Interference Channel with Constrained Partial Group Decoding

Chen Gong, Ali Tajer, *Member, IEEE,* and Xiaodong Wang, *Fellow, IEEE*

***Abstract*—We propose novel coding and decoding methods for a fully connected K-user Gaussian interference channel. Each transmitter encodes its information into multiple layers and transmits the superposition of those layers. Each receiver employs a constrained partial group decoder (CPGD) that decodes its designated message along with a part of the interference. In particular, each receiver performs a twofold task by first identifying which interferers it should decode and then determining which layers of them should be decoded. Determining the layers to be decoded and decoding them are carried out in a successive manner, where in each step a group of layers with a constraint on its group size is identified and jointly decoded while the remaining layers are treated as Gaussian noise. The decoded layers are then subtracted from the received signal and the same procedure is repeated for the remaining layers. We provide a *distributed* algorithm, tailored to the nature of the interference channels, that determines the transmission rate at each transmitter based on some optimality measure and also finds the order of the layers to be successively decoded at each receiver. We also consider practical design of a system that employs the quadrature amplitude modulations (QAM) and rateless codes. Numerical results are provided on the achievable sum-rate under the ideal case of Gaussian signaling with random codes as well as on the system throughput under practical modulations and channel codes. The results show that the proposed multi-layer coding scheme with CPGD offers significant performance gain over the traditional un-layered transmission with single-user decoding.**

***Index Terms*—Interference channel, layered coding, constrained partial group decoding, rate allocation, distributed algorithm, rateless code.**

I. Introduction

INTERFERENCE channel is a fundamental building block of the wireless networks. Due to the ever-shrinking network sizes and the increasing demands for achieving higher spectral efficiency, the emerging wireless networks operate in an interference-limited regime. Motivated by such demands, investigating different aspects of interference channel has resulted in various recent developments for further understanding the fundamental limits of these channels [1], [2].

In the rich literature of interference channel it is well-understood that while a receiver is not ultimately interested in decoding the messages of the interferers, decoding them (fully or partially) is often advantageous for recovering its desired message [3]. Motivated by this premise some recent developments for the K-user interference channels propose that each receiver should partition the interfering transmitters into two groups; one group to be *fully* decoded along with the designated transmitter and the other to be treated as Gaussian noise [4]–[6].

One major advantage of *fully* decoding a transmitter is that it suffices to assign only one codebook to that transmitter, which is appealing for practical purposes. One drawback of such decoding, on the other hand, is that the receivers will give up the freedom of decoding only a *fraction* of an interferer, which in some instances can be more beneficial than fully decoding it. In this paper we consider allocating more than one codebook to each transmitter and provide the receivers with the freedom of deciding about what interferers to decode as well as determining what fraction of such interferers need to be decoded.

Assigning multiple codebooks enables splitting the message of each transmitter to multiple layers, each drawn from one separate codebook. Such rate splitting can be seen as the generalization of the Han-Kobayashi scheme [3] for the 2-user interference channel, that splits the message of each transmitter into two layers. Splitting the messages into multiple layers provides the receiver with the freedom of *partially* decoding the interferers and consequently with the advantage of sustaining reliable communication at higher rates. Attaining such gains is at the cost of facing certain practical complexities and obstacles. In particular when each transmitter has multiple codebooks, each receiver has to identify the best subset of the codebooks to be decoded, which can be computationally prohibitive. Moreover, the interference channel is a distributed system which might allow only very limited coordination among different users, while multi-layer transmission schemes and decoding the interference by receivers necessitates an interplay among the design of the different transmitters and receivers. Finally, as the transmitters need to be decodable at multiple receivers designing the channel codes becomes considerably more complicated.

In this paper we first introduce and discuss the notion of constrained partial group decoder (CPGD) and then address the challenges pertinent to its implementation. We discuss these challenges through solving a fair rate allocation problem for the K-user interference channel. Specifically, in the first step we propose the so-called CPGD, and show that it can solve the rate allocation problem in a computationally efficient way and with limited coordination among different users. In the second step we focus on designing a practical coding scheme that can be used in conjunction with the proposed CPGD structure.

Paper approved by S.-Y. Chung, the Editor for LDPC Coding and Information Theory of the IEEE Communications Society. Manuscript received October 12, 2010; revised April 21, 2011.

C. Gong and X. Wang are with the Department of Electrical Engineering, Columbia University, New York, NY 10027 USA (e-mail: {chengong, wangx}@ee.columbia.edu).

A. Tajer is with the Department of Electrical Engineering, Princeton University, B-311 E-Quad, Olden Street, Princeton, NJ 08544 USA (e-mail: tajer@princeton.edu).

Digital Object Identifier 10.1109/TCOMM.2011.082111.100624

The remainder of the paper is organized as follows. In Section II, we describe the layered transmission model of the K-user interference channel and the CPGD. In Section III, we formulate the rate allocation for CPGD and provide the optimal solution. In Section IV, we develop a practical implementation of CPGD. In Section V, we present the rateless code profile design for CPGD, and show the system throughput performance using the optimized codes. Finally, Section VI provides the concluding remarks.

II. System Descriptions

A. Channel Model

Consider a fully connected K-user interference channel. We denote the wireless channel from the j^{th} transmitter to the i^{th} receiver by $h_{i,j}$ for $i, j \in \{1, \ldots, K\}$. We assume quasi-static block fading channels such that the channel coefficients are fixed during the transmission of N symbols and change to some other independent states afterwards. By defining $x_j[n]$ as the transmitted signal by the j^{th} transmitter for $n \in \{1, \ldots, N\}$ the received signal by the i^{th} receiver is given by

$$y_i[n] = \sum_{j=1}^{K} h_{i,j}\, x_j[n] + v_i[n]\ , \quad \text{for}\ \ i \in \{1, \ldots, K\}\ , \quad (1)$$

where $v_i[n] \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2)$ accounts for the additive white Gaussian noise (AWGN). The term $h_{i,i}\, x_i[n]$ contains the intended signal for the i^{th} receiver and the remaining summands constitute interference and noise. Also denote the transmission power of all transmitters as $P \triangleq \mathbb{E}(|x_j[n]|^2)$.

B. Layered Encoding

For allowing the receivers to decode the transmitters *partially*, the message of each transmitter is split into smaller layers each drawn from an independent codebook (rate splitting). Let us denote the number of codebooks (layers) that we allocate to transmitter j by L_j. Also denote the set of codebooks of transmitter j by $\bar{\mathcal{C}}_j \triangleq \{\mathcal{C}_{j,1}, \ldots, \mathcal{C}_{j,L_j}\}$ for $j = 1, \ldots, K$. By denoting the codeword drawn from codebook $\mathcal{C}_{j,k}$ by $x_{j,k}[n]$ from (1) we obtain

$$y_i[n] = \sum_{j=1}^{K} h_{i,j} \sum_{k=1}^{L_j} x_{j,k}[n] + v_i[n]\ , \quad \text{for}\ \ i \in \{1, \ldots, K\}\ . \quad (2)$$

We adopt equal power allocation for all the layers $x_{j,k}$ at the transmitter j. By defining $R_{j,k}$ and R_j as rates of codebook $\mathcal{C}_{j,k}$ and transmitter j, respectively, we have $R_j = \sum_{k=1}^{L_j} R_{j,k}$. We use (j,k) to denote the index of codebook $\mathcal{C}_{j,k}$ for $1 \leq j \leq K$ and $1 \leq k \leq L_j$, and define $\mathcal{K}$ as the set of all such indices. For any set $\mathcal{A} \subseteq \mathcal{K}$ we define the rate vector $\boldsymbol{R}_{\mathcal{A}} \triangleq [R_{j,k}]_{(j,k)\in\mathcal{A}}$. Throughout the paper all the rates are in bits/sec/Hz and all the logarithms are in base 2

C. Constrained Partial Group Decoding (CPGD)

Motivated by the premise that decoding the interference (fully or partially) along with the desired signal is *sometimes* beneficial, we introduce the notion of constrained partial group decoding. This principle has been the basis of many developments in the rich literature on the interference channel. In particular, the Han-Kobayashi scheme for the 2-user interference channel uses the rate splitting technique in order to allow each receiver to decode a part of the message of the interfering transmitter. In this paper we consider the general K-user interference channel for any arbitrary $K \geq 2$.

Upon determining what codebooks to be decoded at each receiver, which hinges on the utility function that we seek to optimize for the network, each receiver employs a successive decoding procedure. In each stage a subset of the layers are jointly decoded via the maximum likelihood (ML) decoding, after subtracting the already decoded layers from the received signal, and by treating the remaining layers as AWGN. In order to control the complexity of ML decoding we constrain the number of layers being jointly decoded at each stage to be at most μ.

We say that a given *ordered* partition $\underline{\mathcal{Q}}^i \triangleq \{\mathcal{Q}_1^i, \ldots, \mathcal{Q}_{p_i}^i, \mathcal{Q}_{p_i+1}^i\}$ of $\mathcal{K}$ (the set of the indices of all codebooks) is *valid* if all the following conditions are satisfied.

1) $|\mathcal{Q}_m^i| \leq \mu$ for $m \in \{1, \ldots, p_i\}$;
2) All layers of transmitter i, i.e., $\{x_{i,k}\}_{k=1}^{L_i}$, are included in $\{\mathcal{Q}_1^i, \ldots, \mathcal{Q}_{p_i}^i\}$;
3) The rate vector $\boldsymbol{R}_{\mathcal{Q}_m^i}$ is decodable at the m^{th} stage of the successive decoding procedure for $m \in \{1, \ldots, p_i\}$.

For a given valid partition $\underline{\mathcal{Q}}^i$ of $\mathcal{K}$, the i^{th} receiver decodes the layers included in $\{\mathcal{Q}_1^i, \ldots, \mathcal{Q}_{p_i}^i\}$ successively in p_i stages while those in $\mathcal{Q}_{p_i+1}^i$ are always treated as AWGN. More specifically, in the m^{th} stage, the i^{th} receiver jointly decodes the layers in $\mathcal{Q}_m^i$ by treating $\{\mathcal{Q}_{m+1}^i, \ldots, \mathcal{Q}_{p_i+1}^i\}$ as AWGN and then subtracts the decoded messages in $\mathcal{Q}_m^i$ from the received signal. The steps involved in the successive decoding procedure for a given valid partition $\{\mathcal{Q}_1^i, \ldots, \mathcal{Q}_{p_i}^i, \mathcal{Q}_{p_i+1}^i\}$ are as follows.

