On the Achievable Throughput of a Multi-Antenna Gaussian Broadcast Channel

Giuseppe Caire and Shlomo Shamai (Shitz)

July 23, 2001

Abstract

A Gaussian broadcast channel with r single-antenna receivers and t antennas at the transmitter is considered. Both transmitter and receivers have perfect knowledge of the channel. Despite apparent simplicity, this model is in general a non-degraded broadcast channel, for which the capacity region is not fully known. We propose a novel transmission scheme based on "ranked known interference" (RKI). In brief, the transmitter decomposes the channel into an ordered (or ranked) set of interference channels for which the interference signal in the i-th channel is a linear combination of the signals transmitted in channels $j < i$. Since all transmitted signals are generated by the transmitter, interference in each channel is known non-causally. Hence, known techniques of coding for non-causally known interference can be applied to make the interference in each channel harmless without further power penalty. We show that the proposed scheme is throughputwise asymptotically optimal for both low and high SNR, and we compare the throughput achievable by RKI with the throughput achievable by more conventional zero-forcing beamforming (space-division multiple access) and with the throughput of a single-user multiple antenna system obtained by allowing the receivers to cooperate. Also, we provide a modification of the basic RKI scheme which achieves optimal throughput for all SNRs in the special case of two users. For independent Rayleigh fading, closed-form throughput expressions are obtained for the basic RKI strategy and arbitrary t and r. Numerical results are shown for finite r, t and in the large-system limit of $r, t \to \infty$ with fixed ratio $\alpha = r/t$ users per transmit antenna.

Keywords: Gaussian broadcast channel, multiple-antenna systems.

1 Introduction

We consider a wireless communication system with one transmitter and r receivers. The transmitter has t antennas, while the receivers have one antenna each. The transmitter has to deliver to each receiver independent information, as in the downlink of a single-cell system where the base-station is equipped with an antenna array of t elements.

We consider the case where the channel transfer function between the transmitter and all receivers is perfectly known to all terminals. Our main results are independent of the statistics of the channel and they can be directly applied to the more general multi-cell downlink scenario with full cooperation between all base-station antennas. However, in this work we do not consider explicitly the effect of the spatial distribution of the transmit and receive antennas, as for example in the cellular model of [1, 2].

This channel is referred to in the following as the $t \times 1 : r$ GBC (to be read "t times 1 to r Gaussian Broadcast Channel "), in order to stress the fact that the r receivers must process their signals separately, as opposed to a $t \times r$ multi-antenna single-user system where the signals at the r receive antennas can be processed jointly [3]. Despite apparent simplicity, this model is in general a non-degraded broadcast channel, for which the capacity region is not fully known.

Conventional approaches to the $t \times 1 : r$ GBC have been proposed by many authors under different assumptions on the channel knowledge at the transmitter. They are all based on some form of *linear precoding* of the user signals. In [4], the multipath channel knowledge at the transmitter is used to pre-filter the signal of each user by its own space-time matched filter, in order to induce at the receiver the *Maximal-Ratio Combining* (MRC) [5] of the multipath without having to implement the matched filter in the user terminal. We refer to this approach as *MRC beamforming.* [1] In [6, 7, 8, 9, 10, 11] joint linear precoders are found in order to maximize the signal-to-interference plus noise ratio at the user terminals subject to various constraints. These schemes, usually proposed in a CDMA multiple-antenna single-cell downlink scenario, are based on linear transformations of the individual user signals which essentially null-out interference while combining the multipath of the useful signal for each user. In our frequency-flat unspread channel model, this reduces to inverting the channel at the transmitter by using the Moore-Penrose pseudoinverse [12] of the channel matrix. This approach shall be referred to in the following as *Zero-Forcing* (ZF) beamforming.

An upper bound on the achievable throughput of the $t \times 1 : r$ GBC is obtained by letting the receivers cooperate. In this case, the channel reduces to a single-user multiple antenna situation, for which the capacity with perfect channel knowledge at both ends is well-known [3, 13]: the channel can be decomposed into a set of parallel channels corresponding to its eigenmodes without loss of information. It is worthwhile to notice that for both ZF beamforming and the cooperative system the optimal coding strategy reduces to standard single-user Gaussian coding, and the multiple-input multiple-output nature of the system is captured by simple linear signal processing at the transmitter only (for ZF) or at both transmitter and receiver (for cooperative).

A quite different coding strategy is proposed here for the $t \times 1 : r$ GBC. Our scheme, referred to as *Ranked Known Interference* (RKI), makes use of linear precoding in order to decompose the channel into a ranked set of interference channels for which the interference signal in the i-th channel is a linear combination of the signals transmitted in channels $j < i$. Since all signals are generated by the same transmitter, interference in

[1] In the CDMA literature this is also known as "pre-rake".

each channel is known non-causally. Then, known techniques of coding for non-causally known interference [14, 15, 16, 17, 18] can be applied to make the interference in each channel harmless without further power penalty. In particular, the recently proposed scheme of [17] based on modulo-lattice precoding can be applied here. Interestingly, a version of our RKI scheme with suboptimal one-dimensional lattice precoding, analogous to Tomlinson-Harashima precoding (see [19] and references therein), has been recently and independently proposed in [20, 21] for the transmitter-based cancellation of far-end cross-talk in DSL.

We show that the RKI strategy is throughputwise asymptotically optimal for both low and high SNR, and it is equivalent to MRC beamforming to the best user only for low SNR and to the cooperative scheme for high SNR. Also, we provide a modification of the basic RKI scheme which achieves optimal throughput for all SNRs in the special case of two users. In the case of Rayleigh fading, closed-form throughput expressions are obtained for the basic RKI strategy for both finite r, t and in the large-system limit of $r, t \to \infty$ with fixed ratio $\alpha = r/t$ users per transmit antenna.

The paper is organized as follows. In Section 2 the channel model is defined. In Section 3 the basic RKI strategy and its properties are illustrated. Section 4 presents upper and lower bounds on the achievable throughput of the $t \times 1 : r$ GBC and Section 5 shows that a modification of the basic RKI strategy is actually optimal for the $t \times 1 : 2$ GBC. In Section 6 results for the basic RKI scheme in independent Rayleigh fading are given. Finally, Section 7 summarizes our conclusions.

Notation.

- Let $\mathbf{A}$ be a matrix. The i-th row, j-th column and (i, j)-th element of $\mathbf{A}$ are denoted by $\mathbf{a}^i$, $\mathbf{a}_j$ and $a_{i,j}$ or equivalently by $[\mathbf{A}]_{i,j}$, respectively.
- The submatrix obtained by the rows of $\mathbf{A}$ numbered by $i \in \mathcal{S}$, where $\mathcal{S}$ is an index set, is denoted by $\mathbf{A}[\mathcal{S}]$.
- Superscripts ^{T} and ^{H} denote transpose and Hermitian transpose, respectively.
- $\mathbf{z} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ indicates that $\mathbf{z}$ is a complex circularly-symmetric Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.
- $[x]_+ \stackrel{\Delta}{=} \max\{0, x\}$.

2 System model

We consider a discrete-time complex baseband channel model and assume that the propagation channel between each transmit-receive antenna pair is frequency non-selective. Extension to the time-invariant finite-memory frequency-selective case is easily obtained in the frequency domain (see [22] and references therein) and will be briefly addressed in Appendix A. The $t \times 1 : r$ GBC is described by

$$\mathbf{y}_i = \mathbf{H}\mathbf{x}_i + \mathbf{z}_i \tag{1}$$

where $\mathbf{x}_i \in \mathbb{C}^t$ is the transmitted symbol vector at time i, $\mathbf{y}_i$ and $\mathbf{z}_i \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I})$ are the corresponding vectors of received signal and noise and $\mathbf{H} \in \mathbb{C}^{r \times t}$ is the channel matrix, where $h_{k,\ell}$ is the complex channel gain from transmit antenna ℓ to the receiver antenna of user k.

The input is constrained to satisfy

$$\frac{1}{n}\sum_{i=1}^{n} |\mathbf{x}_i|^2 \leq A \tag{2}$$

where A is the maximum allowed total transmit energy per channel use. Since we consider normalized unit noise variance, A takes on the meaning of total *transmit* SNR.

For a given block length n, a code $\mathcal{C}_n$ for the input-constrained $t \times 1 : r$ GBC is defined by a codebook of $\exp(n \sum_{k=1}^{r} R_k)$ *code arrays* of the form $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{C}^{t \times n}$, such that (2) is satisfied for each array, by an encoding function ϕ mapping r-tuples of indexes $(w_1, \ldots, w_r)$ (where $w_k \in \{1, 2, \ldots, \exp(nR_k)\}$) onto the code words, and by r decoding functions $\psi_1, \ldots, \psi_r$, such that $\psi_k : \mathbb{C}^n \to \{1, 2, \ldots, \exp(nR_k)\}$. The received signal at the k-th receiver antenna is given by the k-th row $\mathbf{y}^k$ of the array $\mathbf{Y} = \mathbf{HX} + \mathbf{Z}$. The error probability for the k-th user is given by $\epsilon_k = \Pr(\psi_k(\mathbf{y}^k) \neq w_k)$. A r-tuple of rates $\mathbf{R} = (R_1, \ldots, R_r)$ is achievable if there exist a sequence of codes $\mathcal{C}_n$ with rates approaching $\mathbf{R}$ and vanishing ϵ_k (for all $k = 1, \ldots, r$), as n increases. The system *throughput* R, measured in bit/channel use (or bit/s/Hz), is defined as the rate sum

$$R = \sum_{k=1}^{r} R_k \tag{3}$$

We assume perfect *Channel State Information* at the transmitter (CSIT) and at all receivers (CSIR), i.e., the channel matrix $\mathbf{H}$ is known to everybody, and consider the following scenarios:

1. $\mathbf{H}$ is deterministic and fixed.

2. $\mathbf{H}$ is fixed during the transmission of each code array, but it is randomly and independently selected according to a given probability distribution (composite channel). In this case [23, 24], we can consider both the *short-term* constraint (2) or the *long-term* constraint

$$E_{\mathbf{H}}\left[\frac{1}{n}\sum_{i=1}^{n} |\mathbf{x}_i|^2\right] \leq A \tag{4}$$

   where $E_{\mathbf{H}}$ denotes expectation with respect to $\mathbf{H}$.

3. $\mathbf{H}$ is generated by an ergodic matrix random process and varies during the transmission of each code word so that the channel is *information stable* [24, 25].

Since R depends on $\mathbf{H}$, the *instantaneous* throughput of the composite channel (case 2) is a random variable. In this case we consider the average throughput $\bar{R} = E_{\mathbf{H}}[R]$. This is achievable by a variable-rate coding scheme which adjusts its instantaneous throughput according to the channel realization $\mathbf{H}$. By the law of large numbers, the average throughput achievable over a long sequence of channel realizations is given by $\bar{R}$. More in general, the same result applies to the case of a block-fading channel where $\mathbf{H}$ is constant over blocks of n consecutive channel uses and changes from block to block according to an ergodic matrix random process [24].

With perfect CSIT and CSIR, it turns out that the information stable channel (case 3) has the same average throughput of the composite channel with long-term power constraint (4) (although coding and decoding strategies and error exponents for these channel are generally different [24]). Therefore, we shall focus on cases 1 and 2 only.

