_N}| \leq \varepsilon \rightarrow 0$, **THEN** exit;

ELSE IF $\frac{\partial J}{\partial \gamma_N} > 0$, **THEN** $\gamma_{1,L} \leftarrow \gamma_1$ and go to 2;

ELSE $\gamma_{1,H} \leftarrow \gamma_1$ and go to 2.

Main-function(bisection search algorithm)

1. Initialization. Choose a proper λ_{l_0} and apply SF-C to obtain $P_T(\lambda_{l_0})$ and $D(P_T(\lambda_{l_0}))$, make sure $P_T(\lambda_{l_0}) < P_{\text{bg}}$; choose another λ_{hi} to obtain $P_T(\lambda_{hi})$ and $D(P_T(\lambda_{hi}))$, make sure $P_T(\lambda_{hi}) > P_{\text{bg}}$;
2. Update the lagrange multiplier using: $\lambda_n = \frac{D(P_T(\lambda_{hi})) - D(P_T(\lambda_{l_0}))}{P_T(\lambda_{hi}) - P_T(\lambda_{l_0})}$;
3. Apply the SF-C with λ_n to obtain $P_T(\lambda_n)$ and $D(P_T(\lambda_n))$;
4. **IF** $|P_T(\lambda_n) - P_{\text{bg}}| < \varepsilon \rightarrow 0$, **THEN** exit;
ELSE IF $P_T(\lambda_n) > P_{\text{bg}}$, **THEN** $P_T(\lambda_{hi}) \leftarrow P_T(\lambda_n)$,
 $D(P_T(\lambda_{hi})) \leftarrow D(P_T(\lambda_n))$ and go back to 2;
ELSE $P_T(\lambda_{l_0}) \leftarrow P_T(\lambda_n)$, $D(P_T(\lambda_{l_0})) \leftarrow D(P_T(\lambda_n))$
and go back to 2.

Algorithm 2: Nested iterative algorithm for discrete rate case

1. Initialization. $z \leftarrow 0$, $D_{\min} \leftarrow +\infty$, $\gamma_n^* \leftarrow 0$, $P_n^* \leftarrow 0$,
 $n = 1, \dots, N$;
2. $z \leftarrow z + 1$, Choose the z th bit loading scheme;
3. For a given λ , apply SF-D to find $\{\gamma_n\}_{n=1}^N$ and calculate
corresponding $\{P_n\}_{n=1}^N$, $P_T(\lambda)$ and $D(P_T(\lambda))$;
Apply bisection search algorithm to update λ ;
4. **IF** $|P_T(\lambda) - P_{\text{bg}}| < \varepsilon \rightarrow 0$, **THEN** exit; **ELSE** go to 3.
5. **IF** $D(P_{\text{bg}}) < D_{\min}$, **THEN** $\gamma_n^* \leftarrow \gamma_n$, $P_n^* \leftarrow P_n$,
 $D_{\min} \leftarrow D(P_{\text{bg}})$, $n = 1, \dots, N$;
6. **IF** z is the number of available integer-bit loading schemes,
THEN exit; **ELSE** go to 2.

where $A = \frac{1.5\bar{\gamma}}{-\ln(5\text{BER}_{\text{ig}})}$, $B = \frac{1.2\bar{\gamma}}{-\ln(5\text{BER}_{\text{ig}})}$. Thus, for a given bit loading scheme $\{\tilde{b}_n^{(m)}\}_{n=1}^N$ and power budget P_T , the optimization problem can be expressed as

$$\begin{aligned} & \text{minimize} - \sum_{n=1}^N \alpha_n \tilde{b}_n^{(m)} \\ & \text{subject to} \sum_{n=1}^N \alpha_n (2^{\tilde{b}_n^{(m)}} - c_n)/d_n \gamma_n^{(m)} = P_T \end{aligned} \quad (11)$$

where $c_n = 1$, $d_n = A$, if $\tilde{b}_n$ is an even number, else $c_n = 0.8$, $d_n = B$. (11) can be turned to an unconstrained optimization problem by constructing the Lagrangian

$$J(\lambda) = - \sum_{n=1}^N \alpha_n \tilde{b}_n^{(m)} + \lambda \sum_{n=1}^N \alpha_n (2^{\tilde{b}_n^{(m)}} - c_n)/d_n \gamma_n^{(m)}. \quad (12)$$

Let $D(P_T)$ be the minimum value of (11) with a power budget P_T . Note that for any bit loading scheme, the same result about $D(P_T)$ as that in Lemma 1 exists. This makes the bisection search algorithm applicable, as in the continuous-rate case.