1) Initialize $m = 1$.
2) Compute

$$\Sigma_m^i = \sigma^2 + \sum_{n=m+1}^{p_{i+1}} \sum_{(j,k)\in\mathcal{Q}_n^i} \frac{|h_{i,j}|^2 P}{L_j}\ , \quad (3)$$

and jointly decode the layers in $\mathcal{Q}_m^i$ via ML decoding assuming the AWGN variance is Σ_m^i.

3) Update $y_i[n] \leftarrow y_i[n] - \sum_{j\in\mathcal{Q}_m^i} h_{i,j}\hat{x}_j$ where $\hat{x}_j$ is the decision made on x_j for $j \in \mathcal{Q}_m^i$ and update $m \leftarrow m+1$.
4) If $m = p_{i+1}$ stop and otherwise go to step 2.

Determining the optimal valid partition of $\mathcal{K}$ and the corresponding rate vectors supported by such partition is the task of the CPGD and is the subject of the next section. Whenever we are not using Gaussian codebooks, decodable pertains to the specific channel encoders and decoders deployed.

1) A Simple Example of CPGD: In Fig. 1, we show a simple illustrative example of the proposed CPGD for an interference channel with four users, where users 1 and 3 contain 2 layers and users 2 and 4 contain one layer. The set of all layers is given by $\{(1,1), (1,2), (2,1), (3,1), (3,2), (4,1)\}$. We show the CPGD at the receiver of user 3, which aims

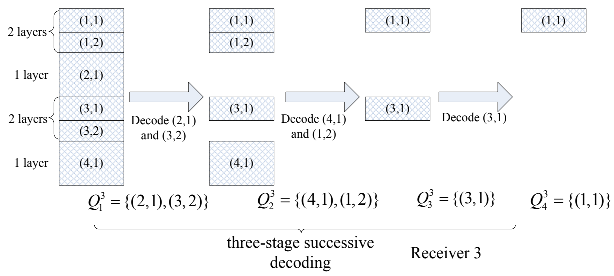


Fig. 1. An simple illustrative example for constrained partial group decoding.

to decode layers $(3,1)$ and $(3,2)$. Shown in Fig. 1, layers $(2,1)$ and $(3,2)$ are decoded in the first stage; layers $(4,1)$ and $(1,2)$ are decoded in the second stage; and layer $(3,1)$ is decoded in the third stage. After the three stages, layers $(3,1)$ and $(3,2)$ are decoded and the decoding terminates. We have $p_3 = 4$, $\mathcal{Q}_1^3 = \{(2,1),(3,2)\}$, $\mathcal{Q}_2^3 = \{(4,1),(1,2)\}$, $\mathcal{Q}_3^3 = \{(3,1)\}$, and $\mathcal{Q}_4^3 = \{(1,1)\}$.

III. Rate Allocation for CPGD

A. Problem Statement

Assume that the users in a K-user interference channel are operating at some *decodable* rate vector $\boldsymbol{R}_{\mathcal{K}} = [R_{j,k}]_{(j,k)\in\mathcal{K}}$, where $R_{j,k}$ is the rate of the k^{th} layer of transmitter j. $\boldsymbol{R}_{\mathcal{K}}$ can be the rate vector achievable by some decoding scheme inferior to the CPGD, e.g., the single-user decoders where all interference is treated as noise. Employing CPGDs allows for further incrementing the rate vector beyond $\boldsymbol{R}_{\mathcal{K}}$. Such further rate increment is of particular interest when they also satisfy a fairness criterion. We consider maximizing the sum-rate of the network such that the rates of all layers are incremented equally, i.e.,

$$\left\{\begin{array}{ll} \max & x\ , \\ \text{s.t.} & \tilde{\boldsymbol{R}}_{\mathcal{K}} = \boldsymbol{R}_{\mathcal{K}} + x\cdot \mathbf{1}_{1\times L} \text{ is decodable}\ , \end{array}\right. \tag{4}$$

where $\mathbf{1}_{1\times L} \triangleq [1,\ldots,1]$ and $L = \sum_{j=1}^{K} L_j$. Note that maximizing x is equivalent to maximizing $\sum_{j=1}^{K}\sum_{k=1}^{L_j}(R_{j,k}+x)$, which is the sum-rate of the K-user interference channel. Parameter x is an intermediate parameter that leverages solving the fair sum-rate optimization problem of interest. The physical meaning of x is the amount of change in the rate of each codebook. Due to the imposed fairness constraint, we require that the rates of all codebooks are incremented/decremented equally and aim to maximize the sum-rate subject to this fairness constraint. As will be made clear later, the number of layers allocated to each transmitter depends on the strength of the channel for that user, such that users with stronger links are assigned more layers. Therefore, such freedom in assigning different number of layers to the transmitters along with incrementing the rates of all layers equally (as required by (4)) has the advantage that it facilitates assigning higher *rates* to the users with stronger links. Let us also define $\tilde{R}_{j,k}$ as the rate of the k^{th} layer of the j^{th} transmitter after solving the above rate allocation problem and further define $r_{j,k}$ as the corresponding rate increment, i.e., $r_{j,k} \triangleq \tilde{R}_{j,k} - R_{j,k}$. For any set $\mathcal{A}\subseteq\mathcal{K}$, we denote the rate vectors $\tilde{\boldsymbol{R}}_{\mathcal{A}} \triangleq [\tilde{R}_{j,k}]_{(j,k)\in\mathcal{A}}$ and $\boldsymbol{r}_{\mathcal{A}} \triangleq [r_{j,k}]_{(j,k)\in\mathcal{A}}$.

B. Optimal Partitions and Rates

We propose an algorithm for solving (4) when the receivers employ CPGDs. The underlying motivation for the proposed algorithms is to alleviate the complexities associated with obtaining the best decodable set on one hand, and to control the amount of information exchange among the users on the other hand. The algorithm is comprised of two sub-routines: the first one aims to find the optimal partition for each receiver in a computationally efficient way and is executed locally by each receiver (Algorithm 1); and the second combines such local optimal solutions to solve the global fair sum-rate optimization problem (Algorithm 2).

For any two *disjoint* sets $\mathcal{U}$ and $\mathcal{V}$ that are subsets of $\mathcal{K}$, let $\mathcal{C}_i(\mathcal{U},\mathcal{V})$ denote the achievable rate region supported by the i^{th} receiver for *jointly* decoding the layers in $\mathcal{U}$ via ML decoding while the layers $\{\mathcal{K}\backslash(\mathcal{U}\cup\mathcal{V})\}$ have already been successfully decoded and subtracted and those in $\mathcal{V}$ are being treated as AWGN. Let us denote the channel that conveys the layer $x_{j,k}$ from the j^{th} transmitter to the i^{th} receiver by $h_{i,j}^k$, for which we clearly have $h_{i,j}^k = h_{i,j}$ for $k=1,\ldots,L_j$. Therefore, $\mathcal{C}_i(\mathcal{U},\mathcal{V})$ can be characterized as

$$\mathcal{C}_i(\mathcal{U},\mathcal{V}) = \left\{\boldsymbol{R}\in\mathbb{R}_+^{|\mathcal{U}|} \mid \sum_{j\in\mathcal{D}} R_j \le \mathcal{R}_i(\mathcal{D},\mathcal{V}),\ \forall\mathcal{D}\subseteq\mathcal{U}\right\}, \tag{5}$$

where upon defining $\boldsymbol{h}_{i,\mathcal{A}} \triangleq \left[\sqrt{\frac{P}{L_j}}\, h_{i,j}^k\right]_{(j,k)\in\mathcal{A}}$ for any $\mathcal{A}\subseteq\mathcal{K}$, the rate $\mathcal{R}_i(\mathcal{D},\mathcal{V})$ is given by

$$\mathcal{R}_i(\mathcal{D},\mathcal{V}) = \log_2\left(1+\frac{\|\boldsymbol{h}_{i,\mathcal{D}}\|^2}{\sigma^2+\|\boldsymbol{h}_{i,\mathcal{V}}\|^2}\right). \tag{6}$$

The normalizing factor $\sqrt{\frac{P}{L_j}}$ is to reflect equal power cross the layers of each transmitter. Let $\underline{\mathcal{Q}}^i = \{\mathcal{Q}_1^i,\ldots,\mathcal{Q}_{p_i}^i,\mathcal{Q}_{p_i+1}^i\}$ be a valid partition of $\mathcal{K}$. Based on this partition all layers $x_{i,k}$ are decodable by the i^{th} receiver if

$$\forall\ m\in\{1,\ldots,p_i\}:\qquad \boldsymbol{R}_{\mathcal{Q}_m^i} \in \mathcal{C}_i(\mathcal{Q}_m^i, \cup_{j=m+1}^{p_i+1}\mathcal{Q}_j^i)\ . \tag{7}$$

For the i^{th} receiver and the valid partition $\underline{\mathcal{Q}}^i$, after setting $\boldsymbol{r}_{\mathcal{Q}_m^i}^i = \left[r_{(j,k)}^i\right]_{(j,k)\in\mathcal{Q}_m^i}$, we define

$$\theta_i(\underline{\mathcal{Q}}^i) \triangleq \left\{\begin{array}{ll} \max & x\ , \\ \text{s.t.} & r_{j,k}^i = x,\quad \forall(j,k)\in\mathcal{K}, \\ & \tilde{\boldsymbol{R}}_{\mathcal{Q}_m^i} = \boldsymbol{R}_{\mathcal{Q}_m^i} + \boldsymbol{r}_{\mathcal{Q}_m^i}^i \in \mathcal{C}_i(\mathcal{Q}_m^i, \cup_{j=m+1}^{p_i+1}\mathcal{Q}_j^i), \\ & \forall m\in\{1,\ldots,p_i\}, \end{array}\right. \tag{8}$$

which suggests equal (fair) rate increment for all the layers included in $\{\mathcal{Q}_1^i,\ldots,\mathcal{Q}_{p_i}^i\}$ such that they are all decodable at the i^{th} receiver. There exists a delicate difference between (4) and (8). (4) characterizes the network-wide fair sum-rate optimization problem, whereas (8) solves the same problem selfishly for each receiver, in the sense that each receiver solves this problem with the purpose of maximizing the fair sum-rate with decodabiliy guarantees only for itself. So each receiver solves problem (8) locally and selfishly. Once all the receivers have solved (8), they exchange some information, through which they collaboratively solve (4). By defining $\widetilde{\boldsymbol{\mathcal{Q}}}^i$ as the ensemble of all valid partitions $\underline{\mathcal{Q}}^i$, the maximum layer rate increment with the fairness constraint from the viewpoint of the i^{th} user is given by

$$\theta_i^* \stackrel{\triangle}{=} \max_{\underline{\mathcal{Q}}^i \in \widetilde{\boldsymbol{\mathcal{Q}}}^i} \theta_i(\underline{\mathcal{Q}}^i) \ . \tag{9}$$

Clearly, solving (9) through the naive exhaustive search over all possible choices of $\underline{\mathcal{Q}}^i$ has a prohibitive complexity. By defining

$$\Delta_i(\mathcal{D},\mathcal{V}) \stackrel{\triangle}{=} \mathcal{R}_i(\mathcal{D},\mathcal{V}) - \sum_{(j,k)\in\mathcal{D}} R_{j,k} \ , \tag{10}$$

for any disjoint arbitrary $\mathcal{D},\mathcal{V} \subseteq \mathcal{K}$ we offer Algorithm 1 that yields two outputs for each receiver i and avoids exhaustive search for obtaining the optimal valid partition $\underline{\mathcal{Q}}^i$ that solves (9) and also provides the corresponding optimal rate increment.