The $1 \times 1 : r$ GBC coincides with the classical *degraded* Gaussian broadcast channel, whose capacity region is well-known (see [26] in the deterministic case and [27, 28] in the composite or ergodic cases). However, the $t \times 1 : r$ GBC for $t > 1$ is in general a non-degraded broadcast channel, for which the capacity region is not fully known, and cannot be reduced to an equivalent set of parallel degraded broadcast channels (studied in [29, 30, 31, 27, 22, 28]). Obviously, an inner bound to the capacity region is provided by exhibiting an explicit coding scheme for which the error probabilities vanish as the block length increases.

It is interesting to notice here that perfect CSIT is a key fact in our model. For example, under the assumption of no CSIT and of symmetric ergodic channel, where $\mathbf{H}$ is random with independently and identically distributed rows, the marginal transition pdfs $p(y_{k,i}, \mathbf{H}|\mathbf{x}_i)$ are identical for all $k = 1, \ldots, r$, and the resulting $t \times 1 : r$ GBC is stochastically degraded [26] also for $t > 1$. In this case, the capacity region is trivially obtained by time-sharing and the optimal throughput coincides with the $t \times 1$ "transmit diversity" capacity obtained in [3], explicitly given by

$$\bar{R}^{\text{tx-div}} = E[\log(1 + |\mathbf{h}^1|^2 A/t)] \tag{5}$$

We hasten to say that the cases of perfect CSIT and of symmetric channel without CSIT are not the only possibilities. Other settings might also yield new and interesting problems.

3 The RKI coding strategy

Let $\mathbf{H} = \mathbf{G}\mathbf{Q}$ be a QR-type decomposition [12] obtained by applying the Gram-Schmidt orthogonalization procedure [12] to the rows of $\mathbf{H}$. Let $m = \text{rank}(\mathbf{H})$, then $\mathbf{G} \in \mathbb{C}^{r \times m}$ is lower triangular (i.e., it has zeros above its main diagonal) and $\mathbf{Q} \in \mathbb{C}^{m \times t}$ has orthonormal rows. By letting the transmitted i-th signal vector be given by $\mathbf{x}_i = \mathbf{Q}^H \mathbf{u}_i$, the original channel is turned into the set of m interference channels

$$y_{k,i} = g_{k,k} u_{k,i} + \sum_{j<k} g_{k,j} u_{j,i} + z_{k,i} \quad , \qquad k = 1, \ldots, m \tag{6}$$

while no information is sent to users $m+1, \ldots, r$. The input constraint translates directly to the new input signal $\mathbf{u}_i$, in fact,

$$\frac{1}{n}\sum_{i=1}^{n} |\mathbf{x}_i|^2 = \frac{1}{n}\sum_{i=1}^{n} \mathbf{u}_i^H \mathbf{Q}\mathbf{Q}^H \mathbf{u}_i = \frac{1}{n}\sum_{i=1}^{n} |\mathbf{u}_i|^2 \tag{7}$$

The interference signal $s_{k,i} = \sum_{j<k} g_{k,j} u_{j,i}$ in the k-th interference channel (6) is given by a linear combination of the transmitted signals $u_{j,i}$ in channels $j < k$. Since all these signals are generated at the transmitter, and the coefficients $g_{k,j}$ are known, the interference signal in each channel is known non-causally by the encoder (if the channels are considered in the order $1, \ldots, m$ dictated by the QR decomposition). For this reason, we refer to this scheme as *Ranked Known Interference* strategy. We have the following result:

Theorem 1. For given $\mathbf{H} = \mathbf{G}\mathbf{Q}$, with $d_k \stackrel{\Delta}{=} |g_{k,k}|^2$, the maximum throughput achieved by the RKI strategy is given by

$$R^{\text{rki}} = \sum_{k=1}^{m} [\log(\xi d_k)]_+ \tag{8}$$

where ξ is the solution of the waterfilling equation [32]

$$\sum_{k=1}^{m} [\xi - 1/d_k]_+ = A \tag{9}$$

Proof. See Appendix B. □

Corollary 1. For the composite channel we have $\bar{R}^{\text{rki}} = E_{\mathbf{H}}[R^{\text{rki}}]$ where ξ solves the short-term input constraint (9), or the long-term input constraint $E_{\mathbf{H}}\left[\sum_{k=1}^{m}[\xi - 1/d_k]_+\right] = A$. □

Remark: on the capacity of channels with interference known to the transmitter. In Appendix B, we outline a coding scheme based on modulo-lattice precoding, first introduced by Erez Shamai and Zamir [17], which is able to achieve the RKI throughput of Theorem 1 and Corollary 1.

Also, in Appendix B we investigate the relation between the RKI scheme and Costa's "writing on dirty paper" capacity result [15], and we show that the throughput of Theorem 1 and Corollary 1 can be also achieved by a simple generalization of Costa's scheme, which is quite different from the lattice precoding scheme of [17]. The main difference between the two schemes is that Costa's relies on the fact that the known interference signal is Gaussian, i.i.d., with given (known) variance, while lattice precoding is *universal* in the sense that it works for any arbitrary interference signal sequence, provided that it is known non-causally to the transmitter (notice that in this case the transmitter knows the interference realization, but need not know the interference statistic). The link between Costa's "writing on dirty paper" capacity and the lattice precoding scheme is provided in Appendix B by combining a well-known result of Ahlswede on the AVC with random parameters [16] non-causally known at the transmitter with a recent result of Cohen and Lapidoth [18]. ♢

For the sake of comparison, we review here the ZF beamforming and the cooperative schemes. ZF beamforming consists of inverting the channel matrix at the transmitter in order to create orthogonal channels between the transmitter and the receivers without receivers cooperation. Let $\mathcal{S} \subseteq \{1, \ldots, r\}$ be a subset of cardinality $\leq m$ for which the corresponding submatrix $\mathbf{H}[\mathcal{S}]$ is full row-rank, i.e., $\text{rank}(H[\mathcal{S}]) = |\mathcal{S}|$. Let

$$\mathbf{H}_{\mathcal{S}}^{+} = \mathbf{H}[\mathcal{S}]^H \left(\mathbf{H}[\mathcal{S}]\mathbf{H}[\mathcal{S}]^H\right)^{-1} \tag{10}$$

be the Moore-Penrose pseudoinverse [12] of $\mathbf{H}[\mathcal{S}]$. In ZF beamforming, the transmit signal is obtained as $\mathbf{x}_i = \mathbf{H}_{\mathcal{S}}^{+}\mathbf{u}_i$ and yields the set of parallel channels

$$y_{k,i} = u_{k,i} + z_{k,i} \quad , \quad k \in \mathcal{S} \tag{11}$$

while no information is sent to users $k \notin \mathcal{S}$. Clearly, the maximum ZF throughput is obtained by $\mathbf{u}_i$ Gaussian with independent components.

The (short-term) input constraint is given by

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} \text{trace}\left(\mathbf{H}_{\mathcal{S}}^{+}\mathbf{u}_i\mathbf{u}_i^H(\mathbf{H}_{\mathcal{S}}^{+})^H\right) &= \text{trace}\left((\mathbf{H}_{\mathcal{S}}^{+})^H\mathbf{H}_{\mathcal{S}}^{+}\frac{1}{n}\sum_{i=1}^{n}\mathbf{u}_i\mathbf{u}_i^H\right) \\ &\to \sum_{k=1}^{|\mathcal{S}|}\frac{\sigma_{u,k}^2}{b_k} \end{aligned} \tag{12}$$

where we let $E[\mathbf{u}_i\mathbf{u}_i^H] = \text{diag}(\sigma_{u,1}^2, \ldots, \sigma_{u,|\mathcal{S}|}^2)$ and where we define

$$b_k \triangleq \frac{1}{\left[\left(\mathbf{H}[\mathcal{S}]\mathbf{H}[\mathcal{S}]^H\right)^{-1}\right]_{k,k}} \tag{13}$$

The maximum ZF throughput is obtained as the solution of the optimization problem

$$\left\{ \begin{array}{ll} \max & \sum_{k=1}^{|\mathcal{S}|}\log(1+\sigma_{u,k}^2) \\ \text{subject to} & \sum_{k=1}^{|\mathcal{S}|}\frac{\sigma_{u,k}^2}{b_k} = A \end{array} \right.$$

which can be conveniently reparameterized by letting $a_k = \frac{\sigma_{u,k}^2}{b_k}$ and yields

$$R^{\text{zf}} = \sum_{k=1}^{|\mathcal{S}|}\left[\log(\xi b_k)\right]_+ \tag{14}$$

where ξ is the solution of $\sum_{k=1}^{|\mathcal{S}|}[\xi - 1/b_k]_+ = A$.

The above ZF maximum throughput can be further optimized with respect to the choice of the *active user set* $\mathcal{S}$. In particular, by optimizing over the sets of cardinality one, we obtain a system based on MRC beamforming to the user with highest individual channel capacity, whose throughput is given by

$$R^{\rm mrc} = \log\left(1 + |\mathbf{h}^{\max}|^2 A\right) \tag{15}$$

where $\mathbf{h}^{\max}$ is the row of $\mathbf{H}$ with largest 2-norm. Interestingly, for $t > 1$ this choice does not yield necessarily the largest throughput, in sharp contrast to the standard degraded broadcast channel ($t = 1$) for which the throughput is maximized by transmitting to the best user only [27].

Now, we consider a single-user multiple antenna system which is obtained from our system by allowing the r receivers to cooperate [3]. Let $\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^H$ be the *Singular Value Decomposition* [12] of $\mathbf{H}$, where $\mathbf{U} \in \mathbb{C}^{r\times r}$ and $\mathbf{V} \in \mathbb{C}^{t\times t}$ are unitary and $\mathbf{S}$ is diagonal with non-negative diagonal elements, the first m of which are strictly positive and denoted by $\sqrt{c_1} \geq \cdots \geq \sqrt{c_m}$. By letting $\mathbf{x}_i = \mathbf{V}\mathbf{u}_i$ and $\mathbf{v}_i = \mathbf{U}^H\mathbf{y}_i$ the channel (1) is diagonalized as

$$\mathbf{v}_i = \mathbf{S}\mathbf{u}_i + \mathbf{z}'_i \tag{16}$$

where $\mathbf{z}'_i = \mathbf{U}^H\mathbf{z}_i$ has the same statistics of $\mathbf{z}_i$. The cooperative throughput is also maximized by $\mathbf{u}_i$ Gaussian with independent components and it is given by [3]

$$R^{\rm coop} = \sum_{k=1}^{m} [\log(\xi c_k)]_+ \tag{17}$$

where ξ is the solution of $\sum_{k=1}^{|\mathcal{S}|}[\xi - 1/c_k]_+ = A$.

The ZF and cooperative throughputs for the composite channel (with short or long-term power constraints) are immediately obtained from (14) and (17) by taking expectation with respect to $\mathbf{H}$.

Remark: on the user ordering problem. Since for any unitary matrix $\mathbf{Q}$ the matrix $\mathbf{QH}$ has the same singular values of $\mathbf{H}$, $R^{\rm coop}$ is obviously independent of the user ordering (permutation matrices are unitary). On the contrary, $R^{\rm zf}$ depends on the choice of the *unordered* active user set $\mathcal{S}$, and $R^{\rm rki}$ depend on the *ordered* active user set $\mathcal{S}$ (whose rows are considered in order to perform Gram-Schmidt orthogonalization). In order to stress this dependence, we introduce the following notation

$$\begin{aligned} R^{\rm rki-max} &\stackrel{\Delta}{=} \max_{\mathcal{S}} R^{\rm rki} \\ R^{\rm zf-max} &\stackrel{\Delta}{=} \max_{\mathcal{S}} R^{\rm zf} \end{aligned} \tag{18}$$

where in the first line $\mathcal{S}$ ranges over the ordered user sets with cardinality m, and in the second line $\mathcal{S}$ ranges over the unordered user sets with cardinality $|\mathcal{S}| \leq m$.