For a given λ , differentiating (12) with respect to $\gamma_n^{(m)}$, one obtains

$$\frac{\partial J}{\partial \gamma_n^{(m)}} = 0, \quad n = 1, \dots, N-1. \quad (13)$$

As discussed below (9), $\frac{\partial J}{\partial \gamma_N^{(m)}} = 0$ can be regarded as a nonlinear equation with respect to $\gamma_1^{(m)}$. So an iterative algorithm (named SF-D) similar to SF-C in algorithm 1 is used to find the optimal thresholds. We skip the detailed description for brevity. The bisection search algorithm is adopted to locate the optimal λ_0 to satisfy $P_T(\lambda_0) = P_{bg}$.

The algorithm described above has to be implemented M times to get M possible solutions $\{\gamma_n^{(m)}, \tilde{\gamma}_n^{(m)}, P_n^{(m)}\}_{n=1}^N$ s and transmission rates $R^{(m)}$ s ($m = 1, \dots, M$). The index of the optimal solution equals $\arg\max_m R^{(m)}$. In summary, the optimization algorithm for discrete rate case is illustrated in Table. 1-algorithm 2.

Finally, we stress that the entire design procedure will be performed offline. Given the number of users and the operating SNR, the optimal thresholds and transmission modes can be obtained in advance so that the optimization is not required in real time.

IV. DESIGN OF TRANSMISSION MODES AND FEEDBACK THRESHOLDS IN CODED MODULATION SYSTEMS

A. TCM

In this subsection, we consider $2L$ -dimension ($2L$ -D) TCM where L is some positive integer [14]. For $L = 1$, the coding scheme reduces to 2-D trellis codes in [8]. As in [8], [13], we focus on adaptive TCM, where a fixed trellis code is superimposed on uncoded adaptive modulation for fading channels. At the transmitter, all users employ trellis encoders with the same structure so that the same Viterbi decoder can be used at different receivers.

Each fading region is assigned one $2L$ -D trellis code. When $x_k \in [\gamma_n, \gamma_{n+1})$, the signal constellation for code n is given by the L -fold Cartesian product of a 2-D signal set with size 2^{b_n} . Each time $Lb_n - 1$ bits pass through encoder to generate Lb_n bits. The coded bits are then mapped onto L 2-D 2^{b_n} -QAM symbols. Thus, the average information rate of user k is

$$R_k = \sum_{n=1}^N (b_n - 1/L) \alpha_n \quad (14)$$

We adopt the approximate BER for $2L$ -D trellis-coded MQAM as [14]

$$\text{BER} \approx c_n \exp(-\beta_n g_n \rho_n) \quad (15)$$

where c_n and β_n depend on the chosen trellis code and the signal constellation. c_n and β_n in 4-D trellis code can be found in [14]. Similar to the formulation in (4a)-(4d), our ultimate objective is to find optimal thresholds $\{\gamma_n\}_{n=1}^N$ and transmission modes $\{(b_n, P_n)\}_{n=1}^N$ to maximize R_k , subject to 1. $\sum_{n=1}^N \alpha_n P_n \leq P_{bg}$; 2. $\text{BER}_n = c_n \exp(-\beta_n g_n \bar{r} \gamma_n P_n) \leq \text{BER}_{ig}$; 3. $b_n \in \{1, 2, 3, \dots\}$. It is straightforward to apply the optimization algorithm in section III to solve this problem.