For brevity and also for focusing the attention on the more practical issues pertinent to the implementation of these decoders, we describe the steps involved in Algorithm 1 and provide their properties and skip the proofs. The proofs hold essentially along the same line of arguments as those provided in [5] and[7], albeit with significant differences. In contrast to the rate allocation schemes in [5] and[7] that allocate one codebook to each transmitter, we allocate multiple codebooks to each transmitter, where the number of the codebooks is determined by the strength of the channels corresponding to that transmitter. Moreover, in [5] and[7] there is no constraint on the size of the number of transmitters than can be jointly decoded, which gives rise to a prohibitive complexity due to ML decoding. In contrast, here we constrain the number of layers to be decoded jointly by μ which is determined based on the computational complexity level that the receivers can afford.

It is noteworthy that the structure of Algorithm 1 rules finding the ordered partitions $\underline{\mathcal{Q}}^i = \{\mathcal{Q}_1^i, \dots, \mathcal{Q}_{p_i}^i, \mathcal{Q}_{p_i+1}^i\}$, in a reverse order, i.e., it first identifies $\mathcal{Q}_{p_i+1}^i$ and $\mathcal{Q}_1^i$ is the last to be identified. $\mathcal{Q}_{p_i+1}^i$ contains the indices of the layers that will not be decoded in any stage of the successive decoding procedure of the CPGD algorithm. $\mathcal{Q}_1^i$ is the set of the indices of the layers that will tolerate the least amount of increase in their rates (subject to the fairness constraint). Subsequently, after decoding the layers in $\mathcal{Q}_1^i$, among the remaining layers in $\mathcal{Q}_2^i$ become the set of the indices of the layers that will tolerate the least amount of increase in their rates and the same property holds for all other sets $\mathcal{Q}_3^i, \dots, \mathcal{Q}_{p_i}^i$.

We remark that the complexity of Algorithm 1 is determined by those of the optimization problems in lines 3 and 4. As the size μ is controlled to be small, even an exhaustive search for finding at most μ layers among the (at most) L existing layers will suffice to obtain a computationally efficient (of the order at most L^μ) way of solving these two problems. This is contrary to the approaches in [5] and[7] that impose no constraint on decoding size, for which the complexity of the exhaustive search grows exponentially with $|\mathcal{K}|$ and solving them necessitates resorting to the submodular optimization tools. Also the algorithm has at most $L = \sum_{j=1}^K L_j$ iterations, which consequently induced another level of computations with linear complexity in L.

Algorithm 1 - The Optimal Valid Partition for the Receiver i

```
1:  Initialize D = K, G = ∅, p_i = 0 , and ℓ = 1
2:  repeat
3:     Find δ_ℓ = min_{V≠∅, V⊆D, |V|≤μ} Δ_i(V,G)/|V|
4:     Find G^i_ℓ = arg min_{V≠∅, V⊆D, |V|≤μ} Δ_i(V,G)/|V|
5:     D ← D\G^i_ℓ and G ← G ∪ G^i_ℓ
6:     if (i,k) ∈ D for all k ∈ {1,...,L_i}
7:        r̃^i_{j,k} = +∞ for all (j,k) ∈ G^i_ℓ
8:     else
9:        r̃^i_{j,k} = δ_ℓ for all (j,k) ∈ G^i_ℓ , p_i ← p_i + 1
10:    end if
11:    ℓ ← ℓ + 1
12: until D = ∅
13: Set Q^i_m ← G^i_{ℓ−m} for 1 ≤ m ≤ p_i, and
    Q^i_{p_i+1} ← ∪_{m>p_i} G^i_{ℓ−m}
14: Output the rates {r̃^i_{j,k}} for (j,k) ∈ K and the partitions
    Q^i = {Q^i_1, ..., Q^i_{p_i}, Q^i_{p_i+1}}.
```

The main property of Algorithm 1 is formalized in the following theorem. The intuition behind Algorithm 1 and Theorem 1 is that it each receiver aims to identify the best set of interfering codebooks to decode that yields the highest sum-rate throughput. By focusing on each possible set of codebooks, these codebooks constitute a multiple access channel to the receiver. Therefore, the best set of codebooks is the one whose corresponding multiple access channel sustains a higher level of rate increment (or lower level of rate decrement). A naive, and computationally complex approach is the exhaustive search. Alternatively, each receiver employ a successive decoding approach such that successively identifies and discards the codebooks that removing them allows for further increasing the sum-rate. Such successive identification of the codebooks resumes until discarding a codebook penalizes the sum-rate.

Theorem 3.1: Algorithm 1 identifies the partition that maximizes $\theta_i(\underline{\mathcal{Q}}^i)$ over all possible valid partitions $\widetilde{\boldsymbol{\mathcal{Q}}}^i$ and obtains θ_i^*, i.e.,

$$\{\mathcal{Q}_1^i, \dots, \mathcal{Q}_{p_i+1}^i\} = \arg \max_{\underline{\mathcal{Q}}^i \in \widetilde{\boldsymbol{\mathcal{Q}}}^i} \theta_i(\underline{\mathcal{Q}}^i)$$
$$\text{and} \qquad \theta_i^* = \max_{\underline{\mathcal{Q}}^i \in \widetilde{\boldsymbol{\mathcal{Q}}}^i} \theta_i(\underline{\mathcal{Q}}^i) = \min_{(j,k)\in\mathcal{K}} \tilde{r}_{j,k}^i \ . \tag{11}$$

Through Algorithm 1 each receiver identifies its best partition in a distributed way by performing some local processing without any information exchange among different users. Such local processing can interestingly lead to solving the fair sum-rate optimization problem cast in (4) with limited information exchange. For this purpose, the i^{th} receiver reports the scalar θ_i^*, which is obtained in Algorithm 1 and is related to the maximum rate-increment possible for the i^{th} receiver, to all other transmitters. The globally optimal rate increment is obtained by Algorithm 2, which is simply the smallest rate increment that all users suggest.

Theorem 1: The rate vector yielded by Algorithm 2 satisfies $\tilde{\boldsymbol{R}}_\mathcal{K} \succeq \hat{\boldsymbol{R}}_\mathcal{K}$, where $\hat{\boldsymbol{R}}_\mathcal{K}$ is any decodable rate-vector such that $\hat{\boldsymbol{R}}_\mathcal{K} = \boldsymbol{R}_\mathcal{K} + x \cdot \mathbf{1}_{1\times L}$ for some $x \in \mathbb{R}$.

According to Theorem 1, by running Algorithm 1 each receiver specifies the best set of interferers to be decoded, which maximizes the sum-rate throughput. Given the different sum-rate throughputs that different users can sustain based on their local processing in Algorithm 1 and Theorem 1, the network-wide optimal sum-rate throughput is determined by the bottleneck user, which is the one that sustains the smallest sum-rate throughput.

Algorithm 2 - Fair Rate Allocation

```
Input R = [R_{j,k}]_{(j,k)∈K}
for i = 1, ..., K do
    Run Algorithm 1 to determine {r̃^i_{j,k}} for (j,k) ∈ K and Q^i
end for
Obtain θ* = min_i θ*_i where θ*_i = min_{(j,k)∈K} r̃^i_{j,k}.
Update R̃_{j,k} ← R_{j,k} + θ* for (j,k) ∈ K
Output R̃ = [R̃_{j,k}]_{(j,k)∈K}
and the partitions {Q^1, ..., Q^K}.
```

C. Iterative Rate Allocation

Note that Algorithm 2 solves (4) by increasing the rate of each layer by an amount of θ^*, which is the same for all layers. As stated in [4], this strict notion of fairness will penalize the system throughput in terms of the sum-rate. In this subsection, we propose Algorithm 3 to further increase the rates of some layers so that the sum-rate can be further increased.

In contrast to Algorithm 2, Algorithm 3 iteratively makes a new rate increment recommendation for all layers based on the allocated rates obtained from the previous iteration, denoted as $r^i_{j,k}$ for all receivers i in a distributive manner. The rate increment for the layer (j,k) is $\tilde{r}_{j,k} = \min_i \tilde{r}^i_{j,k}$ which ensures that all layers are decodable after each iteration of rate increment recommendation. The following Theorem 2 shows that for any layer the allocated rate vector from Algorithm 3 is larger than or equal to that from Algorithm 2.

Algorithm 3 - Iterative Rate Allocation

```
Input R = [R_{j,k}]_{(j,k)∈K}
repeat
    for i = 1, ..., K do
        Run Algorithm 1 to determine {r̃^i_{j,k}} for (j,k) ∈ K and Q^i
    end for
    Obtain r̃_{j,k} = min_{i=1}^K r̃^i_{j,k} for (j,k) ∈ K.
    Update R̃_{j,k} ← R̃_{j,k} + r̃_{j,k} for (j,k) ∈ K, R ← R̃
until R̃ converges
Output R̃ = [R̃_{j,k}]_{(j,k)∈K} and the partitions {Q^1, ..., Q^K}.
```

Theorem 2: Denote the output rate vector yielded by Algorithms 2 and 3 as $\tilde{\boldsymbol{R}}_{\mathcal{K},2}$ and $\tilde{\boldsymbol{R}}_{\mathcal{K},3}$, respectively. We have $\tilde{\boldsymbol{R}}_{\mathcal{K},3} \succeq \tilde{\boldsymbol{R}}_{\mathcal{K},2}$.

Remark: Algorithm 4 relaxes the rate constraint that the rate increments of all layers are not restricted to the minimal rate increment as recommended by Algorithm 2. Therefore, by running Algorithm 4 the rate increments of all layers are larger than or equal to those obtained from Algorithm 2. Based on Theorem 2, *Algorithm 3* is employed for the design of practical coding scheme in the remainder of this paper.