If $\text{rank}(\mathbf{H}) = m$ then $R^{\rm rki-max}$ is achieved by an ordered set of m users, for every SNR $A \geq 0$. In fact, suppose that for a given A the maximum of (8) is achieved by an ordered set $\mathcal{S}'$ of cardinality $k < m$, such that $\mathbf{H}[\mathcal{S}'] = \mathbf{G}'\mathbf{Q}'$. Then, there exists an ordered set of users $\mathcal{S} = \mathcal{S}' \cup \{i_1, \dots, i_{m-k}\}$ such that $\mathbf{H}[\mathcal{S}] = \mathbf{G}\mathbf{Q}$ where $|g_{i,i}|^2 = |g'_{i,i}|^2$ for $i = 1, \dots, k$ and $|g_{i,i}|^2 > 0$ for $i = k+1, \dots, m$. Therefore, the RKI strategy applied to $\mathbf{H}[\mathcal{S}']$ and to $\mathbf{H}[\mathcal{S}]$ yields the same throughput.

Also, the user ordering is irrelevant for the RKI strategy for asymptotically large SNR if $\mathbf{H}$ is full row-rank, i.e., if $m = r$. In fact, let $\{d_1, \dots, d_r\}$ and $\{d'_1, \dots, d'_r\}$ be the two sets of ordered squared diagonal elements of the matrices $\mathbf{G}$ and $\mathbf{G}'$ in the QR decompositions $\mathbf{H} = \mathbf{G}\mathbf{Q}$ and $\mathbf{\Pi}\mathbf{H} = \mathbf{G}'\mathbf{Q}'$, respectively, where $\mathbf{\Pi}$ is a $r \times r$ permutation matrix. We have that

$$\prod_{i=1}^{r} d_i = |\det(\mathbf{G})|^2 = \det(\mathbf{H}\mathbf{H}^H) = \det(\mathbf{\Pi}\mathbf{H}\mathbf{H}^H\mathbf{\Pi}^H) = |\det(\mathbf{G}')|^2 = \prod_{i=1}^{r} d'_i$$

Define the following arithmetic means

$$M_a \triangleq \frac{1}{r}\sum_{i=1}^{r}\frac{1}{d_i}, \quad M'_a \triangleq \frac{1}{r}\sum_{i=1}^{r}\frac{1}{d'_i}$$

and the geometric mean $M_g \triangleq (\prod_{i=1}^{r} 1/d_i)^{1/r} = (\prod_{i=1}^{r} 1/d'_i)^{1/r}$. There exist $A_0 < \infty$ such that the equation

$$\sum_{i=1}^{r}[\xi - 1/d_i]_+ = A_0$$

has solution $\xi_0 = A_0/r + M_a$ and the equation

$$\sum_{i=1}^{r}[\xi - 1/d'_i]_+ = A_0$$

has solution $\xi'_0 = A_0/r + M'_a$. Then, for all $A \geq A_0$, the RKI throughputs corresponding to the original and permuted row orders are given by $r \log \frac{A/r + M_a}{M_g}$ and by $r \log \frac{A/r + M'_a}{M_g}$, respectively, implying that their difference vanish as $A \to \infty$.

With ZF beamforming, for a given SNR A the maximum throughput $R^{\text{zf-max}}$ might be achieved by a user subset $\mathcal{S}$ of cardinality strictly less than $m = \text{rank}(\mathbf{H})$. However, it is easy to see from the properties of the waterfilling power allocation in (14) that there exists a finite value A_0 (which depends on $\mathbf{H}$) for which, for all $A \geq A_0$, $R^{\text{zf-max}}$ is achieved by a subset of cardinality m. ◊

Since by constraining the receivers to process their signals independently the throughput cannot be increased, R^{coop} upperbounds both $R^{\text{zf-max}}$ and $R^{\text{rki-max}}$. The RKI scheme yields generally a larger maximal throughput than ZF beamforming, as stated in the following:

Theorem 2. For any channel matrix $\mathbf{H}$, $R^{\text{rki-max}} \geq R^{\text{zf-max}}$.

Proof. Assume that, after a suitable row permutation, the first k rows of $\mathbf{H}$ are linearly independent, choose the user subset $\mathcal{S} = \{1, \ldots, k\}$. The columns $\mathbf{v}_i$ of $\mathbf{H}_{\mathcal{S}}^{+}$ satisfy

$$\mathbf{h}^j \mathbf{v}_i = \delta_{i,j} \quad , \quad j = 1, \ldots, k$$

Therefore, $\mathbf{v}_i^H$ must lie in the orthogonal complement of the subspace $\mathcal{V}_i = \text{span}\{\mathbf{h}^j : j = 1, \ldots, k, j \neq i\}$. Let $\mathbf{P}_i^{\perp}$ be the orthogonal projector [12] on $\mathcal{V}_i^{\perp}$. From the above orthonormality condition we get

$$\mathbf{v}_i^H = \frac{\mathbf{h}^i \mathbf{P}_i^{\perp}}{\mathbf{h}^i \mathbf{P}_i^{\perp} (\mathbf{h}^i)^H}$$

The inverse of the i-th diagonal element of $(\mathbf{H}[\mathcal{S}]\mathbf{H}[\mathcal{S}]^H)^{-1} = (\mathbf{H}_{\mathcal{S}}^{+})^H \mathbf{H}_{\mathcal{S}}^{+}$ is given by

$$b_i = \frac{1}{|\mathbf{v}_i|^2} = \frac{|\mathbf{h}^i \mathbf{P}_i^{\perp} (\mathbf{h}^i)^H|^2}{\mathbf{h}^i \mathbf{P}_i^{\perp} (\mathbf{h}^i)^H} = |\mathbf{h}^i \mathbf{P}_i^{\perp}|^2 \tag{19}$$

where we used the fact that orthogonal projectors are idempotents [12]. The rows $\mathbf{q}^i$ of $\mathbf{Q}$ in the QR decomposition $\mathbf{H} = \mathbf{G}\mathbf{Q}$ are obtained by applying Gram-Schmidt orthogonalization to the ordered rows $\mathbf{h}^1, \mathbf{h}^2, \ldots, \mathbf{h}^k$. We obtain

$$\mathbf{h}^i = \sqrt{\mathbf{h}^i \widetilde{\mathbf{P}}_i^{\perp} (\mathbf{h}^i)^H} \mathbf{q}^i + \sum_{j=1}^{i-1} \mathbf{h}^i (\mathbf{q}^j)^H \mathbf{q}^j$$

where $\widetilde{\mathbf{P}}_i^{\perp}$ is the orthogonal projector on the orthogonal complement of $\widetilde{\mathcal{V}}_i = \text{span}\{\mathbf{h}^1, \ldots, \mathbf{h}^{i-1}\}$. From the definition of d_i in (8) and the formula above we obtain

$$d_i = \mathbf{h}^i \widetilde{\mathbf{P}}_i^{\perp} (\mathbf{h}^i)^H = |\mathbf{h}^i \widetilde{\mathbf{P}}_i^{\perp}|^2 \tag{20}$$

Since $\mathcal{V}_i \supset \widetilde{\mathcal{V}}_i$, then $b_i \leq d_i$ for all $i = 1, \ldots, k$. Finally, since both k and the user ordering were arbitrary, this implies that $R^{\text{zf-max}} \leq R^{\text{rki-max}}$. □

We conclude that $R^{\text{coop}} \geq R^{\text{rki-max}} \geq R^{\text{zf-max}}$ holds for any channel matrix. The next result makes this statement stronger in the limits for high and low SNR.

Theorem 3. For any channel matrix $\mathbf{H}$ with full row-rank,

$$\lim_{A \to \infty} \left(R^{\text{coop}} - R^{\text{rki-max}} \right) = 0 \tag{21}$$

For any channel matrix $\mathbf{H}$,

$$\lim_{A \to 0} \frac{R^{\text{rki-max}}}{R^{\text{zf-max}}} = 1 \tag{22}$$

Proof. Consider first (21). Let $c_1, \ldots, c_r$ denote the (non-zero) squared singular values of $\mathbf{H}$ and $d_1, \ldots, d_r$ the (non-zero) squared diagonal elements of $\mathbf{G}$ in the QR decomposition $\mathbf{H} = \mathbf{GQ}$. Define the following arithmetic means

$$M_a \triangleq \frac{1}{r}\sum_{i=1}^{r}\frac{1}{c_i}, \quad \widetilde{M}_a \triangleq \frac{1}{r}\sum_{i=1}^{r}\frac{1}{d_i}$$

and the geometric mean $M_g \triangleq (\prod_{i=1}^{r} 1/c_i)^{1/r} = (\prod_{i=1}^{r} 1/d_i)^{1/r}$, where the last equality follows from

$$\prod_{i=1}^{r} c_i = \det(\mathbf{H}\mathbf{H}^H) = |\det(\mathbf{G})|^2 = \prod_{i=1}^{r} d_i$$

It is immediate to see that there exist $A_0 < \infty$ such that the equation

$$\sum_{i=1}^{r}[\xi - 1/c_i]_+ = A_0$$

has solution $\xi_0 = A_0/r + M_a$ and the equation

$$\sum_{i=1}^{r}[\xi - 1/d_i]_+ = A_0$$

has solution $\widetilde{\xi}_0 = A_0/r + \widetilde{M}_a$. Then, for all $A \geq A_0$, the maximum throughputs can be written as

$$\begin{aligned} R^{\text{coop}} &= r\log\frac{A/r + M_a}{M_g} \\ R^{\text{rki}} &= r\log\frac{A/r + \widetilde{M}_a}{M_g} \end{aligned} \quad (23)$$

By substituting these expressions in the limit (21) we obtain

$$\lim_{A\to\infty} r\log\frac{1 + rM_a/A}{1 + r\widetilde{M}_a/A} = 0$$

In order to show (22), consider first the case where there is a single row $\mathbf{h}^{\max}$ in $\mathbf{H}$ of maximum squared Euclidean norm. We notice both RKI and ZF achieve the MRC throughput $R^{\text{mrc}} = \log(1 + |\mathbf{h}^{\max}|^2 A)$, by choosing an active user set containing only the user corresponding to the row $\mathbf{h}^{\max}$. Let $\mathcal{S}$ denotes an arbitrary user subset of cardinality k for which $\mathbf{H}[\mathcal{S}]$ has rank k, let

$$b_i(\mathcal{S}) = \frac{1}{\left[(\mathbf{H}[\mathcal{S}]\mathbf{H}[\mathcal{S}]^H)^{-1}\right]_{i,i}}$$

and let $R^{\mathrm{zf}}(\mathcal{S})$ denote the maximum of $\sum_{i=1}^{k} \log(1 + b_i(\mathcal{S})a_i)$ subject to $\sum_{i=1}^{k} a_i \leq A$, $a_i \geq 0$. There exists $A_1(\mathcal{S}) > 0$ such that, for all $A \leq A_1(\mathcal{S})$,

$$R^{\mathrm{zf}}(\mathcal{S}) = \log(1 + \max_i \{b_i(\mathcal{S})\}A)$$

Hence, by definition of $\mathbf{h}^{\max}$, for all $A \leq A_1(\mathcal{S})$ we have $R^{\mathrm{zf}}(\mathcal{S}) \leq R^{\mathrm{mrc}}$. By considering all possible subsets $\mathcal{S}$ of cardinality $k = 1, \ldots, m$ we conclude that $R^{\mathrm{zf-max}} = R^{\mathrm{mrc}}$ for $0 < A \leq \min_{\mathcal{S}} A_1(\mathcal{S})$.