Particularly, in section VI we consider a 2-D eight-state TCM using a core rate $2/3$ encoder. Out of $b_n - 1$ information bits, 2 bits enter the encoder to produce 3 coded bits to choose a subset from the partition of the signal constellation. The remaining $b_n - 3$ bits are used to specify one 2^{b_n} -QAM signal from the selected subset. To guarantee that each subset

contains at least one point, the constellations with size ≥ 8 are chosen. For such a TCM, $c_n = 6$ and $\beta_n = 5$ for all constellations [13]. To satisfy the BER constraint, we assume all dominating errors events are confined within one feedback interval.

B. Capacity-achieving codes

In subsection A, the optimization of thresholds and transmission modes is under BER constraints for real constellations. In this subsection, we follow the "zero information outage" philosophy [7], [12] and find a set of capacity-achieving rates $\{C_n\}_{n=1}^N$, thresholds $\{\gamma_n\}_{n=1}^N$, and power levels $\{P_n\}_{n=1}^N$ to maximum the average reliable throughput of each user. Using the result of [9], data should only be buffered when $\max x_k < \gamma_1$. In the other normalized CNR regions, the assigned rate supports the worst channel state of that region. The problem is formulated as

$$\begin{aligned} \text{maximize } C &= \sum_{n=1}^N \alpha_n C_n = \sum_{n=1}^N \alpha_n \log_2 (1 + P_n \bar{r} \gamma_n) \\ \text{subject to } \sum_{n=1}^N \alpha_n P_n &\leq P_{bg}. \end{aligned} \quad (16)$$

By applying the same optimization algorithm as that in A, we can get a set of optimal $\{C_n, \gamma_n, P_n\}_{n=1}^N$.

V. EXTENSION TO THE HETEROGENEOUS NETWORK

Obviously, the scheduling algorithm proposed in section II can guarantee the fairness in the heterogeneous network with different average CNRs. Assume that all users deploy same feedback thresholds $\{\tilde{\gamma}_n\}_{n=1}^N$. Forward power control is used to guarantee that the average receive SNR of each user is equal, i.e., $P_k^{bg} = \bar{r} P_{bg} / \bar{r}_k$, where P_k^{bg} denote the power budget of user k . The optimization problem can be expressed using (4a)-(4d) with the substitution $\bar{r} \rightarrow \bar{r}_k$, $P_{bg} \rightarrow P_k^{bg}$. It is observed that the optimization results from the homogeneous network can be readily applied to the heterogeneous network with $\tilde{\gamma}_n = \gamma_n$, $b_{k,n} = b_n$, $P_{k,n} = \bar{r} P_n / \bar{r}_k$.

VI. NUMERICAL RESULTS

In this section, we present numerical results for multiuser systems with finite-rate feedback. Due to the similarity of the design between in the homogeneous network and heterogeneous network, we only give the results in the homogeneous network. We set $\bar{r} = 1$. The average throughput of the system is KR . QAM constellations with size 2^i , $i \in [1, 8]$ are available. The thresholds and transmission modes for both continuous rate case and discrete rate case are designed by Algorithm 1 and 2. For example, the optimal values of the 3-tuple (threshold, constellation, transmit power) in the system with $P_{bg} = 10\text{dB}$, $\text{BER} = 10^{-3}$, $K = 10$ are

- $B = 1$ case: (1.54, 4QAM, 8.38dB), (3.11, 16QAM, 13.21dB)
- $B = 2$ case: (1.27, 4QAM, 9.2dB), (1.95, 4QAM, 7.34dB), (2.98, 16QAM, 12.5dB), (3.83, 16QAM, 11.4dB).