D. Numerical Results

Consider a single-antenna interference channel with $K = 6$ transceivers. For $i, j \in \{1, \ldots, K\}$, the channel coefficients $h_{i,j}$ are distributed as $\mathcal{N}_{\mathbb{C}}(0, 1)$. We run Algorithms 2 and 3 to assess the performance of the proposed coding scheme for Gaussian modulation and infinite-length random codes. The initial rate $\boldsymbol{R}_{\mathcal{K}}$ is selected to be the rate vector achievable by using single-user decoders that treat all interference as noise. Fig. 2 shows the sum-rate obtained from Algorithms 2 and 3 for the un-layered coding scheme with the group size $\mu = 1$, 2, and 3. It is seen that the sum-rate obtained from Algorithm 3 is larger than that from Algorithm 2. In the remainder of this paper, *Algorithm 3* is employed to obtain all numerical and simulation results.

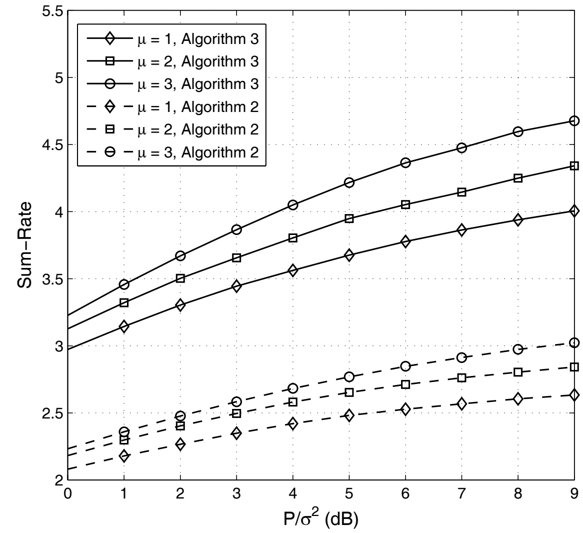


Fig. 2. The sum-rate given by Algorithms 2 and 3 assuming Gaussian signaling and infinite-length random codes.

Fig. 3 shows the performance of the layered coding scheme with $\mu = 1$ for different number of layers $L_j = 2$, 5, 10, and 15 for all users. It is seen that the coding schemes with $L_j = 5$ significantly outperform those with $L_j < 5$, and further increasing L_j brings little performance gain over $L_j = 5$.

Furthermore, we compare the sum-rate of users for different group sizes $\mu = 1$, 2, and 3, as illustrated in Fig. 4. For the layered coding scheme, we set $L_j = 5$ layers for all users. At each channel SNR value P/σ^2, one hundred channel realizations are simulated and the average sum-rate is plotted against the channel SNR. Also plotted is the performance of the single-user decoding, denoted by "MMSE decoding" (where "MMSE" stands for minimum mean square error). It is seen that CPGD schemes exhibit significant performance gains over the single-user decoding.

We have the following important observation from Fig. 4. Without layering, an increase in the group size μ leads to an increase in the sum-rate. Moreover, for a fixed group size, the layered coding can indeed provide higher sum-rate than the un-layered coding. More interestingly, it is seen that with layering, when the group size $\mu = 1$, there is a substantial gain in sum-rate compared with the un-layered coding scheme, and further increasing μ shows little performance gain over $\mu = 1$. Hence, the layered coding scheme with the group size $\mu = 1$ achieves a good tradeoff between the performance and

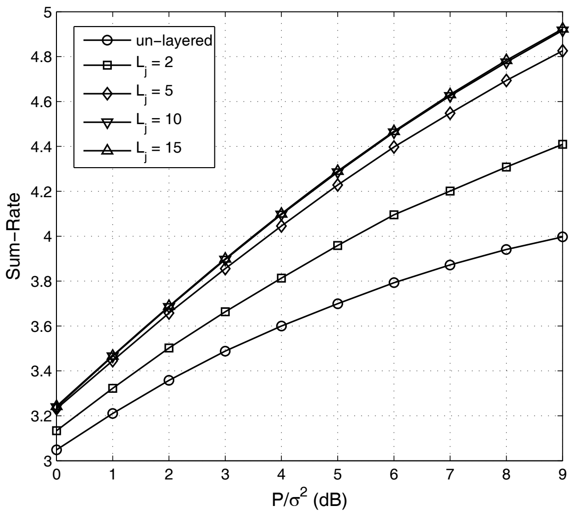


Fig. 3. The sum-rate given by Algorithm 3 assuming Gaussian signaling and infinite-length random codes for different number of layers.

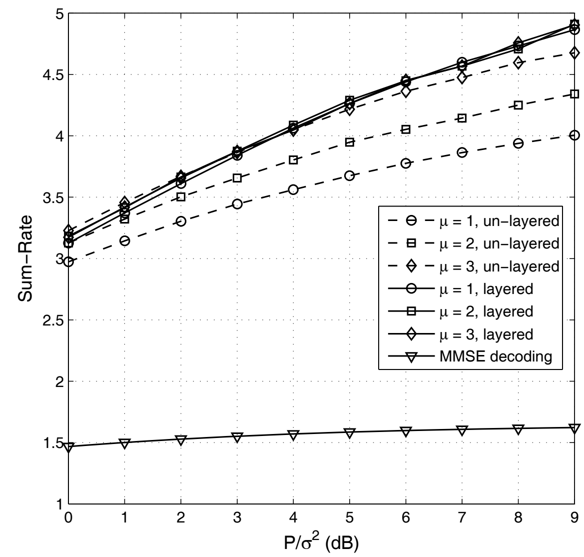


Fig. 4. The sum-rate given by Algorithm 3 assuming Gaussian signaling and infinite-length random codes.

complexity. It will be seen in the next section that this is also the case when practical modulations and channel codes are employed.

Remark: For $\mu = 1$, the group selection in Algorithm 1 can be simplified. In lines 3 and 4, given the selected layers $\mathcal{G}$, we find the optimal layer (j^*, k^*) as follows,

$$\begin{aligned} (j^*, k^*) &= \arg \min_{(j,k)\in\mathcal{D}} \Delta_i((j,k),\mathcal{G}) \\ &= \arg \min_{(j,k)\in\mathcal{D}} \Big(\mathcal{R}_i((j,k),\mathcal{G}) - R_{j,k}\Big)\ , \quad (12) \end{aligned}$$

where the rate $\mathcal{R}_i((j,k),\mathcal{G})$ is given by

$$\mathcal{R}_i((j,k),\mathcal{G}) = \log_2\left(1 + \frac{P}{L_j}\cdot\frac{\|h_{i,j}\|^2}{\sigma^2 + \|\boldsymbol{h}_{i,\mathcal{G}}\|^2}\right)\ , \quad (13)$$

and the rate increment is given by

$$\delta_\ell = \arg \min_{(j,k)\in\mathcal{D}} \Delta_i((j,k),\mathcal{G})\ . \quad (14)$$

Thus both steps involve linear search over the set $\mathcal{D}$, whose cardinality, in the worst case, is equal to the total number of layers L. The same will hold true when practical modulations and channel codes are employed.

IV. Practical Transmission Scheme

The rate increment solutions obtained through the CPGD procedure in Algorithms 1 and 3 are not necessarily viable in practice. There are two major challenges involved. First in characterizing the rate regions in (5) and the achievable rates in (6) it is implicitly assumed that the codewords are drawn from Gaussian codebooks and we have ideal infinite-length random codes for achieving these rates. In practice, however, there is always a gap between the ideal Gaussian rates and the rates achieved by practical modulation and coding schemes, that we aim to minimize. More specifically, for practical rate allocation the rate $\mathcal{R}_i(\mathcal{D},\mathcal{V})$ is given by

$$\mathcal{R}_i(\mathcal{D},\mathcal{V}) = \log_2\left(1 + \gamma\frac{\|\boldsymbol{h}_{i,\mathcal{D}}\|^2}{\sigma^2 + \|\boldsymbol{h}_{i,\mathcal{V}}\|^2}\right), \quad (15)$$

where γ denotes the gap between the Gaussian modulation with infinite-length random codes and the practical QAM with finite-length codes. The other challenge is that, in contrast to the output rates of Algorithm 3 which can be set arbitrarily such that the system needs to perform online channel code construction which incurs unacceptable computational complexity, in practical transmission the channel codes are constructed offline and only a finite number of channel codes can be stored and selected according to current channel realizations.

A. Rate Selection Procedure

To address these challenges, we carry out a three-step procedure that selects a transmission rate for each user and its layers. In the first step, we examine whether all users can operate at rates well above zero that can be implemented in practice. In case there exist users that can operate only at very low rates we inactivate them. Then we determine how many layers each active transmitter should have. Subsequently, based on the number of layers obtained for each transmitter, by running Algorithm 3 we determine the rates for the layers of each transmitter, and inactivate the layers which operate at very low rates. In the second step, in order to find some implementable rates close to those yielded by Algorithm 3 we quantize the rates according to a quantization codebook that satisfies some optimality measure. These quantized rates are subsequently considered as *coarse* approximations for possible practical rates. As the quantized rates must be decodable, each quantized rate is smaller than its original counterpart which incurs some loss in spectral efficiency. In the final step, we compensate for such loss, through devising a *fine* tuning scheme for increasing the rates beyond their quantized values and make them as close as possible to the original rates yielded by Algorithm 3 while being also practical.

We employ a table of spectrum rate $\mathcal{T} = \{d_1, ..., d_T\}$ for the user and layer inactivation in the first step and the rate quantization in the second step. Each element $d_j \in \mathcal{T}$ is associated with a pre-designed channel code with Nd_j information bits and a pre-designed modulation scheme which maps the coded bits to channel symbols. In the user (layer) inactivation, the users (layers) with allocated rates smaller than d_1 will be inactivated. If the rate of a layer is quantized to the rate d_j for some $1 \leq j \leq T$, then we employ a codebook with the number of information bits Nd_j and the corresponding modulation scheme for that layer. To achieve the rate enhancement in the final step, we choose the number of transmitted symbols smaller than N while keeping the number of information bits Nd_j unchanged. The rate enhancement could also be done by increasing the number of information bits for each layer while keeping the number of transmitted symbols unchanged. However, for a practical capacity-approaching codebook, changing the number of coded bits flexibly without regenerating the codebook can be achieved using rateless codes, while there is no practical capacity-approaching codebook which allows changing the number of information bits flexibly.

1) User Inactivation and Message Layering: For initializing the rate allocation procedure we start off with assigning one codebook to each transmitter, i.e., $\forall j$ we have $L_j = 1$. We run Algorithm 3 and inactivate user j for which $\tilde{R}_j < d_1$. Let us denote the set of users retained as active users by $\mathcal{H} \subseteq \{1, \ldots, K\}$. By inactivating some users, the rest are exposed to less interference and can possibly sustain higher rates. For this reason, after inactivating the users with rates smaller than d_1 we run Algorithm 3 again to obtain the new set of rate increment suggestions for all active users, denoted as $\{\tilde{R}_j^{(1)}\}_{j\in\mathcal{H}}$.