Similarly, consider an ordered set $\mathcal{S}$ of cardinality m, let $d_1(\mathcal{S}), \ldots, d_m(\mathcal{S})$ denote the squared diagonal elements of $\mathbf{G}$ in the QR decomposition $\mathbf{H}[\mathcal{S}] = \mathbf{GQ}$ and let $R^{\mathrm{rki}}(\mathcal{S})$ denote the maximum of $\sum_{i=1}^{m} \log(1 + d_i(\mathcal{S})a_i)$ subject to $\sum_{i=1}^{m} a_i = A$. There exists $A_2(\mathcal{S}) > 0$ such that, for all $A \leq A_2(\mathcal{S})$,

$$R^{\mathrm{rki}}(\mathcal{S}) = \log(1 + \max_i \{d_i(\mathcal{S})\}A)$$

Hence, by definition of $\mathbf{h}^{\max}$, for all $A \leq A_2(\mathcal{S})$ we have $R^{\mathrm{rki}}(\mathcal{S}) \leq R^{\mathrm{mrc}}$. By considering all possible such subsets $\mathcal{S}$ we conclude that $R^{\mathrm{rki-max}} = R^{\mathrm{mrc}}$ for $0 < A \leq \min_{\mathcal{S}} A_2(\mathcal{S})$. Then, there exists an $A_3 > 0$ such that for $A \in [0, A_3]$ we have $R^{\mathrm{rki-max}} = R^{\mathrm{zf-max}} = R^{\mathrm{mrc}}$.

In the case where $\mathbf{H}$ has more than one row with maximum squared Euclidean norm we have to distinguish the case where there exists a subset of mutually orthogonal rows with maximal norm from the case where any subset of the maximal norm rows is mutually non-orthogonal. In the latter case, the above proof still holds, and the MRC throughput can be obviously achieved by transmitting to anyone of the users corresponding to the maximal norm rows. In the former case, it is not difficult to show that there exists $A_4 > 0$ for which for every $A \in [0, A_4]$ both the ZF and the RKI throughputs are maximized by transmitting with equal power to the users corresponding to the subset of mutually orthogonal rows with maximal norm. This concludes the proof. □

4 Throughput bounds for the $t \times 1 : r$ GBC

In this section we consider $\mathbf{H}$ deterministic and fixed and we find upper and lower bounds to the maximum achievable throughput R of the corresponding $t \times 1 : r$ GBC. These allow us to establish the asymptotic optimality of the RKI scheme for high and low SNR provided that the channel matrix has full row-rank. For finite and non-vanishing SNR these bounds are generally difficult to evaluate explicitly, however, they will be useful to prove the maximum throughput for the $t \times 1 : 2$ GBC (two-users case) in Section 5.

An upperbound on R is obtained by noticing that the capacity region of a general broadcast channel depends only on the marginal transition probabilities $\{p(y_k|\mathbf{x}) : k = 1, \ldots, r\}$ and not on the joint transition probability $p(\mathbf{y}|\mathbf{x})$ [33, 26]. In our case, the marginal transition pdfs are given by

$$p(y_k|\mathbf{x}) = \frac{1}{\pi} e^{-|y_k - \mathbf{h}^k \mathbf{x}|^2}, \quad k = 1, \ldots, r$$

Any set of marginal transition pdfs

$$p'(y_k|\mathbf{x}) = \frac{1}{\pi \nu_k} e^{-|y_k - \mathbf{h}^k \mathbf{x}|^2/\nu_k}, \quad k = 1, \ldots, r$$

with $\nu_k \leq 1$ yields a GBC capacity region containing that of the original GBC, since any user k in the new channel can emulate artificially the corresponding output of the original channel by adding independent Gaussian noise with variance $1 - \nu_k$. This implies that the channels in the family (1) for given $\mathbf{H}$ and with $\mathbf{z} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \boldsymbol{\Sigma}_z)$, where $\boldsymbol{\Sigma}_z$ is any non-negative definite Hermitian matrix whose diagonal elements are not larger than 1 (we refer to this constraint as the *sub-unit diagonal constraint*), have all broadcast capacity region containing the region of the original GBC (and therefore throughput not smaller than R). In order to tighten the cooperative throughput bound, we can choose the worst-case (cooperative throughput-wise) channel in this family. By using the general capacity formula for the cooperative throughput [3], we have:

Lemma 1. For any channel matrix $\mathbf{H}$,

$$R \leq \max_{\boldsymbol{\Sigma}_x} \min_{\boldsymbol{\Sigma}_z} \log \frac{\det\left(\mathbf{H}\boldsymbol{\Sigma}_x\mathbf{H}^H + \boldsymbol{\Sigma}_z\right)}{\det \boldsymbol{\Sigma}_z} \tag{24}$$

where minimization is over all noise covariances $\boldsymbol{\Sigma}_z$ satisfying the sub-unit diagonal constraint and maximization is over all input covariance matrices $\boldsymbol{\Sigma}_x$ satisfying $\text{trace}(\boldsymbol{\Sigma}_x) = A$. □

From Theorem 3 and Lemma 1 we can prove the following

Theorem 4. If $\mathbf{H}$ has full row-rank, then

$$\lim_{A \to \infty} (R - R^{\text{rki-max}}) = 0 \tag{25}$$

and

$$\lim_{A \to 0} R/R^{\text{rki-max}} = 1 \tag{26}$$

Proof. It is clear that $R^{\text{rki-max}}$ and R^{coop} are a lower and an upper bound on R. The first statement follows directly from the first part of Theorem 3, since $R^{\text{rki-max}} \leq R \leq R^{\text{coop}}$ and $\lim_{A\to\infty}(R^{\text{coop}} - R^{\text{rki-max}}) = 0$ imply the statement.

In order to prove the second statement, let $\boldsymbol{\Pi}$ be the $r \times r$ permutation matrix which sorts the rows of $\mathbf{H}$ such that $|\mathbf{h}^1| \geq \cdots \geq |\mathbf{h}^r|$, and consider the QR decomposition $\boldsymbol{\Pi}\mathbf{H} = \mathbf{G}\mathbf{Q}$. We apply Lemma 1 by choosing as noise covariance $\boldsymbol{\Sigma}_z = \mathbf{G}\mathbf{D}^{-2}\mathbf{G}^H$, where $\mathbf{D} = \text{diag}(|\mathbf{h}^1|, \ldots, |\mathbf{h}^r|)$. By construction, $\boldsymbol{\Sigma}_z$ is positive definite (recall that $\mathbf{H}$ has rank r) and satisfies the sub-unit diagonal constraint, in fact,

$$[\boldsymbol{\Sigma}_z]_{k,k} = \sum_{j=1}^{k} \frac{|g_{k,j}|^2}{|\mathbf{h}^j|^2} \leq \sum_{j=1}^{k} \frac{|g_{k,j}|^2}{|\mathbf{h}^k|^2} = 1$$

With this choice, the RHS in (24) becomes

$$\log \frac{\det\left(\mathbf{G}\mathbf{Q}\boldsymbol{\Sigma}_x\mathbf{Q}^H\mathbf{G}^H + \mathbf{G}\mathbf{D}^{-2}\mathbf{G}^H\right)}{\det\mathbf{G}\mathbf{D}^{-2}\mathbf{G}^H} = \log \frac{\det\left(\mathbf{Q}\boldsymbol{\Sigma}_x\mathbf{Q}^H + \mathbf{D}^{-2}\right)}{\det\mathbf{D}^{-2}}$$

From Hadamard inequality [26] we obtain the maximizing signal covariance in the form $\boldsymbol{\Sigma}_x = \mathbf{Q}^H\text{diag}(a_1,\ldots,a_r)\mathbf{Q}$, which yields the bound

$$R \leq \sum_{k=1}^{r}[\log(\xi a_k)]_+ \tag{27}$$

where ξ is the solution of $\sum_{k=1}^{r}[\xi - 1/|\mathbf{h}_k|^2]_+ = A$.

If $|\mathbf{h}^1| > |\mathbf{h}^2|$, then there exists a value $0 < A_1 < \infty$ such that, for all $A \leq A_1$, the RHS in (27) is equal to $\log(1+|\mathbf{h}^1|^2A) = R^{\text{mrc}}$. This is clearly achievable by RKI and by ZF, therefore for $A \in [0, A_1]$ the RKI (and ZF) strategy is optimal (not only asymptotically for $A \to 0$). If there exist $\kappa > 1$ rows with maximal 2-norm $|\mathbf{h}^{\max}|$, then there exists a value $0 < A_2 < \infty$ such that, for all $A \leq A_2$, the RHS in (27) is equal to $\kappa\log(1+|\mathbf{h}^{\max}|^2A/\kappa)$. In this case,

$$\lim_{A\to 0} \frac{\kappa\log(1+|\mathbf{h}^{\max}|^2A/\kappa)}{\log(1+|\mathbf{h}^{\max}|^2A)} = 1$$

and the statement of Theorem 4 still holds (but only in the limit for vanishing A). □

Remark: downlink strategies. Theorem 4 shows an interesting feature of the $t\times 1 : r$ GBC and of the RKI strategy, which might have a relevant impact on the design of the downlink of wireless communication systems. If the base-station is strongly power limited, then the throughput-maximizing strategy consists of MRC beamforming to the best user, which is the same optimal strategy for the standard degraded GBC ($t = 1$). In this case, the transmit antenna array is used to enhance the received SNR of the best user but does not expand the useful dimensions for transmission. Practical downlink protocols for high-rate packet communications are currently proposed and implemented according to this principle: only one user in each time-slot is served according to a channel-driven scheduling allocating the channel to the user enjoying the instantaneous highest individual capacity [34, 35]. [2]

On the contrary, if the base-station can transmit at large power, the same throughput of a single-user multiple-antenna system can be approached even if the receivers cannot cooperate. In particular, by letting the number of served users per time-slot equal to the number of transmit antennas, under mild conditions on the channel matrix statistics the slope of the throughput as a function of SNR in dB is equal to t. Hence, in the GBC setting, the "capacity boost" typical of multiple-antenna systems depends strongly on the available transmit power. Notice also that the same throughput slopes for low and

[2] In practice, the scheduler must maximize the cell throughput subject to some fairness constraint [34, 35], therefore the channel allocation rule differs from the simple "best user" rule.

high SNR are obtained by the conventional ZF beamforming, although this is generally asymptotically suboptimal for high SNR. $\Diamond$

Assume that $\mathbf{H}$ has rank $m \leq r$ and, after a suitable row permutation $\mathbf{\Pi}$, can be partitioned as

$$\mathbf{\Pi H} = \left[\begin{array}{c} \mathbf{H}_1 \\ \mathbf{H}_2 \end{array} \right] \tag{28}$$

where $\mathbf{H}_1 \in \mathbb{C}^{m \times t}$ has rank m and $\mathbf{H}_2 \in \mathbb{C}^{(r-m)\times t}$. Any row of $\mathbf{H}_2$ can be expressed as a linear combination of rows of $\mathbf{H}_1$, i.e., we can write

$$\mathbf{H}_2 = \mathbf{B}\mathbf{H}_1$$

where $\mathbf{B} = \mathbf{H}_2\mathbf{H}_1^+$. The noise vector is also partitioned as $\mathbf{z} = [\mathbf{z}_1^T, \mathbf{z}_2^T]^T$, where $\mathbf{\Sigma}_{z1}$ and $\mathbf{\Sigma}_{z2}$ are the $m \times m$ upper left and $(r-m) \times (r-m)$ lower right diagonal blocks of $\mathbf{\Sigma}_z$ (both $\mathbf{\Sigma}_{z1}$ and $\mathbf{\Sigma}_{z2}$ must satisfy the sub-unit diagonal constraint). Then, we have the following:

Lemma 2. If $\mathbf{\Sigma}_{z2} - \mathbf{B}\mathbf{\Sigma}_{z1}\mathbf{B}^H$ is non-negative definite, then $R \leq R_1^{\rm coop}$, where $R_1^{\rm coop}$ is the cooperative throughput of the $t \times m$ channel $\mathbf{y}_1 = \mathbf{H}_1\mathbf{x} + \mathbf{z}_1$.