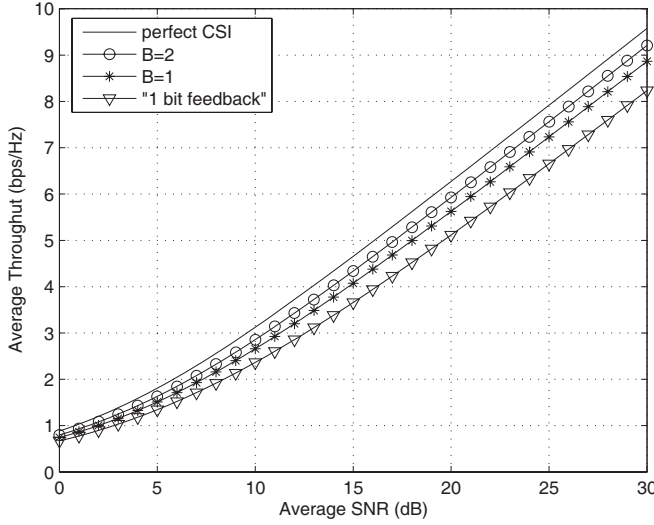


Fig. 1. Achievable data throughput vs. average SNR (continuous-rate case, $K = 10$, $\text{BER} = 10^{-3}$)

Experiment 1-Achievable data throughput versus average receive SNR. Fig.1 shows the achievable data throughput for continuous-rate modulation systems with different feedback bits. The throughput upper bound is also plotted for reference, which is achieved by assuming that the transmitter knows perfect CSI of each user. Then, the user with best channel is scheduled and the transmission parameters are optimized based on the instantaneous CNR. The optimal adaptive modulation solution for such case is available in [9]. We would point that the feedback with $B=1$ is different from “1 bit feedback” proposed in [7] since the latter only designs one threshold and one transmission mode. Compared with the case with perfect CSI, “1 bit feedback” scheme causes about 4dB SNR loss. Such a loss is reduced to 2dB and 1dB when $B=1$ and $B=2$. It suggests that even increasing more feedback bits, the improvement will be marginal. A tradeoff between the throughput and the number of feedback bits therefore emerges. In practice, a small number of feedback bits, such as one or two bits per block might be favorable.

Fig.2 displays the transmission rate achieved by the discrete rate design. Similar to its continuous rate counterpart, the throughput with $B=2$ comes within approximately 1dB of the throughput upper bound. However, different from the continuous-rate case, the bit loading over each region could remain invariant with the increase of SNR, and the throughput is improved by adjusting thresholds. The average throughput exhibits fluctuating curves due to the alteration of bit loading scheme. This phenomenon is obvious in “1 bit feedback” scheme. We compare the average throughput for trellis-coded adaptive MQAM with that for uncoded adaptive modulation in Fig. 3. It is observed that an 8-state code yields about 3dB coding gain over uncoded modulation when $B=2$ and $B=1$. The maximum coding gain can reach 4dB in “1 bit feedback” scheme.

Experiment 2-Achievable data throughput versus number of users. It is well known that the sum-rate capacity of a SISO downlink with K users grows as $\log \log K$ when “the best user” is scheduled at each transmission. In multiuser systems

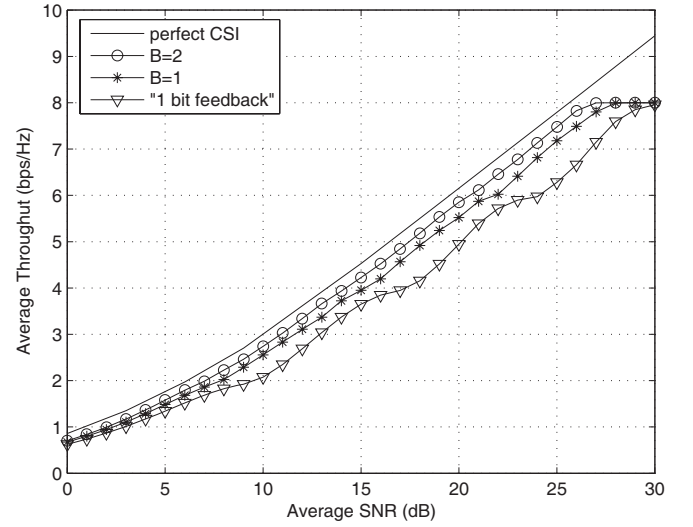


Fig. 2. Achievable data throughput vs. average SNR (discrete-rate case, $K = 10$, $\text{BER} = 10^{-3}$)