Now based on the rates $\{\tilde{R}_j^{(1)}\}_{j\in\mathcal{H}}$ we decide how many codebooks should be assigned to each transmitter as follows,

$$\forall j \in \mathcal{H}: \qquad L_j = \max\Big\{1, \Big\lfloor \frac{\tilde{R}_j^{(1)}}{\Delta R} + 0.5 \Big\rfloor\Big\}, \tag{16}$$

where ΔR is the unit rate to be sustained by each codebook. Given the number of layers, we can now assign multiple codebooks to each active transmitter and implement the CPGD in order to find the new rate for each layer by running Algorithm 3. Let us denote the rate of the k^{th} layer of the active transmitter $j \in \mathcal{H}$ by $\tilde{R}_{j,k}^{(2)}$. It is possible that the rates of some *layers* are not large enough to be implementable. We then inactivate the layers with rates smaller than d_1, and update the number of layers L_j by the number of layers with rates larger than d_1, i.e.,

$$L_j \leftarrow \sum_{k=1}^{L_j} \mathbb{1}_{\{\tilde{R}_{j,k}^{(2)} \geq d_1\}}, \quad 1 \leq j \leq K. \tag{17}$$

Based on the updated number of layers L_j, we run Algorithm 3 again to compute the rates $\tilde{R}_{j,k}^{(3)}$ for all layers. We define

$$R_{\text{sum}}^{(3)} = \sum_{j=1}^{K} \sum_{k=1}^{L_j} \tilde{R}_{j,k}^{(3)} \cdot \mathbb{1}_{\{\tilde{R}_{j,k}^{(3)} \geq d_1\}}. \tag{18}$$

We compare $R_{\text{sum}}^{(3)}$ with $R_{\text{sum}}^{(1)} = \sum_{j\in\mathcal{H}} \tilde{R}_j^{(1)}$. If the layered scheme provides larger rate, i.e., $R_{\text{sum}}^{(3)} > R_{\text{sum}}^{(1)}$, we employ the layered scheme; otherwise we employ the un-layered scheme. Hence we have the following sum rate

$$\tilde{R}_{\text{sum}}^{(3)} = \max\Big\{R_{\text{sum}}^{(3)}, R_{\text{sum}}^{(1)}\Big\}. \tag{19}$$

2) Rate Quantization and Modulation Selection (Coarse Tuning): Assume that we get the rate $\bar{R}_{j,k}$ for the layer (j, k) after the message layering and the rate allocation for each layer. Now we map the rate $\bar{R}_{j,k}$ to a pre-designed channel node and modulation scheme associated with some $\hat{R}_{j,k} \in \mathcal{T} = \{d_1, \ldots, d_T\}$ as follows,

$$\hat{R}_{j,k} = \max_{d_q \in \mathcal{T}, d_q \leq \bar{R}_{j,k}} d_q, \tag{20}$$

and employ the channel code with $N\hat{R}_{j,k}$ information bits and the corresponding modulation scheme which maps $m_{j,k}$ coded bits to one symbol. To encode the layer (j, k), we encode the $N\hat{R}_{j,k}$ information bits to $N\hat{m}_{j,k}$ coded bits, and map those coded bits to N channel symbols.

The modulation scheme is selected from 4-, 16-, and 64-QAMs according to the rules in Table I. Let us explain the rules in Table I. Note that for the 4-, 16- and 64-QAMs, the capacity of bit-interleaved coded modulation with Gray mapping well approximates that of the coded modulation, and the capacity of coded modulation well approximates that of Gaussian modulation when the spectral efficiency is below 1, 2, and 3 bits per channel use, respectively [8]. Therefore in Table I, when $\mu = 1$, for $d_j \leq 1.0$ we associate it with 4-QAM, for $1.0 \leq d_j \leq 2.0$ we associate it with 16-QAM, and for $d_j \geq 2.0$ we associate it with 64-QAM. For controlling the complexity of the ML decoders, we constrain the number of layers to be decoded jointly in each iteration of the successive decoder of Algorithm 1 to be $\mu \leq 3$. For controlling the demodulation complexity when $\mu = 3$, i.e., three users can be jointly decoded, all users can employ only 4-QAM. For the same reason, for $\mu = 2$, all users only employ 4- and 16-QAMs, and the joint decoding of two users both employing 16-QAM is not allowed, since it incurs large complexity to the 2-user a posteriori probability (APP) detector.

For $\mu = 2$, we provide a post-processing procedure (Algorithm 4) to avoid the joint decoding of two users (layers) both employing 16-QAM. For a decoding group $\mathcal{Q}_\ell^i$ with $|\mathcal{Q}_\ell^i| = 2$, if both layers employs 16-QAM, we change the rate of smaller one to 4-QAM. The post-processing procedure is detailed in Algorithm 4.

TABLE I
MODULATION SCHEMES FOR DIFFERENT RATE VALUES

Group Size	$d_j \leq 1$	$1 \leq d_j \leq 2$	$2 < d_j$	Symbol Mapping
1	4-QAM	16-QAM	64-QAM	Gray mapping
2	4-QAM	16-QAM	16-QAM	Natural mapping
3	4-QAM	4-QAM	4-QAM	Natural mapping

Algorithm 4 - Rate Post-processing for $\mu = 2$

for $i \in \mathcal{H}$ and $\mathcal{Q} \in \{\mathcal{Q}_\ell^i\}_{1 \leq \ell \leq p_i}$;
  if $|\mathcal{Q} = \{(j_1, k_1), (j_2, k_2)\}| = 2$
    and $\hat{R}_{j_1,k_1} > 1.0, \hat{R}_{j_2,k_2} > 1.0$
    if $\hat{R}_{j_1,k_1} < \hat{R}_{j_2,k_2}$
      Let $\hat{R}_{j_1,k_1} \leftarrow \max_{d_q \in \mathcal{T}, d_q \leq 1.0} d_q$
    else
      Let $\hat{R}_{j_2,k_2} \leftarrow \max_{d_q \in \mathcal{T}, d_q \leq 1.0} d_q$
    end if
  end if
end for

3) Rate Enhancement (Fine Tuning): Recall that after rate quantization, there is a reduction in the rates of layers. Such reduction incurs a loss in terms of the spectral efficiency of the network. To compensate for such loss, we boost each rate by a factor η, scaling up the quantized rate $\hat{R}_{j,k}$ to $\eta \cdot \hat{R}_{j,k}$. We need to determine the largest possible η denoted as η^*, such that the new scaled rates do not violate the Shannon limits. For each subset $\mathcal{U} \subseteq \mathcal{Q}_m^i$ let us define the sum-rates

$$\begin{aligned} R_{i,m}^{\text{sum}}(\mathcal{U}) &= \log_2\left(1 + \gamma \frac{\|\boldsymbol{h}_{i,\mathcal{U}}\|^2}{\sigma^2 + \|\boldsymbol{h}_{i,\mathcal{V}}\|^2}\right) \\ \text{and} \quad \hat{R}_{i,m}^{\text{sum}}(\mathcal{U}) &= \sum_{(j,k)\in\mathcal{U}} \hat{R}_{j,k} \ , \end{aligned} \tag{21}$$

where $\mathcal{V} = \cup_{j=m+1}^{p_{i+1}} \mathcal{Q}_j^i$. Given the subset $\mathcal{U}$ of $\mathcal{Q}_m^i$, $R_{i,m}^{\text{sum}}(\mathcal{U})$ is Shannon limit of the sum-rate of the layers in $\mathcal{U}$, where the parameter γ is incorporated to account for the loss due to practical constellation and coding schemes, and $\hat{R}_{i,m}^{\text{sum}}(\mathcal{U})$ is the sum of the quantized rates of the same layers obtained in Section IV-A2. For $\mathcal{Q}_m^i$, the rates $\hat{R}_{j,k}$ can be increased to $\eta_{i,m} \cdot \hat{R}_{j,k}$ for $(j,k) \in \mathcal{Q}_m^i$ where $\eta_{i,m}$ is given by

$$\eta_{i,m} = \min_{\mathcal{U}\subseteq\mathcal{Q}_m^i} \frac{R_{i,m}^{\text{sum}}(\mathcal{U})}{\hat{R}_{i,m}^{\text{sum}}(\mathcal{U})}. \tag{22}$$

Based on these definitions, the quantized rates $\{\hat{R}_{i,k}\}$ can be increased as much as the scaled sum-rate $\eta^* \cdot \hat{R}_{i,m}^{\text{sum}}$ does not exceed its limit $R_{i,m}^{\text{sum}}$ for all i, m, where

$$\eta^* = \min_{i,m} \eta_{i,m} = \min_{i,m} \min_{\mathcal{U}\subseteq\mathcal{Q}_m^i} \frac{R_{i,m}^{\text{sum}}(\mathcal{U})}{\hat{R}_{i,m}^{\text{sum}}(\mathcal{U})}. \tag{23}$$

After the above three steps, the transmission data is assembled as follows. For the layer (j,k), we choose the predesigned channel code with the number of information bits $N\hat{R}_{j,k}$, encode those information bits to $Nm_{j,k}$ coded bits using the rateless codes to be optimized in Section V, and map these coded bits to N symbols. The rate enhancement is implemented via transmitting

$$N_s = \Delta N \cdot \lceil \frac{N}{\eta^* \Delta N} \rceil \tag{24}$$

symbols, where ΔN is the number of symbols in each transmission burst.

Remark: In real transmission, the role of fine tuning is to set up a starting point after which the receiver is allowed to perform decoding and sending feedback. Doing so significantly reduces the feedback load compared to the transmission scheme in which the receivers begin decoding and sending feedback from the beginning of the transmission after each transmission burst of ΔN symbols.

4) An Illustrative Example of the Proposed Coding Scheme: In Fig. 5 we show an illustrative example for the procedure of the proposed coding scheme in an interference channel with four users. Let the spectrum rate table $\mathcal{T} = \{d_1, d_2, d_3, d_4\}$. After the user inactivation and message layering based on the threshold rate d_1, user 2 is inactivated and users 1 and 3 are divided into 2 layers. In the rate quantization and modulation selection, the rates of the five layers $(1,1)$, $(1,2)$, $(3,1)$, $(3,2)$, and $(4,1)$ are quantized to be d_1, d_2, d_3, d_1, and d_4, respectively, i.e., the number of information bits assigned to these layers are Nd_1, Nd_2, Nd_3, Nd_1, and Nd_4, respectively, where N is the nominal number of transmitted symbols. In the rate enhancement, the number of transmitted symbols is decreased to $\eta^* N$ for some $\eta^* < 1$, and thus the rate of the five layers are enhanced to d_1/η^*, d_2/η^*, d_3/η^*, d_1/η^*, and d_4/η^*, respectively.