Proof. Define the auxiliary channel with input $\mathbf{y}_1$ and output

$$\mathbf{y}_2 = \mathbf{B}\mathbf{y}_1 + \boldsymbol{\nu}$$

where $\boldsymbol{\nu} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{\Sigma}_\nu)$, independent of $\mathbf{z}_1$ and with $\mathbf{\Sigma}_\nu = \mathbf{\Sigma}_{z2} - \mathbf{B}\mathbf{\Sigma}_{z1}\mathbf{B}^H$, which by assumption is a valid covariance matrix. Among all channels with assigned marginal transition pdfs, there exists the "cascade" channel with joint transition pdf $p(\mathbf{y}_1, \mathbf{y}_2|\mathbf{x}) = p'(\mathbf{y}_2|\mathbf{y}_1)p(\mathbf{y}_1|\mathbf{x})$, where

$$\begin{array}{rcl} p(\mathbf{y}_1|\mathbf{x}) & = & \mathcal{N}_{\mathbb{C}}(\mathbf{H}_1\mathbf{x}, \mathbf{\Sigma}_{z1}) \\ p'(\mathbf{y}_2|\mathbf{y}_1) & = & \mathcal{N}_{\mathbb{C}}(\mathbf{B}\mathbf{y}_1, \mathbf{\Sigma}_\nu) \end{array}$$

For the cascade channel, $\mathbf{x} \rightarrow \mathbf{y}_1 \rightarrow \mathbf{y}_2$ is a Markov chain and we have

$$I(\mathbf{x}; \mathbf{y}_1, \mathbf{y}_2) = I(\mathbf{x}; \mathbf{y}_1) + I(\mathbf{x}; \mathbf{y}_2|\mathbf{y}_1) = I(\mathbf{x}; \mathbf{y}_1)$$

By maximizing $I(\mathbf{x}; \mathbf{y}_1)$ with respect to the input distribution (subject to the input constraint), we get by definition $\max_{\mathbf{x}:{\rm trace}(\mathbf{\Sigma}_x)=A} I(\mathbf{x}; \mathbf{y}_1) = R_1^{\rm coop}$. $\square$

Lemma 2 is actually "included" in Lemma 1, since $R_1^{\rm coop}$ is achieved by a particular choice of $\mathbf{\Sigma}_z$ in the RHS of (24). However, it puts in evidence the interesting fact that, under certain conditions on $\mathbf{H}$, the RKI strategy is asymptotically optimal for high SNR even if the channel has rank $m < r$. In particular, we have the following

Corollary 2. Let $\mathbf{H}$ have rank $m \leq r$ and, after a suitable row permutation, assume that

$$\mathbf{\Pi H} = \left[\begin{array}{c} \mathbf{H}_1 \\ \mathbf{B H}_1 \end{array} \right] \tag{29}$$

where $\mathbf{H}_1 \in \mathbb{C}^{m \times t}$ has rank m and $\|\mathbf{B}\|^2 \leq 1$. Then, the RKI scheme is asymptotically optimal for the $t \times 1 : r$ GBC in the limit for large SNR, i.e., $\lim_{A \to \infty}(R - R^{\mathrm{rki-max}}) = 0$.

Proof. We apply Lemma 2 with the choice $\mathbf{\Sigma}_{z1} = \mathbf{I}_m$ and $\mathbf{\Sigma}_{z2} = \mathbf{I}_{r-m}$. If $\|\mathbf{B}\|_2 \leq 1$, then $\mathbf{I}_{r-m} - \mathbf{B}\mathbf{B}^H$ is non-negative definite and, by Lemma 2, $R \leq R_1^{\mathrm{coop}}$, the cooperative throughput of the $t \times m$ subchannel defined by $\mathbf{H}_1$. Let $R_1^{\mathrm{rki-max}}$ be the maximum RKI throughput for this subchannel. Since $\mathbf{H}_1$ is full row-rank, by Theorem 3 we have that $\lim_{A \to \infty}(R_1^{\mathrm{coop}} - R_1^{\mathrm{rki-max}}) = 0$. Since $R_1^{\mathrm{rki-max}} \leq R^{\mathrm{rki-max}} \leq R$, this implies that $\lim_{A \to \infty}(R - R^{\mathrm{rki-max}}) = 0$, as desired. □

The decomposition (29) (if it exists) is essentially unique, as stated by the following result in linear algebra (new up to the authors' knowledge):

Lemma 3. Let $\mathbf{H} \in \mathbb{C}^{r \times t}$ have rank m and assume that, after a suitable row permutation, it can be decomposed into the submatrices $\mathbf{H}_1 \in \mathbb{C}^{m \times t}$ of rank m and $\mathbf{H}_2 = \mathbf{B}\mathbf{H}_1$ as in (29), with $\|\mathbf{B}\|_2 \leq 1$. Then, this decomposition is unique , in the sense that for any $\mathbf{H}_1' \in \mathbb{C}^{m \times t}$ and $\mathbf{H}_2'$ obtained by exchanging some rows of $\mathbf{H}_1$ with some rows of $\mathbf{H}_2$ such that $\mathrm{rank}(\mathbf{H}_1') = m$, we have $\|\mathbf{H}_2'(\mathbf{H}_1')^+\|_2 \geq 1$.

Proof. See Appendix C. □

Establishing if a given matrix $\mathbf{H}$ admits the decomposition (29) is a problem of independent interest which can be formulated as follows: given a set of vectors $\mathcal{S} = \{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ in $\mathbb{C}^t$, spanning an $m \leq \min\{r, t\}$ dimensional subspace, find (if it exists) a set $\mathcal{S}'$ of m linearly independent vectors such that all other vectors in $\mathcal{S}$ can be written as linear combinations of the vectors in $\mathcal{S}'$ and such that the matrix of combination coefficients has 2-norm not larger than 1.

The following simple example shows that there exist matrices for which decomposition (29) is not possible. For these channels, we cannot claim that the RKI strategy is asymptotically optimal.

Example. Let $\mathbf{H} \in \mathbb{C}^{3 \times t}$ of rank 2, and assume that $\mathbf{h}^3 = \beta_1 \mathbf{h}^1 + \beta_2 \mathbf{h}^2$. If $|\beta_1|^2 + |\beta_2|^2 > 1$ and $|\beta_2|^2 < |\beta_1|^2 < 1 + |\beta_2|^2$ or $|\beta_1|^2 < |\beta_2|^2 < 1 + |\beta_1|^2$, it is not possible to express any of the rows $\mathbf{h}^1$, $\mathbf{h}^2$ and $\mathbf{h}^3$ as a linear combination of the other two with coefficients β' and β'' such that $|\beta'|^2 + |\beta''|^2 \leq 1$. The set of coefficients β_1, β_2 satisfying the above condition is clearly non-empty, so, such matrices exist. ◊

Modified RKI strategy. In order to tighten the RKI throughput lower bound we can consider the following modified RKI strategy. Let $\mathbf{H} = \mathbf{G}\mathbf{Q}$ and construct the transmitted signal as $\mathbf{x}_i = \mathbf{Q}^H \mathbf{R} \mathbf{u}_i$ where $\mathbf{R}$ is a $m \times m$ upper triangular matrix satisfying trace($\mathbf{R}\mathbf{R}^H$) $\leq A$ and where $E[\mathbf{u}_i\mathbf{u}_i^H] = \mathbf{I}_m$. This yields the set of m interference channels

$$y_{k,i} = w_{k,k}u_{k,i} + \sum_{j<k} w_{k,j}u_{j,i} + \sum_{j>k} w_{k,j}u_{j,i} + z_{k,i} \quad , \quad k = 1, \ldots, m \tag{30}$$

where $\mathbf{W}$ denotes the upper $m \times m$ block of $\mathbf{GR}$, and where no information is sent to users $k = m+1, \ldots, r$. The encoder considers the interference signal $\sum_{j<k} w_{k,j}u_{j,i}$ caused by users $j < k$ as known non-causally and the k-th decoder treats the interference signal $\sum_{j>k} w_{k,j}u_{j,i}$ caused by users $j > k$ as additional noise. By applying a coding for known interference strategy (e.g., the lattice precoding scheme of [17]) and by using minimum Euclidean distance decoding at each k-th receiver, from the results of Appendix B and the results of [36] it is not hard to show that the following throughput is achievable

$$R^{\mathrm{mod-rki}} = \sum_{k=1}^{m} \log\left(1 + \frac{|w_{k,k}|^2}{1 + \sum_{j>k} |w_{k,j}|^2}\right) \tag{31}$$

This can be further maximized over all matrices $\mathbf{R}$ satisfying the trace constraint and over all ordered user subsets $\mathcal{S}$ of size m. The maximum throughput of the modified RKI strategy provides a lower bound to R generally tighter than the basic RKI strategy, since the former reduces to the latter by constraining $\mathbf{R}$ to be diagonal.

5 The optimal throughput of the $t \times 1 : 2$ GBC

The simplest non-trivial (i.e., non-degraded) GBC with multiple transmit antennas is the two-user case. For this channel, we have the following closed-form result:

Theorem 5. The maximum achievable throughput of the $t \times 1 : 2$ GBC is given by

$$R = \begin{cases} \log(1 + |\mathbf{h}^1|^2 A) & A \leq A_1 \\ \log \frac{\left(A\det(\mathbf{H}\mathbf{H}^H) + \mathrm{trace}(\mathbf{H}\mathbf{H}^H)\right)^2 - 4|\mathbf{h}^2(\mathbf{h}^1)^H|^2}{4\det(\mathbf{H}\mathbf{H}^H)} & A > A_1 \end{cases} \tag{32}$$

where without loss of generality we assume $|\mathbf{h}^1| \geq |\mathbf{h}^2|$ and where

$$A_1 = \frac{|\mathbf{h}^1|^2 - |\mathbf{h}^2|^2}{\det(\mathbf{H}\mathbf{H}^H)}$$

Proof. Assume $|\mathbf{h}^1| \geq |\mathbf{h}^2|$. The case of rank($\mathbf{H}$) $= 1$ is trivial, since in this case the two rows of $\mathbf{H}$ are linearly dependent, then, the $t \times 1 : 2$ GBC reduces to a standard degraded GBC with input $x = \mathbf{h}^1\mathbf{x}$ and outputs

$$y_1 = x + z_1, \quad y_2 = (|\mathbf{h}^2|/|\mathbf{h}^1|)x + z_2$$

The throughput is clearly maximized by transmitting to the best user only [27], i.e., to user 1, and is given by $R = \log(1 + |\mathbf{h}^1|^2 A)$, which coincides with the first line in (32) since in this case $A_1 = +\infty$ and only the first line in (32) is relevant, therefore, Theorem 5 holds in the rank 1 case.