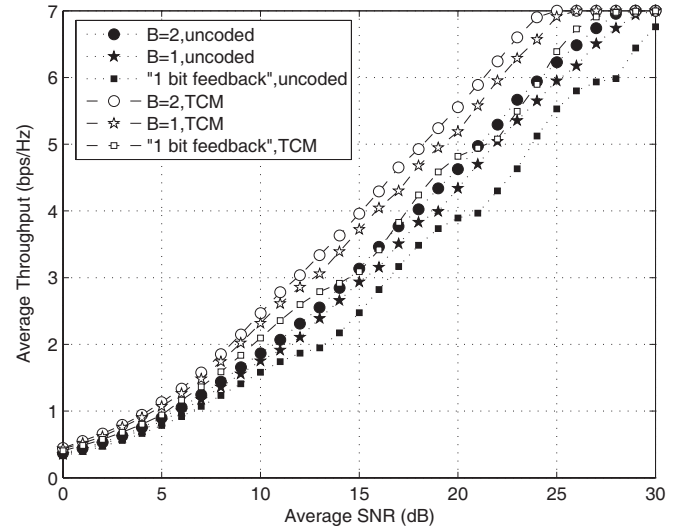


Fig. 3. Throughput improvement with trellis coded modulation (discrete rate case, $K = 10$, $\text{BER} = 10^{-6}$)

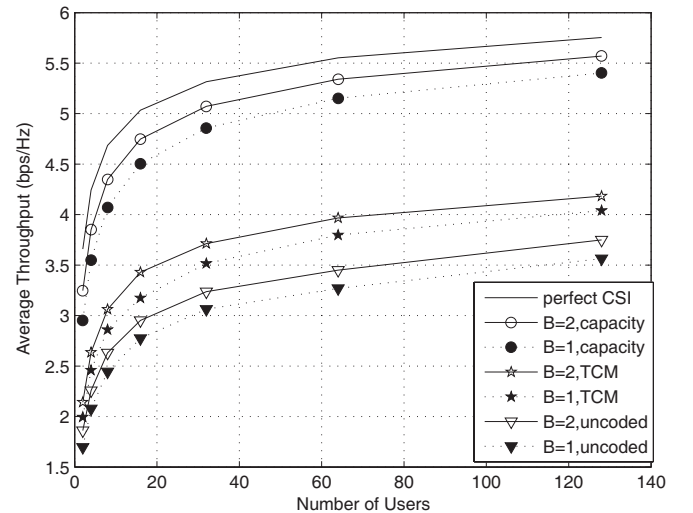


Fig. 4. Achievable data throughput vs. number of users ($K P_{bg} = 10\text{dB}$, $\text{BER} = 10^{-3}$)

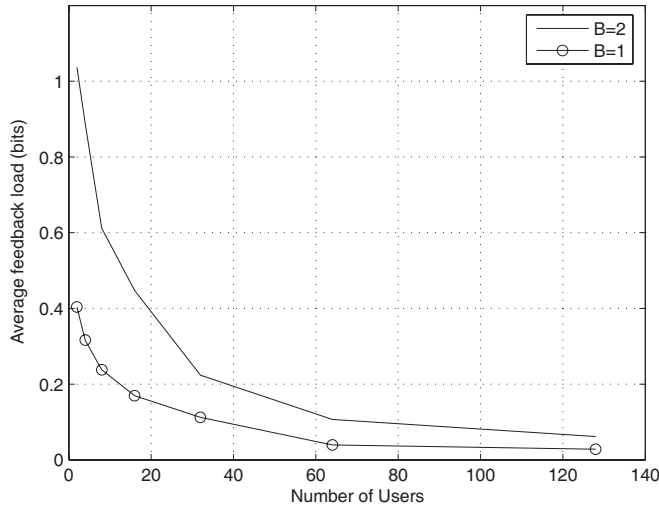


Fig. 5. Average feedback load of each user

with finite-rate feedback, it is not necessary that "the best user" is chosen in each scheduling round, which reduces the multiuser diversity gain. However, we see from the curve with circle in Fig.4, the system with $B=2$ obtains approximately $\log K$ multiuser diversity gain. The gap between the capacity and the rate achieved by uncoded modulation (or trellis-coded modulation) can shrink further by using more powerful code. Fig. 5 shows the average feedback load of each user, which can be calculated as $B \exp(-\gamma_1)$. With the increase of the number of users, γ_1 increases too, which leads to the reduction of average feedback load. As a result, the total amount of feedback bits doesn't exceed 8 bits in the system with $K = 128$, $B = 2$. It suggests only the users with "good channel" will feedback information and most of users keep silent.