The optimization of the table can be done in two steps. In the first step, we optimize d_1 based on the user inactivation and message layering, to maximize the sum rate of all layers; and in the second step, we optimize the remaining rates d_2, d_3, and d_4 based on optimal scalar quantization, to minimize the rate loss due to the rate quantization. The design of the rate table $\mathcal{T}$ is addressed in the next subsection.

B. Parameter Design

In this section, we design the rate table $\mathcal{T}$. In contrast to the above rate selection procedure which is performed online according to each channel realization, designing the rate table $\mathcal{T}$ is done offline for the ensemble of channel realizations considered. We collect enough samples of channel realizations as a large training set, and design the rate table $\mathcal{T}$ according to the training set. Assume that all elements d_q in $\mathcal{T}$ belongs to some candidate set $\mathcal{V} = \{v_1, \ldots, v_S\}$, and we set $v_0 = 0$ and $v_{S+1} = +\infty$.

As we can see, given the channel realization, the sum-rate $\tilde{R}_{\text{sum}}^{(3)}$ in (19) is a function of the threshold rate d_1, denoted as $\tilde{R}_{\text{sum}}^{(3)}(d_1)$. Let

$$\bar{R}_{\text{sum}}^{(3)}(d_1) \triangleq \mathbb{E}\left[\tilde{R}_{\text{sum}}^{(3)}(d_1)\right] \ , \tag{25}$$

where the expectation is over the ensemble of channel realizations, which is computed as the averaged $\tilde{R}_{\text{sum}}^{(3)}(d_1)$ based

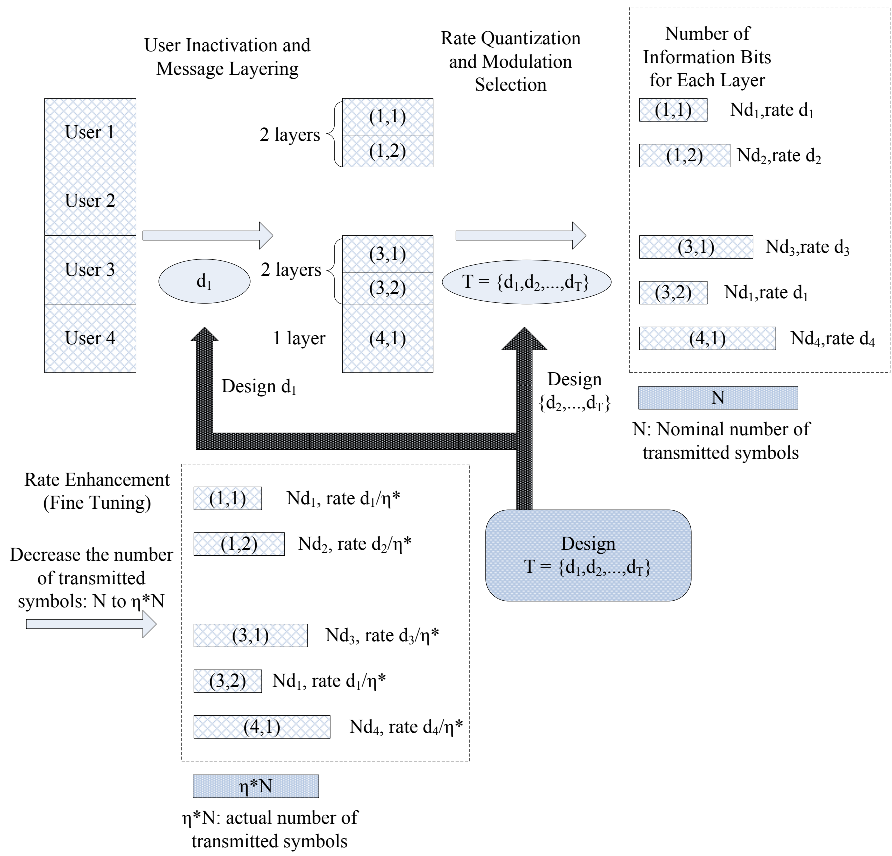


Fig. 5. An simple illustrative example for the proposed coding scheme.

on all channel realizations in the training set. We select d_1 as follows,

$$d_1 = \begin{cases} \max_{u \in \mathcal{V}} & u\ , \\ \text{s.t.} & \bar{R}^{(3)}_{\text{sum}}(u) \geq \max_{v \in V} \bar{R}^{(3)}_{\text{sum}}(v) - \epsilon \end{cases} \tag{26}$$

where ϵ is to avoid the extreme low rate of d_1 when $\arg\max_v \bar{R}^{(3)}_{\text{sum}}(v)$ is very small.

In the next step we find the optimal quantized rates d_2, d_3, ...,d_T to minimize the rate loss due to the quantization operation (20). Based on the optimal d_1 from (26), we run the user inactivation and message layering for all the channel realizations in the training data set, and get the empirical distribution of the rate $\bar{R}_{j,k}$ (unquantized version) for each layer (j,k). Let p_q be the fraction of rates $\bar{R}_{j,k}$ that fall into the interval $[v_{q-1}, v_q)$ for $q \in \{1, \ldots, S+1\}$. For any two consecutive quantization levels d_q and d_{q+1} we define $D(d_q, d_{q+1})$ as the rate distortion due to quantizing the rates that lie in the interval $[d_q, d_{q+1})$. As d_q and d_{q+1} are to be selected from $\mathcal{V}$, then $\exists \quad m < n \in \{1, \ldots, S\}$ such that $d_q = v_m$ and $d_{q+1} = v_n$. Therefore, the distortion of interest is the weighted sum of the distortions over the intervals $[v_m, v_{m+1})\ ,\ \ \ldots, [v_{n-1}, v_n)$, i.e.,

$$D(d_q, d_{q+1}) = \sum_{i:\ d_q < v_i < d_{q+1}} (v_i - d_q) p_i\ .$$

Then given d_1 obtained from (26), we design the rest of $\mathcal{T}$ by solving the following optimization problem,

$$\begin{cases} \max_{\{d_2,\ldots,d_T\} \subset \mathcal{V}} & \sum_{q=1}^{T} D(d_q, d_{q+1})\ , \\ \text{s.t.} & d_1 < d_2 < \cdots < d_T < d_{T+1} = +\infty. \end{cases} \tag{27}$$

This problem can be solved using a dynamic programming approach as the distortion function can be solved recursively in T stages where the optimal solution of each stage does not affect the optimal solution of later stages.

C. Numerical Results

We consider the same scenario as that in Section III-D. Assume that the size of the rate quantization table is $|\mathcal{T}| = 4$. From the code profile optimization results, we have $\gamma = 0.63$ for $\mu = 1$, $\gamma = 0.45$ for $\mu = 2$. We set $\gamma = 0.45$ for $\mu = 3$, which is an upper bound for the error free decoding, since we need to ensure the error-free decoding of any two users among the three.

We choose the maximum group size μ based on the sum-rate $\bar{R}^{(3)}_{sum}$ according to (25). Let the set of candidates $\mathcal{V} = \{0.001, 0.002, 0.003, ..., 4.000\}$. In Fig. 6, we plot the sum-rates $\bar{R}^{(3)}_{\text{sum}}(d_1)$ for $\mu = 1, 2$ and 3 and the cutoff rate d_1 from 0.0 to 0.9 with and without layering. The average sum-rates $\bar{R}^{(3)}_{\text{sum}}(d_1)$ are computed for the SNR from 0dB to 9dB each with 1000 random channel realizations. It can be seen that at the optimum cutoff rate which maximizes $\bar{R}^{(3)}_{\text{sum}}(d_1)$, the coding schemes for $\mu = 1$ outperform those for $\mu = 2$ and 3.

Next we optimize the rate quantization table $\mathcal{T}$ according to (27) in Section IV-B. The optimum cutoff rates d_1 are 0.45 and

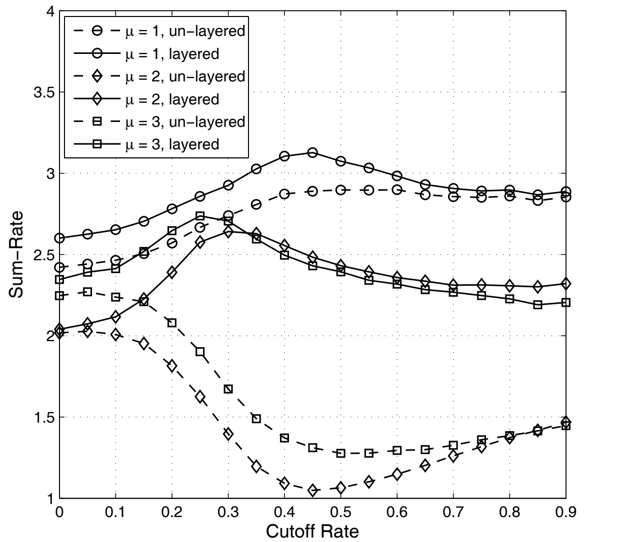


Fig. 6. The averaged sum-rate $\bar{R}^{(3)}_{sum}(d_1)$ for different cutoff rate d_1.

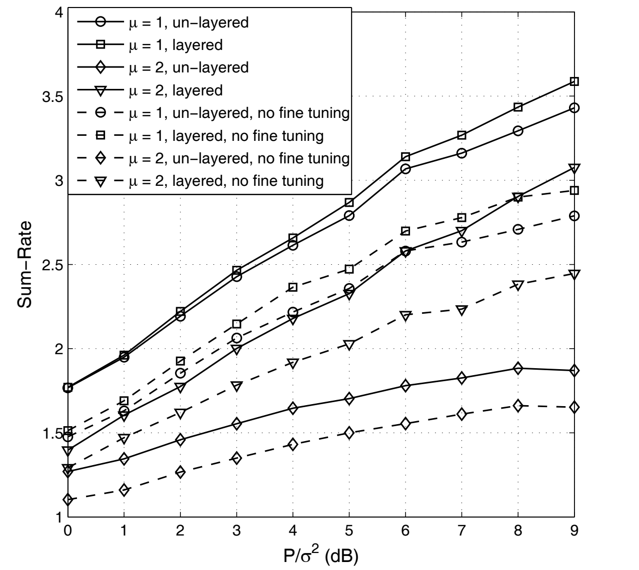


Fig. 7. The optimal predicted sum-rate under practical modulation and coding schemes.