Next, we consider the case of $\mathbf{H}$ of rank 2. We shall use Lemma 1 to find an upperbound and the modified RKI strategy to find a lowerbound, and show that these bounds coincide with (32).

Cooperative throughput upperbound. A looser version of Lemma 1 yields the upperbound

$$R \leq \min_{\mathbf{\Sigma}_z \in \mathcal{U}} \max_{\mathbf{\Sigma}_x \in \mathcal{A}} \log \frac{\det\left(\mathbf{H}\mathbf{\Sigma}_x\mathbf{H}^H + \mathbf{\Sigma}_z\right)}{\det \mathbf{\Sigma}_z} \tag{33}$$

where $\mathcal{A}$ is the set of all $t \times t$ non-negative definite covariance matrices with trace $\leq A$, and $\mathcal{U}$ is the set of all positive definite covariance matrices satisfying the unit-diagonal constraint (diagonal elements strictly equal to 1). Consider first the maximization of mutual information with respect to $\mathbf{\Sigma}_x$ for a given $\mathbf{\Sigma}_z$. By letting $\mathbf{\Sigma}_z = \mathbf{U}\mathbf{\Lambda}_z\mathbf{U}^H$, with $\mathbf{U}$ unitary and $\mathbf{\Lambda}_z$ diagonal, we obtain the equivalent problem

$$\max_{\mathbf{\Sigma}_x \in \mathcal{A}} \log\det\left(\mathbf{H}_z\mathbf{\Sigma}_x\mathbf{H}_z^H + \mathbf{I}\right)$$

where $\mathbf{H}_z = \mathbf{\Lambda}_z^{-1/2}\mathbf{U}^H\mathbf{H}$. More explicitly, since

$$\mathbf{\Sigma}_z = \begin{bmatrix} 1 & \rho \\ \rho^* & 1 \end{bmatrix}$$

we obtain

$$\mathbf{U} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ \theta & -\theta \end{bmatrix}, \quad \mathbf{\Lambda}_z = \begin{bmatrix} 1+r & 0 \\ 0 & 1-r \end{bmatrix}$$

where we let $r = |\rho|$ and $\theta = \exp(-j\angle\rho)$.

Let λ_1 and λ_2 denote the eigenvalues of $\mathbf{H}_z\mathbf{H}_z^H$. Then, the maximization with respect to $\mathbf{\Sigma}_x \in \mathcal{A}$ yields the function of r, θ and A

$$f(r,\theta,A) = \sum_{i=1}^{r} [\log(\xi\lambda_i)]_+ \tag{34}$$

where ξ solves the equation

$$\sum_{i=1}^{2}\left[\xi - \frac{1}{\lambda_i}\right]_+ = A \tag{35}$$

Explicitly, we have

$$\mathbf{H}_z\mathbf{H}_z^H = \frac{1}{2}\begin{bmatrix} \frac{h^+ + 2\sigma}{1+r} & \frac{h^- + j2\omega}{\sqrt{1-r^2}} \\ \frac{h^- - j2\omega}{\sqrt{1-r^2}} & \frac{h^+ - 2\sigma}{1-r} \end{bmatrix}$$

where we define $h^+ = |\mathbf{h}^1|^2 + |\mathbf{h}^2|^2$, $h^- = |\mathbf{h}^1|^2 - |\mathbf{h}^2|^2$ and $\theta\mathbf{h}^2(\mathbf{h}^1)^H = \sigma + j\omega$. The eigenvalues $\lambda_{1,2}$ are given by

$$\lambda_{1,2} = \frac{1}{2}\left(T \pm \sqrt{T^2 - 4D}\right)$$

with $T = \text{trace}(\mathbf{H}_z\mathbf{H}_z^H)$ and $D = \det(\mathbf{H}_z\mathbf{H}_z^H)$. It is immediate to check that D is independent of θ, while T is a decreasing function of σ. We conclude that θ minimizing capacity is given by $\theta^\star = \exp(-j\angle\mathbf{h}^2(\mathbf{h}^1)^H)$, so that $\sigma = |\mathbf{h}^2(\mathbf{h}^1)^H|$ is maximum. With this choice, we have

$$T = \frac{h^+ - 2r|\mathbf{h}^2(\mathbf{h}^1)^H|}{1-r^2}, \quad D = \frac{|\mathbf{h}^1|^2|\mathbf{h}^2|^2 - |\mathbf{h}^2(\mathbf{h}^1)^H|^2}{1-r^2}$$

In order to obtain the eigenvalues in a convenient form, it is useful to represent the rows $\mathbf{h}^1$ and $\mathbf{h}^2$ in an orthonormal basis. Applying Gram-Schmidt orthogonalization to $\mathbf{h}^1$ and $\mathbf{h}^2$ (in the order) we obtain $\mathbf{h}^1 = g_{1,1}\mathbf{q}^1$, $\mathbf{h}^2 = g_{2,1}\mathbf{q}^1 + g_{2,2}\mathbf{q}^2$, and where $\mathbf{q}^1$ and $\mathbf{q}^2$ (the rows of $\mathbf{Q}$) are orthonormal. The explicit expression of the eigenvalues is now given by

$$\begin{aligned}\lambda_{1,2} &= \frac{1}{2(1-r^2)}\Big[|g_{1,1}|^2 + |g_{2,1}|^2 + |g_{2,2}|^2 - 2r\sigma \\ &\pm \Big((|g_{1,1}|^2 - |g_{2,1}|^2 - |g_{2,2}|^2)^2 - 4r\sigma(|g_{1,1}|^2 + |g_{2,1}|^2 + |g_{2,2}|^2) \\ &+ 4r^2|g_{1,1}|^2(|g_{2,1}|^2 + |g_{2,2}|^2) + 4\sigma^2\Big)^{1/2}\Big] \end{aligned} \tag{36}$$

Next, we have to minimize the maximum mutual information defined by (34) and by (35) with respect to the noise correlation parameter $r \in [0,1)$ (recall that minimization with respect to θ is already achieved). We partition the SNR range $[0,\infty)$ into two intervals, called in following the *low-SNR* and the *high-SNR* regions, and defined by the range of A for which $\xi \le 1/\lambda_2$ or $\xi > 1/\lambda_2$, respectively (notice that $\lambda_1 \ge \lambda_2$ holds for any channel matrix and value of r). Then, we consider separately the minimization of the upperbound on the two regions.

Low SNR region. Let $r = r^\star = |g_{2,1}/g_{1,1}| = \sigma/|g_{1,1}|^2$. Then, we obtain

$$\lambda_1 = |g_{1,1}|^2, \quad \lambda_2 = \frac{|g_{1,1}g_{2,2}|^2}{|g_{1,1}|^2 - |g_{2,1}|^2}$$

For

$$0 \le A \le A_1 = \frac{|g_{1,1}|^2 - |g_{2,1}|^2 - |g_{2,2}|^2}{|g_{1,1}g_{2,2}|^2} = \frac{|\mathbf{h}^1|^2 - |\mathbf{h}^2|^2}{\det(\mathbf{H}\mathbf{H}^H)} \tag{37}$$

the resulting mutual information is $\log(1 + |g_{1,1}|^2 A)$, which coincides with the first line of (32). This is achievable under the broadcast channel constraint (non-cooperative receivers) by MRC beamforming to user 1 only (the best user), and therefore it is clearly a tight upper bound. For $|\mathbf{h}^1|^2 = |\mathbf{h}^2|^2$ we have $A_1 = 0$ and under this condition the low-SNR case is irrelevant.

High SNR region. In this case, (34) and (35) become

$$\begin{aligned} f(r,\theta^\star,A) &= \log(\xi\lambda_1) + \log(\xi\lambda_2) \\ \xi &= \frac{1}{2}\left(A + \frac{1}{\lambda_1} + \frac{1}{\lambda_2}\right) \end{aligned} \tag{38}$$

By substituting we obtain

$$
\begin{aligned}
f(r,\theta^\star,A) &= 2\log\left(\sqrt{\lambda_1\lambda_2}\left(A+\frac{\lambda_1+\lambda_2}{\lambda_1\lambda_2}\right)\right)-2\log 2\\
&= 2\log\left(\sqrt{D}\,(A+T/D)\right)-2\log 2 \qquad (39)
\end{aligned}
$$

Trace and determinant T and D are written in terms of the $g_{i,j}$'s as

$$
T=\frac{|g_{1,1}|^2+|g_{2,1}|^2+|g_{2,2}|^2-2r\sigma}{1-r^2},\qquad D=\frac{|g_{1,1}g_{2,2}|^2}{1-r^2}
$$

By direct substitution of the above expressions into (39) and by differentiating with respect to r we obtain a stationary point in

$$
r^\star=\frac{2|g_{1,1}g_{2,1}|}{A|g_{1,1}g_{2,2}|^2+|g_{1,1}|^2+|g_{2,1}|^2+|g_{2,2}|^2}=\frac{2|\mathbf{h}^2(\mathbf{h}^1)^H|}{A\det(\mathbf{H}\mathbf{H}^H)+\mathrm{trace}(\mathbf{H}\mathbf{H}^H)}
$$

Notice that $r^\star$ is a decreasing function of A, and $\lim_{A\to\infty} r^\star=0$. Therefore, the worst-case noise in the large-SNR case is white. Also, for $A=A_1$ we obtain $r^\star=|g_{2,1}/g_{1,1}|\leq 1$. Then, the solution $r^\star$ is compatible with the positive definiteness constraint on $\mathbf{\Sigma}_z$ for all finite A. Finally, we observe that the worst-case noise correlation $r^\star$ is continuous in A for all $A\geq 0$, since the limits of $r^\star$ for $A\to A_1$ from the left and from the right are the same. Eventually, by substituting $r^\star$ found above into (38) we obtain the second line of (32).

Modified RKI throughput lowerbound. The throughput achievable by the modified RKI strategy is given by

$$
R^{\mathrm{mod-rki}}=\log\left(1+\frac{|g_{1,1}r_{1,1}|^2a_1}{1+|g_{1,1}r_{1,2}|^2a_2}\right)+\log\left(1+|g_{2,1}r_{1,2}+g_{2,2}r_{2,2}|^2a_2\right) \qquad (40)
$$

where we let

$$
\mathbf{R}=\left[\begin{array}{cc}\sqrt{a_1}r_{1,1} & \sqrt{a_2}r_{1,2}\\ 0 & \sqrt{a_2}r_{2,2}\end{array}\right]
$$

subject to the constraint $\mathrm{trace}(\mathbf{R}\mathbf{R}^H)=A$, which is written explicitly as

$$
|r_{1,1}|^2a_1+(|r_{1,2}|^2+|r_{2,2}|^2)a_2=A
$$

This must be maximized with respect to a_1,a_2 and the coefficients $r_{1,1},r_{1,2},r_{2,2}$. To this purpose, we reparameterize the problem by letting $b=|g_{2,1}/g_{2,2}|$, $z=|r_{1,2}/r_{2,2}|$, $q=z^2/(1+z^2)$ and $p=(bz+1)^2/(1+z^2)$, $X_1=|r_{1,1}|^2a_1$ and $X_2=(|r_{1,2}|^2+|r_{2,2}|^2)a_2$. Then, we obtain

$$
R^{\mathrm{mod-rki}}=\log\left(1+\frac{|g_{1,1}|^2X_1}{1+|g_{1,1}|^2qX_2}\right)+\log(1+|g_{2,2}|^2pX_2) \qquad (41)
$$

with the constraint $X_1 + X_2 = A$. For $z = 0$, the modified RKI strategy reduces to the standard RKI strategy and for $X_1 = A$ and $X_2 = 0$ it achieves (32) in the low-SNR region. Therefore, we shall consider only the high-SNR region $A \geq A_1$.