VII. CONCLUSION

In this paper, we proposed a design scheme for feedback thresholds and transmission modes in adaptive multiuser systems under rate-constrained feedback links. Subject to the average power and BER constraints, we intended to maximize the average throughput of each user. Both continuous-rate modulation and discrete-rate modulation were discussed. A

nested iterative algorithm was proposed to yield a set of optimal thresholds and transmission modes. This optimization scheme can be applied to trellis-coded modulation systems to improve the throughput further. Due to its inherent fairness, it also can be readily applied to the heterogeneous network. Numerical results demonstrated that the finite-rate feedback reduces the feedback load considerably, but only causes a slight loss to both multiuser diversity gain and throughput.

REFERENCES

- [1] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. IEEE International Conference on Communication 1995 (ICC'95)*, vol. 1, pp. 331-335.
- [2] D. Gesbert and M.-S. Alouini, "How much feedback is multi-user diversity really worth?" in *Proc. IEEE Int. Conf. Commun.*, Mar. 2004, pp. 234-238.
- [3] V. Hassel, D. Gesbert, M.-S. Alouini, and G. E. Oien, "A threshold-based channel state feedback algorithm for modern cellular systems," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2422-2426, July 2007.
- [4] F. Floren, O. Edfors, and B.-A. Molin, "The effect of feedback quantization on the throughput of a multiuser diversity scheme," in *Proc. IEEE Global Commun. Conf.*, Dec. 2003, vol. 1, pp. 497-501.
- [5] O. Somekh, A. M. Haimovich, and Y. Bar-Ness, "Sum rate analysis of downlink channels with 1-bit feedback," *IEEE Commun. Lett.*, vol. 11, no. 2, pp. 137-139, Feb. 2007.
- [6] S. Sanayei and A. Nosratinia, "Exploiting multiuser diversity with only 1-bit feedback," in *Proc. IEEE Wireless Commun. and Network. Conf.*, May 2005, pp. 978-983.
- [7] A. J. Goldsmith and S. G. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, no. 10, pp. 1218-1230, Oct. 1997.
- [8] A. J. Goldsmith and S. G. Chua, "Adaptive coded modulation for fading channels," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 595-602, May 1998.
- [9] L. Lin, R. D. Yates, and P. Spasojevic, "Adaptive transmission with discrete code rates and power levels," *IEEE Trans. Commun.*, vol. 51, pp. 2115-2125, Dec. 2003.
- [10] P. Xia, S. Zhou, and G. B. Giannakis, "Multiantenna adaptive modulation with beamforming based on bandwidth constrained feedback," *IEEE Trans. Commun.*, vol. 53, no. 3, pp. 526-536, Mar. 2005.
- [11] V. Lau and T. Wu, "On the encoding rate and discrete modulation adaptive design for MIMO links," *IEEE Trans. Wireless Commun.*, vol. 6, no. 1, Jan. 2007.
- [12] J. F. Paris, M. del Carmen A.-Torres, and J. T. Entrambasaguas, "Optimum discrete-power adaptive QAM scheme for Rayleigh fading channels," *IEEE Commun. Lett.*, vol. 5, pp. 281-283, July 2001.
- [13] S. Zhou and G. B. Giannakis, "Adaptive modulation for multiantenna transmissions with channel mean feedback," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1626-1636, Sept. 2004.
- [14] K. J. Hole, and H. Holm, and G. E. Oien, "Adaptive multidimensional coded modulation over flat fading channels," *IEEE J. Select. Areas Commun.*, vol. 18, no. 7, pp. 1153-1158, July 2000.