0.60 for the layered and un-layered coding schemes with $\mu = 1$. We have that for $\mu = 1$, $\mathcal{T} = \{0.45, 1.08, 1.87, 2.80\}$ for the layered coding scheme and $\mathcal{T} = \{0.60, 1.28, 2.02, 2.88\}$ for the un-layered coding scheme. We also optimize the rate quantization table $\mathcal{T}$ for $\mu = 2$ and $\mu = 3$ with and without layering. Fig. 7 shows the sum quantized rates with and without the fine turning for different coding schemes with group size $\mu = 1$ and 2. It is seen that the layered coding scheme significantly outperforms the un-layered scheme. Moreover, the layered scheme with $\mu = 1$ performs the best among all the schemes. This is because for $\mu = 1$ the performance gap ($\gamma = 0.63$) is significantly smaller than that for $\mu = 2$ ($\gamma = 0.45$). It is also seen that the fine tuning offers a rate enhancement about 20%. For the coding schemes with $\mu = 3$, with and without layering, we observe that they perform worse than $\mu = 1$, and the fine tuning offers significant rate enhancement.

V. Rateless Code Design

A. Transmission Using Raptor Codes

We employ the Raptor code, which is a concatenated code with an inner LT code and an outer linear code, to implement the rateless codebook. In particular, we use the doped Raptor code [9], [10] with a rate-0.95 IRA precode. The reason we employ the Raptor code is as follows. Firstly, it exhibits near-capacity performance for both the single-user channel and the two-user multiple-access channel. Secondly, the flexible online fine tuning can be easily implemented by fixing the precode and adjusting the number of LT coded bits.

For layer (j, k), we first encode the $N\hat{R}_{j,k}$ information bits using a rate-0.95 IRA precode, and then perform the LT encoding to generate $m_{j,k}N_s$ coded bits (N_s symbols). It may happen that the decoding of some layer based on the N_s received symbols fails. In this case, we perform an incremental transmission where each time each layer transmits additional ΔN parity symbols, until successful decoding. In most cases the decoding succeeds after transmitting N_s symbols; and even if the incremental transmission happens, usually only one additional burst of ΔN symbols suffices.

In the remainder of this section we design the code profiles to increase the parameter γ in (21) and reduce the loss in spectral efficiency due to the practical code design. We consider the code profile design for $\mu = 1$ and $\mu = 2$.

B. Code Design for $\mu = 1$

Since each time one user is decoded, we only need to consider the code profile optimization for the single-user AWGN channel.

1) Code Profile Optimization: When $|\mathcal{Q}^i_m| = 1$, only one layer is decoded by the i^{th} receiver. Therefore, the employed profile should perform well for the single-user channel over a wide range of code rates. We find a profile that minimizes the maximum gap between the profile threshold and the capacity yielded by Gaussian signaling among the rates $\mathcal{R}_0 = \{0.2, 0.3, 0.4, 0.5\}$. For each code rate $r \in \mathcal{R}_0$, the threshold SNR for Gaussian signaling is given by

$$\mathcal{J}^{(1)}_G(r) = 2^r - 1 \ . \tag{28}$$

We define $\mathcal{J}^{(1)}_P(r, \{\lambda_j\})$ as the SNR threshold for the LT output node profile $\{\lambda_j\}$ (node perspective) for BPSK modulation, which is computed using the extrinsic information transfer (EXIT) function [11]. We optimize the profile via finding

$$\{\lambda^*_j\} = \arg\min_{\{\lambda_j\}} \ \max_{r\in\mathcal{R}_0} \frac{\mathcal{J}^{(1)}_P(r, \{\lambda_j\})}{\mathcal{J}^{(1)}_G(r)} \ , \tag{29}$$

using differential evolution (DE) method [12]. Although in our system 16- and 64-QAMs are also employed, since Gray mapping is employed, the optimized code profile for BPSK (QPSK) is also a good profile for the 16- and 64-QAMs. Let $\mathcal{R}_0^{M_1}$ be a subset of the combination of the code rate and modulation schemes $R_0 \times \{$4-QAM, 16-QAM, 64-QAM$\}$ where each elements $(r, m) \in \mathcal{R}_0^{M_1}$ satisfy the constraint of the spectrum efficiency and modulation schemes for $\mu = 1$ given in Table I. For the employed codes with the optimized profile $\{\lambda_j^*\}$, we compute

$$\gamma = \max_{(r,m)\in\mathcal{R}_0^{M_1}} \frac{\mathcal{J}_S^{(1)}(r, m, \{\lambda_j^*\})}{\mathcal{J}_G^{(1)}(r)}, \quad (30)$$

where $\mathcal{J}_S^{(1)}(r, m, \{\lambda_j^*\})$ is the error-free decoding SNR for the profile $\{\lambda_j^*\}$ from simulations for $(r, m) \in \mathcal{R}_0^{M_1}$. For each $\mathcal{J}_S^{(1)}(r, m, \{\lambda_j^*\})$, one hundred channel realizations are simulated and the decoding is error-free if there is no decoding error.

2) EXIT Function: The threshold $\mathcal{J}_P^{(1)}(r, \{\lambda_j\})$ is computed using EXIT functions as follows. Let $\{\rho_j\}$ be the LT input node profile also from the node perspective, which can be derived from the profile $\{\lambda_j\}$. Let $\{\tilde{\lambda}_j\}$ and $\{\tilde{\rho}_j\}$ be the corresponding profiles from the edge perspective. Let $\{\gamma_i\}$ and $\{\tilde{\gamma}_i\}$ denote the profiles of the outer IRA code from the node and edge perspectives, respectively, and $d_c = 59$ denote the concentrated check node degree for the IRA code. We have that $\{\rho_j\}$ satisfies a Poisson distribution with the average degree

$$\bar{\rho} = \frac{0.95 - r}{r}\bar{\lambda}, \quad (31)$$

where $\bar{\lambda}$ is the average degree of the profile $\{\lambda_j\}$. Let I_{vc} and I_{cv} be the mutual information (MI) from the LT input nodes to the LT output nodes and the MI from the LT output nodes to the LT input nodes, respectively, and I_{vp} and I_{pv} be the MI from the LT input nodes to the IRA parity check nodes and the MI from the IRA parity check nodes to the LT input nodes, respectively. Let I_{dv} and I_{dc} be the MI from the channel to the LT input and output nodes, respectively. For the channel with AWGN of variance σ^2, we have [11]

$$\begin{aligned} I_{dv} &= I_{dc} = J(\sigma) \\ &\triangleq 1 - \int_{-\infty}^{\infty} \frac{e^{-(\xi-\sigma^2/2)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} \log_2(1 + e^{-\xi}) d\xi. \quad (32) \end{aligned}$$

Following the update of MI in [10], we present the update of the above MI as follows from equations (33) - (36) on the top of the next page.

Given an SNR value (defined as P/σ^2), the performance of iterative decoder can be evaluated as follows. First initialize $I_{dv} = I_{dc} = J(2/\sigma)$ according to (32) and all other MI to be zero, then iteratively update I_{vc} and I_{vp} according to (33) and (34), respectively, and then update I_{cv} and I_{pv} according to (35) and (36), respectively. The decoding is successful if I_{vc} approaches one. The threshold $\mathcal{J}_P^{(1)}(r, \{\lambda_j\})$ is the minimum SNR that ensures successful decoding.

3) Code Profile Optimization Results: We show the code profile optimization results for $\mu = 1$. Fig. 8 shows the gap between the error free decoding SNR (no error decoding for 100 channel realizations) and the Gaussian signaling SNR (28) for both the optimized profile (29) and Luby's profile, for different transmission rates. We fix the number of transmitted symbols to be 10000, and test the performance of the optimized profile for the code rates in $\mathcal{R}_0$ for 4-, 16-, and 64-QAMs. Specifically, we simulate the optimized code profile and Luby's profile for the code rates 0.2, 0.3, 0.4, and 0.5 for 4-QAM, the code rates 0.3, 0.4, and 0.5 for 16-QAM, and the code rates 0.4 and 0.5 for 64-QAM, which correspond to the transmission rates 0.4, 0.6, 0.8, 1.0, 1.2, 1.6, 2.0, 2.4, and 3.0. Fig. 8 shows that the maximal gaps for the optimized profile and Luby's profile are 2.0dB and 3.5dB, corresponding to $\gamma = 0.63$ and $\gamma = 0.45$, respectively.

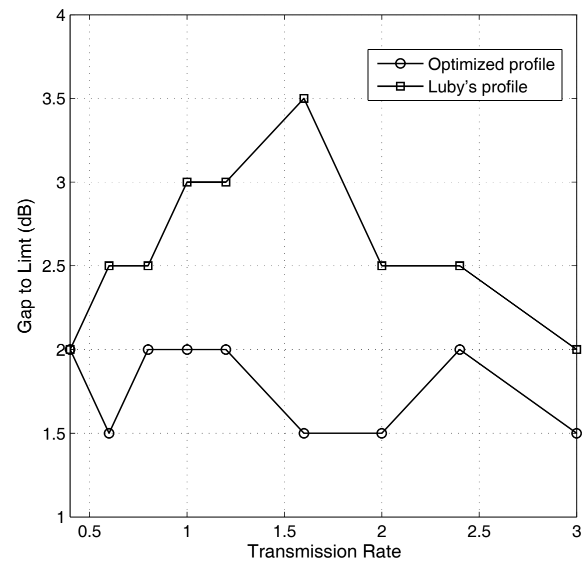


Fig. 8. The gap to Shannon limit for the optimized profile and Luby's profile.

Remark - Code Design for $\mu = 2$***:*** For $\mu = 2$, since one user may be decoded individually or jointly with another user, we need to find a good code profile for both the single-user decoding and the two-user joint decoding. Furthermore, for the two-user joint decoding, we need to find a good profile for the combinations of all channel gains, all modulation schemes, and all code rates for the two users. This makes the profile optimization rather complicated, and the gap of the optimized profile $\gamma = 0.45$ is significantly larger than that ($\gamma = 0.63$) for $\mu = 1$. This is the reason that practical coding schemes with group sizes $\mu = 2$ and 3 perform worse than that with group size $\mu = 1$.