By dividing all elements of $\mathbf{G}$ by $|g_{1,1}|$ and by replacing the input constraint A by $a \triangleq A|g_{1,1}|^2$ we obtain the equivalent problem

$$R^{\mathrm{mod-rki}} = \log\left(1 + \frac{x_1}{1 + qx_2}\right) + \log(1 + px_2) \tag{42}$$

where $x_1 + x_2 = a$ and where, by letting $z = \sqrt{q/(1-q)}$ and $b = \sqrt{\alpha/\beta}$ with $\alpha = |g_{2,1}/g_{1,1}|^2$ and $\beta = |g_{2,2}/g_{1,1}|^2$, we have the relation

$$p = \left(\sqrt{\alpha q} + \sqrt{\beta(1-q)}\right)^2$$

The high-SNR condition $A \geq A_1$ translates into the condition $a \geq (1 - \alpha - \beta)/\beta$.

For any fixed $q \in [0, 1]$, we let $x_1 = a - x$ and $x_2 = x$ in (42) and maximize the resulting expression for $x \in [0, a]$. By letting $\frac{\partial}{\partial x} R^{\mathrm{mod-rki}} = 0$ and making the substitution $y = 1 + qx$, we obtain the solution

$$y = \sqrt{\frac{(1 + aq)(p - q)}{p(1 - q)}} \tag{43}$$

which is valid if $p \geq q$ and $y \geq 1$. The first condition yields the inequality

$$\left(\sqrt{\alpha q} + \sqrt{\beta(1-q)}\right)^2 \geq q$$

which implies

$$0 \leq q \leq q_{\max} = \frac{\beta}{(1 - \sqrt{\alpha})^2 + \beta} \leq 1$$

The second condition yields the inequality

$$\left(\sqrt{\alpha q} + \sqrt{\beta(1-q)}\right)^2 \geq \frac{1 + aq}{1 + a}$$

By letting $q = z^2/(1 + z^2)$ this is turned into the second order inequality

$$(\alpha - 1)z^2 + 2\sqrt{\alpha\beta}z + \beta - 1/(1 + a) \geq 0$$

which implies $z \in [z_1, z_2]$, where

$$z_{1,2} = \frac{\sqrt{\alpha\beta} \pm \sqrt{\beta - \frac{1-\alpha}{1+a}}}{1 - \alpha}$$

For $a \geq (1-\alpha-\beta)/\beta$ it is easy to check that $\beta-(1-\alpha)/(1+a) \geq 0$, therefore the above solution always exists in the high-SNR region. It is easy to check that

$$0 \leq q_1 \triangleq \frac{z_1^2}{1+z_1^2} \leq q_2 \triangleq \frac{z_2^2}{1+z_2^2} \leq q_{\max}$$

Therefore, the solution (43) is valid in the interval $q \in [q_1, q_2]$ and the maximum throughput can be obtained by first substituting (43) into (42) and then maximizing with respect to $q \in [q_1, q_2]$. After substitution, the function to be maximized is

$$2\log\left[\frac{1}{q}\left(\sqrt{(\sqrt{\alpha q}+\sqrt{\beta(1-q)})^2(1+aq)}-\sqrt{\left((\sqrt{\alpha q}+\sqrt{\beta(1-q)})^2-q\right)(1-q)}\right)\right] \tag{44}$$

By substituting again $q = z^2/(1+z^2)$ into the above expression and letting the derivative with respect to z equal to zero we obtain a 5-th order equation, whose roots can be given (quite fortuitously!) in closed form as z_1, z_2 given above, $z_{3,4} = -\sqrt{\beta/\alpha}$ and

$$z_5 = \frac{2\sqrt{\alpha\beta}}{\beta(a+1)+1-\alpha}$$

It can be checked that $z_5 \in [z_1, z_2]$, and therefore it is the sought maximum.

Finally, by substituting the resulting

$$q^\star \triangleq \frac{z_5^2}{1+z_5^2} = \frac{4\alpha\beta}{(\beta(a+1)+1-\alpha)^2+4\alpha\beta}$$

into (44), we obtain the maximum throughput as

$$R^{\mathrm{mod-rki}} = \log\frac{(a\beta+1+\alpha+\beta)^2-4\alpha}{4\beta} \tag{45}$$

which coincides with the second line of (32). This concludes the proof. □

Example: average throughput in Rayleigh fading. We use Theorem 5 to compute the maximum average throughput $\bar{R}$ of the composite channel when $\mathbf{H}$ has i.i.d. entries $\sim \mathcal{N}_{\mathbb{C}}(0,1)$ (independent Rayleigh fading). Subject to the short-term constraint, we have simply $\bar{R}(A) = E_{\mathbf{H}}[R(A)]$ where $R(A)$ is given by (32) and where we put in evidence its dependence on the input constraint A. The maximum average throughput subject to a long-term constraint is obtained by solving

$$\left\{\begin{array}{ll}\max & E_{\mathbf{H}}[R(a)] \\ \text{subject to} & E_{\mathbf{H}}[a] = A, \quad a \geq 0\end{array}\right. \tag{46}$$

By using the standard Lagrange-Kuhn-Tucker technique [32], after some algebra we obtain the optimal transmit power allocation in the form

$$a(\mathbf{H}) = \left\{\begin{array}{ll}[\xi - 1/|\mathbf{h}^1|^2]_+ & \text{for } \frac{|\mathbf{h}^1|^4-|\mathbf{h}^2(\mathbf{h}^1)^H|^2}{|\mathbf{h}^1|^2\det(\mathbf{H}\mathbf{H}^H)} \geq \xi \\ \xi + \sqrt{\xi^2+\frac{4|\mathbf{h}^2(\mathbf{h}^1)^H|^2}{\det(\mathbf{H}\mathbf{H}^H)^2}} - \frac{\mathrm{trace}(\mathbf{H}\mathbf{H}^H)}{\det(\mathbf{H}\mathbf{H}^H)} & \text{otherwise}\end{array}\right. \tag{47}$$

where ξ satisfies $E_{\mathbf{H}}[a(\mathbf{H})] = A$. The condition $\frac{|\mathbf{h}^1|^4 - |\mathbf{h}^2(\mathbf{h}^1)^H|^2}{|\mathbf{h}^1|^2 \det(\mathbf{H}\mathbf{H}^H)} \geq \xi$ is equivalent to $a(\mathbf{H}) \leq A_1$, where A_1 is given in Theorem 5. Hence, by substituting (47) in the place of A in (32), we obtain explicitly the optimal throughput subject to the long-term power constraint as $\bar{R}(A) = E_{\mathbf{H}}[f_\xi(\mathbf{H})]$ where

$$f_\xi(\mathbf{H}) = \begin{cases} [\log(\xi|\mathbf{h}^1|^2)]_+ & \text{for } \frac{|\mathbf{h}^1|^4 - |\mathbf{h}^2(\mathbf{h}^1)^H|^2}{|\mathbf{h}^1|^2 \det(\mathbf{H}\mathbf{H}^H)} \geq \xi \\ \log\left[\det(\mathbf{H}\mathbf{H}^H)\xi\left(\xi + \sqrt{\xi^2 + \frac{4|\mathbf{h}^2(\mathbf{h}^1)^H|^2}{\det(\mathbf{H}\mathbf{H}^H)^2}}\right)\right] - \log 2 & \text{otherwise} \end{cases} \tag{48}$$

Figs. 1 and 2 show $\bar{R}$ in the case $t = r = 2$ for the short and the long-term constraints, respectively, vs. E_b/N_0. For the sake of comparison, we show also the 2×2 cooperative and ZF throughputs.

In this work, E_b/N_0 for the $t \times 1 : r$ GBC is defined as [37]

$$\frac{E_b}{N_0} \triangleq \frac{tA}{R} \tag{49}$$

the factor t in the numerator of (49) is introduced to take into account that, under mild conditions on $\mathbf{H}$, the average *received* energy per channel use increases linearly with t for MRC beamforming to any given user (e.g., if $\mathbf{H}$ has i.i.d. elements with unit second order moment the individual user channels have average gain t). For the cooperative system, since $t \times r$ and $r \times t$ channels yield the same throughput [3], we adopt the definition

$$\frac{E_b}{N_0} \triangleq \frac{\max\{r, t\}A}{R} \tag{50}$$

For the short-term constraint, there exists a minimum $(E_b/N_0)_{\min} > 0$ below which $\bar{R}$ is zero. [3] This can be calculated by letting $A \downarrow 0$ in (49) and in (50). For the $2 \times 1 : 2$ GBC (basic RKI, ZF and modified RKI schemes) we obtain

$$\left(\frac{E_b}{N_0}\right)_{\min} = \frac{2\log 2}{E[|\mathbf{h}^1|^2]} = \frac{2\log 2}{(11/4)} = -2.97 \text{ dB}$$

where we used the fact that $|\mathbf{h}^1|^2$ is distributed as the maximum of two independent and identically distributed central Chi-squared random variables with 4 degrees of freedom. For the 2×2 cooperative system we have

$$\left(\frac{E_b}{N_0}\right)_{\min} = \frac{2\log 2}{E[c_1]} = \frac{2\log 2}{(7/2)} = -4.02 \text{ dB}$$

where we used the fact that c_1 is the maximum eigenvalue of the 2×2 Wishart matrix [3] $\mathbf{H}\mathbf{H}^H$.

From Figs. 1 and 2 it is clearly visible that the basic RKI scheme is optimal both for $E_b/N_0 \downarrow (E_b/N_0)_{\min}$ and for $E_b/N_0 \to \infty$. For small SNR, it is (asymptotically) equivalent

[3] For the long-term constraint $(E_b/N_0)_{\min} = 0$ since $|\mathbf{h}^1|^2$ has a distribution with unbounded support [38].