C. Simulation Results for System with Optimized Codes

1) SISO Channel: We use the same setup as that in Section IV-C with group size $\mu = 1$ and employ the codes with the optimized profile and the optimized quantization tables corresponding to the layered and un-layered coding. Fig. 9 shows the corresponding throughput, denoted as "layered, optimum" and "un-layered, optimum", respectively, for the channel SNR P/σ^2 from 0dB to 9dB. The number of transmitted symbols is computed according (23) in Section IV-A3. At each channel SNR, 1000 channel realizations are simulated and the throughput is the total number of information bits

$$I_{vc} = \sum_{i,j} \gamma_i \tilde{\rho}_j J\Big(\sqrt{(j-1)[J^{-1}(I_{cv})]^2 + [J^{-1}(I_{dv})]^2 + i[J^{-1}(I_{pv})]^2}\Big), \quad (33)$$

$$I_{vp} = \sum_{i,j} \tilde{\gamma}_i \rho_j J\Big(\sqrt{j[J^{-1}(I_{cv})]^2 + [J^{-1}(I_{dv})]^2 + (i-1)[J^{-1}(I_{pv})]^2}\Big), \quad (34)$$

$$I_{cv} = 1 - \sum_{j} \widetilde{\lambda}_j J\Big(\sqrt{(j-1)[J^{-1}(1-I_{vc})]^2 + [J^{-1}(1-I_{dc})]^2}\Big), \quad (35)$$

$$I_{pv} = 1 - J\Big(\sqrt{(d_c-1)[J^{-1}(1-I_{vp})]^2}\Big) . \quad (36)$$

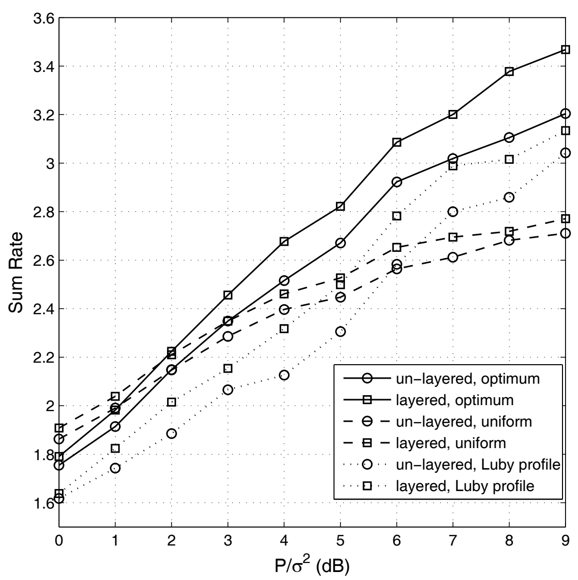


Fig. 9. The simulated throughput with QAM and optimized rateless codes.

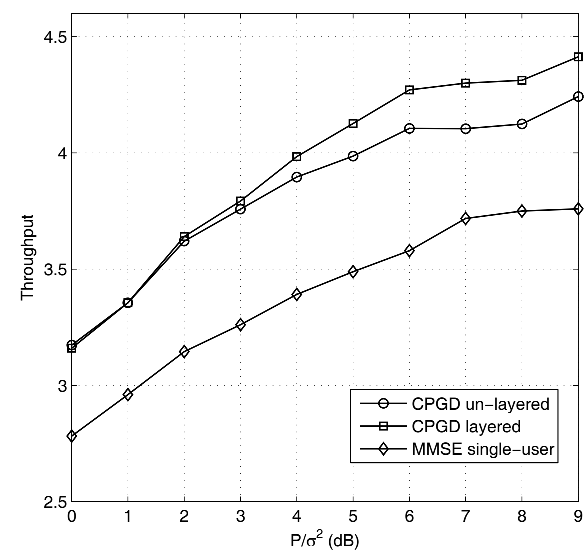


Fig. 10. The simulated throughput in MISO channel with beamforming.

divided by the total number of transmitted symbols. Recall that based on the incremental transmission in Section V-A, for each channel realization the number of transmitted symbols is given by $N_s + b\Delta N$ where b is the number of additional bursts to guarantee successful decoding.

For comparison, we plot the throughput performance of the rate table $\mathcal{T} = \{0.40, 0.80, 1.20, 1.60\}$ for the layered and un-layered coding schemes, denoted as "layered uniform" and "un-layered uniform", respectively, in Fig. 9. It is seen that the optimized $\mathcal{T}$ provides larger throughputs. Furthermore, we show the performance gain brought by the code profile optimization. We employ the same optimized rate quantization table $\mathcal{T}$ and only substitute the optimized profile with Luby's robust soliton distribution profile in [13] (c.f. Table I in [13], pp.2560) given by

$$\Omega(x) = 0.008x + 0.494x^2 + 0.166x^3 + 0.073x^4 + 0.083x^5 + 0.056x^8 + 0.037x^9 + 0.056x^{19} + 0.025x^{65} + 0.003x^{66}. \quad (37)$$

The throughput of Luby's profile for the layered and un-layered coding schemes, denoted as "layered, Luby profile" and "un-layered, Luby profile", respectively, is plotted in Fig. 9. It is seen that the optimized profile significantly outperforms Luby's profile.

2) Extension to MISO Channel with Beamforming: Finally we consider a MISO interference channel where each transmitter employs three antennas and each receiver employs one antenna. The number of transmitter-receiver pairs is $K = 6$. The channel responses between any two antennas is an i.i.d. complex Gaussian random variable $\mathcal{N}_{\mathbb{C}}(0, 1)$. Assume that each user employs a transmission power P. Let $|\mathcal{T}| = 4$.

We compare two schemes: one employs the simple channel matching (CM) beamforming together with the proposed CPGD with $\mu = 1$ and the optimized rate quantization and rateless code; the other scheme employs the linear MMSE beamforming together with the single-user decoding. For the latter, we design the corresponding practical coding scheme. We follow the procedure given in Section IV to find the optimal cutoff rate d_1, which maximizes the sum-rate of all active users where the users with rates smaller than d_1 are inactivated, and then optimize the rate quantization set $\mathcal{T}$. Fine tuning with incremental transmission is also employed with the initial number of transmitted symbols derived based on the same idea as that in Section IV-A3. We also optimize the Raptor code profile according to Section V-B.

Simulations are performed for both schemes at the SNR from 0dB to 9dB each with 1000 channel realizations. The throughput is plotted in Fig. 10 against the channel SNR P/σ^2. The CM beamforming with CPGD with and without layering, denoted as "CPGD layered" and "CPGD un-layered", respectively, exhibits significant performance improvement over MMSE beamforming-based single-user coding scheme denoted as "MMSE single-user". The CM beamforming which brings about stronger interference is more appropriate to the interference decoding since strong interference can be decoded before the decoding of intended messages.

VI. Conclusions

In this paper, we have first proposed a generalization to the Han-Kobayashi coding scheme for the general K-user interference channel, where each user encodes its information into different layers and transmits a superposition of the signals corresponding to these layers. We have also proposed a constrained partial group decoder (CPGD) which successively decodes the different layer groups (which contain both the useful signal and interference) to facilitate the decoding of the useful message. We have provided a distributed algorithm for determining the rate allocation among different layers and the optimal ordered grouping of layers for CPGD. We have further developed a practical design for the proposed system that employs QAM and rateless code. Our results show that the proposed layered coding scheme with CPGD can offer significant sum-rate improvement over the traditional un-layered transmission with single-user decoding, especially in the strong interference scenario. More importantly, it suffices to have a group size of one in CPGD, which has a low complexity to achieve most of the performance gain.

VII. Acknowledgment

This work was supported in part by the U.S. National Science Foundation (NSF) under grant CCF-0726480.

References

[1] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the k-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[2] M. Maddah-Ali, A. Motahari, and A. Khandani, "Communication over MIMO X channels: interference alignment, decomposition, and performance analysis," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3457–3470, Aug. 2008.

[3] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 49–60, Jan. 1981.

[4] N. Prasad and X. Wang, "Outage minimization and rate allocation for the multiuser Gaussian interference channels with successive group decoding," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5540–5557, Dec. 2009.

[5] A. Tajer, N. Prasad, and X. Wang, "Beamforming and rate allocation in MISO cognitive radio networks," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 362–377, Jan. 2010.

[6] A. Motahari and A. Khandani, "To decode the interference or to consider it as noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1274–1283, Mar. 2011.

[7] A. Tajer, N. Prasad, and X. Wang, "Fair rate adaptation in multiuser interference channels," in *Proc. IEEE International Symposium Information Theory*, June 2010.

[8] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.

[9] O. Etesami and A. Shokrollahi, "Raptor codes on binary memoryless symmetric channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2033–2051, May 2006.

[10] C. Gong, G. Yue, and X. Wang, "Analysis and optimization of a rateless coded joint relay system," *IEEE Trans. Wireless Commun.*, vol. 9, no. 3, pp. 1175–1185, Mar. 2010.

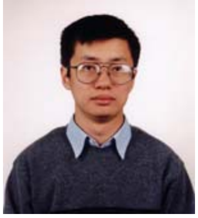
[11] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 670–678, Apr. 2004.

[12] T. Richardson, M. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.

[13] A. Shokrollahi, "Raptor codes," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2551–2567, June 2006.

Chen Gong received the B.S. degree in electrical engineering and mathematics (minor) from Shanghai Jiaotong University, Shanghai, China in 2005, and the M.S. degree in electrical engineering from Tsinghua University, Beijing, China in 2008. He is now pursuing his Ph.D. degree in Columbia University, New York City, New York, USA. His research interests are in the areas of channel coding and modulation techniques for wireless communications.

Ali Tajer received the B.Sc. and M.Sc. degrees in electrical engineering from Sharif University of Technology, Tehran, Iran, in 2002 and 2004, respectively, and the M.Phil. and Ph.D. degrees in electrical engineering from Columbia University, New York, NY in 2010. He is currently a Postdoctoral Research Associate in the Department of Electrical Engineering, Princeton University, Princeton, NJ, and an Adjunct Assistant Professor in the Department of Electrical Engineering, Columbia University, New York, NY. He spent the summers 2008 and 2009 as a Research Assistant with NEC Laboratories America, Princeton, NJ. His research interests lie in the general areas of network information theory and statistical signal processing with applications in wireless communication.



Xiaodong Wang (S'98-M'98-SM'04-F'08) received the Ph.D. degree in electrical engineering from Princeton University. He is a Professor of electrical engineering at Columbia University in New York. Dr. Wang's research interests fall in the general areas of computing, signal processing, and communications, and he has published extensively in these areas. Among his publications is a recent book entitled *Wireless Communication Systems: Advanced Techniques for Signal Reception*, published by Prentice Hall in 2003. His current research interests include wireless communications, statistical signal processing, and genomic signal processing. Dr. Wang received the 1999 NSF CAREER Award and the 2001 IEEE Communications Society and Information Theory Society Joint Paper Award. He has served as an Associate Editor for the IEEE Transactions on Communications, the IEEE Transactions on Wireless Communications, the IEEE Transactions on Signal Processing, and the IEEE Transactions on Information Theory. He is a Fellow of the IEEE.