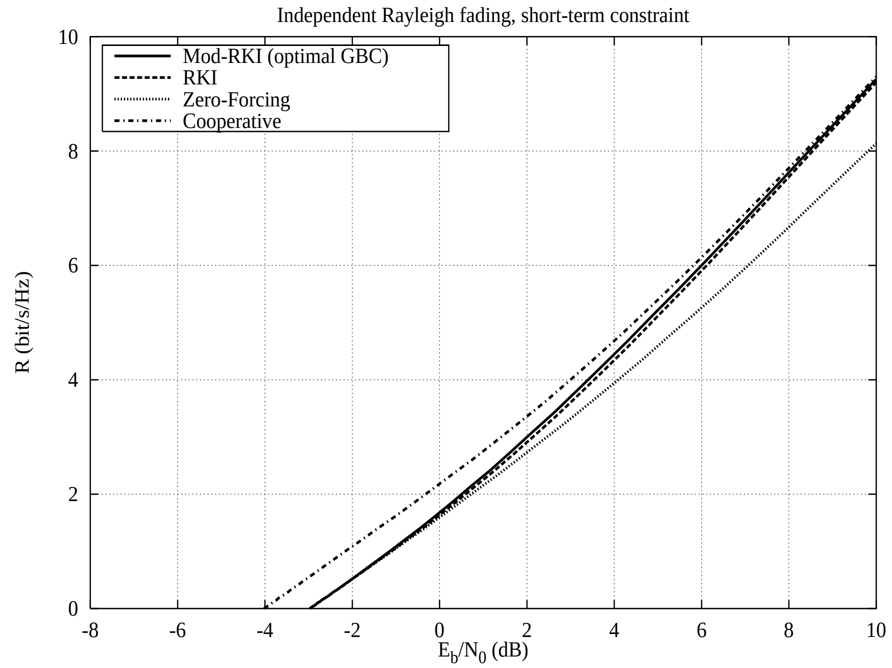

Figure 1: Throughput of the $2 \times 1 : 2$ GBC with independent Rayleigh fading and short-term input constraint.

to ZF and both strategies reduce to simple MRC beamforming to the best user. For large SNR, it is (asymptotically) equivalent to the cooperative single-user multiple-antenna capacity. This is a consequence of the fact that for independent Rayleigh fading the channel matrix $\mathbf{H}$ has full-rank almost surely, therefore Theorem 3 applies to almost all realizations of the channel. $\Diamond$

6 Performance of the basic RKI strategy with Rayleigh fading

In this section we focus on the basic RKI scheme when the channel matrix has i.i.d. entries $\sim \mathcal{N}_{\mathbb{C}}(0, 1)$, and we consider the throughput of the composite channel subject to a long-term power constraint. We consider both the finite dimensional and the large-system limit, i.e., when $r, t \to \infty$ while the ratio r/t users/transmit antenna converges to a given constant α (referred to in the following as *antenna loading*). Moreover, we assume that no effort is made to optimize the user ordering. In other words, the receiver works with the natural ordering of the rows of $\mathbf{H}$ and considers the coefficients $\{d_i : i = 1, \ldots, m\}$ corresponding to the QR decomposition $\mathbf{H} = \mathbf{G}\mathbf{Q}$.

It interesting to notice that since the composite channel is symmetric with respect to any user by time-sharing with uniform probability over all possible user subsets and

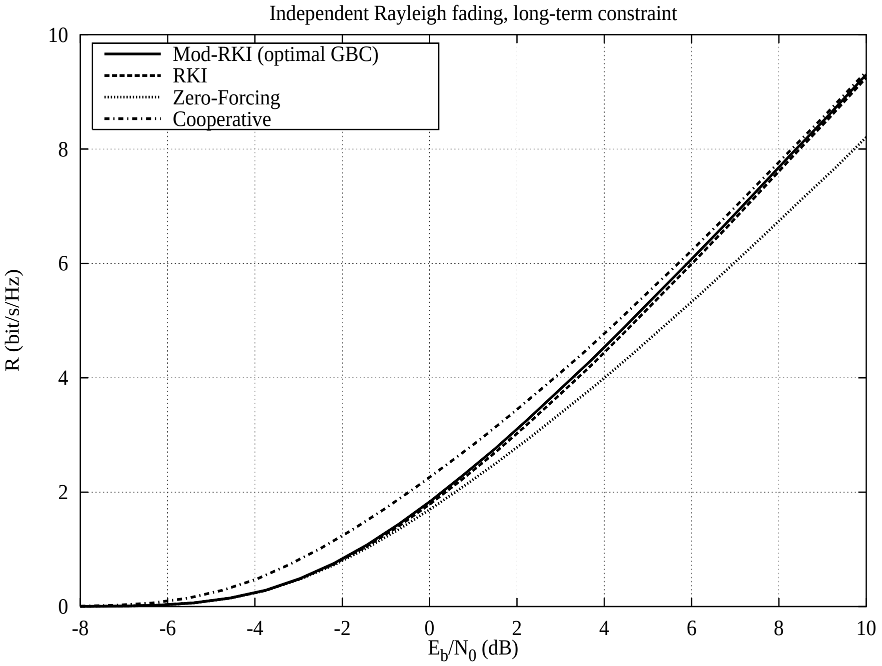

Figure 2: Throughput of the $2 \times 1 : 2$ GBC with independent Rayleigh fading and long-term input constraint.

orderings, every user in the system achieves the same average per-user rate $\rho \stackrel{\Delta}{=} \bar{R}/r$ with no loss of optimality in the total throughput.

We make use of the following results [39, 40, 41, 42]:

Lemma 4. Let $\mathbf{H} \in \mathbb{C}^{r \times t}$ have i.i.d. entries $\sim \mathcal{N}_{\mathbb{C}}(0,1)$, and let $g_{i,i}$ be the i-th diagonal element of $\mathbf{G}$ in the QR decomposition $\mathbf{H} = \mathbf{G}\mathbf{Q}$. Then, the random variables $d_i = |g_{i,i}|^2$ are statistically independent and $d_i \sim \chi^2_{2(t-i+1)}$, where χ^2_{2k} denotes the central Chi-squared distribution with $2k$ degrees of freedom, whose pdf is $f(z) = z^{k-1}e^{-z}/(k-1)!$. □

Lemma 5. Let $\nu = i/r \in [0,1]$ denote the normalized user index. If $\mathbf{H} \in \mathbb{C}^{r \times t}$ (for $r/t = \alpha$) has i.i.d. circularly-symmetric elements with mean 0, variance 1 and bounded fourth moment, then

$$\lim_{r \to \infty} \frac{1}{t} d_i = [1 - \alpha\nu]_+ \tag{51}$$

with probability 1. □

In the case of finite r, t, Corollary 1 yields ξ solution of $\sum_{i=1}^m E\left[[\xi - \frac{1}{d_i}]_+\right] = A$, where, from Lemma 4, we have

$$\begin{aligned}\sum_{i=1}^{m} E\left[\left[\xi - \frac{1}{d_i}\right]_+\right] &= \sum_{i=1}^{m} \int_{1/\xi}^{\infty} \left(\xi - \frac{1}{z}\right) \frac{z^{t-i}e^{-z}}{(t-i)!} dz \\ &= \sum_{i=1}^{m} \frac{1}{(t-i)!} \left[\xi\Gamma(t-i+1, 1/\xi) - \Gamma(t-i, 1/\xi)\right] \qquad (52)\end{aligned}$$

where $\Gamma(n,x) \triangleq \int_x^\infty z^{n-1}e^{-z}dz$, and where $\Gamma(0,x) = \mathrm{E_i}(1,x)$ and, for integer $n \geq 1$, $\Gamma(n,x) = (n-1)!e^{-x}\sum_{j=0}^{n-1} x^j/j!$.

The resulting RKI average throughput is given by

$$\begin{aligned}\bar{R}^{\mathrm{rki}} &= \sum_{i=1}^{m} E\left[[\log(\xi d_i)]_+\right] \\ &= \sum_{i=1}^{m} \frac{1}{\xi^{t-i+1}} \int_1^\infty \log(z) \frac{z^{t-i}e^{-z/\xi}}{(t-i)!} dz \\ &= \sum_{i=1}^{m} \mathcal{J}_{t-i+1}(\xi) \qquad (53)\end{aligned}$$

where we let [43]

$$\begin{aligned}\mathcal{J}_k(a) &\triangleq a^{-k} \int_1^\infty \log(z) \frac{z^{k-1}e^{-z/a}}{(k-1)!} dz \\ &= \mathrm{E_i}(1, 1/a) + \sum_{\ell=1}^{k-1} \frac{1}{\ell} e^{-1/a} \sum_{j=0}^{\ell-1} \frac{a^{-j}}{j!} \qquad (54)\end{aligned}$$

With the same assumptions on the statistics of $\mathbf{H}$, the elements b_i in the ZF throughput formula (14) are identically distributed $\sim \chi^2_{2(t-k+1)}$, for any subset of active users of cardinality $k \leq m$. Here, the cardinality k of the active user set can be chosen in order to optimize further the ZF throughput. By replicating the calculation done above in the case of k active users we obtain the ZF throughput subject to the long-term constraint as

$$\bar{R}^{\mathrm{zf}} = \max_{k=1,\ldots,m} k\mathcal{J}_{t-k+1}(\xi_k) \qquad (55)$$

where ξ_k is the solution of the waterfilling equation

$$\frac{k}{(t-k)!}\left[\xi\Gamma(t-k+1, 1/\xi) - \Gamma(t-k, 1/\xi)\right] = A$$

analogous to (52).

The throughput of the cooperative system can be obtained from the pdf of a single eigenvalue of the $m \times m$ Wishart matrix

$$\mathbf{W} = \begin{cases} \mathbf{H}\mathbf{H}^H & \text{for } r \leq t \\ \mathbf{H}^H\mathbf{H} & \text{for } r > t \end{cases} \qquad (56)$$

given by [3]

$$p_\lambda(z) = \frac{1}{m}\sum_{i=0}^{m-1} \frac{i!}{(i+n-m)!}\left[L_i^{(n-m)}(z)\right]^2 z^{n-m}e^{-z} \tag{57}$$

where $n \triangleq \max\{r,t\}$ and where $L_k^\ell(x) \triangleq \frac{1}{k!}e^x x^{-\ell}\frac{d^k}{dx^k}(e^{-x}x^{\ell+k})$ is the associated Laguerre polynomial of order k. We have

$$\bar{R}^{\rm coop} = m\int_{1/\xi}^\infty \log(\xi z)p_\lambda(z)dz \tag{58}$$

where ξ is the solution of the waterfilling equation

$$\int_{1/\xi}^\infty \left(\xi - \frac{1}{z}\right)p_\lambda(z)dz = A/m$$

Interestingly, both the LHS of the above equation and (58) can be put easily in closed form. In particular, (58) can be given as a linear combination of functions $\mathcal{J}_k(a)$ is defined in (54). For example, for $t = r = 2$ we obtain

$$\bar{R}^{\rm coop} = 2\left(\mathcal{J}_3(\xi) - \mathcal{J}_2(\xi) + \mathcal{J}_1(\xi)\right)$$

where ξ is the solution of

$$\left(1 + 2\xi - 2{\rm E_i}(1,1/\xi)e^{1/\xi}\right)e^{-1/\xi} = A$$

Fig. 3 shows the RKI throughput for $t = 4, r = 2$ and $t = 4, r = 4$ cases. The ZF and cooperative throughputs are shown for comparison. The throughput gain of the basic RKI strategy over ZF beamforming is very significant for $t = 4, r = 4$, and less significant for $t = 4, r = 2$.

Next, we study the RKI normalized throughput ρ in the *large-system* regime, i.e., we let $r \to \infty$ with $r/t = \alpha$, where $\alpha \geq 0$ is the *antenna loading*. From Corollary 1 and Lemma 5, we have immediately that $\rho^{\rm rki}$ is given as the solution of

$$\left\{\begin{array}{ll} \max & \int_0^\mu \log\left(1 + [1-\alpha\nu]_+ a(\nu)\right)\, d\nu \\ \text{subject to} & \int_0^\mu a(\nu)\, d\nu = A/\alpha, \quad a(\nu) \geq 0 \end{array}\right. \tag{59}$$

where $a(\nu)$ is the transmit SNR of the $\lfloor \nu r \rfloor$-th signal and $\mu \triangleq m/r = \min\{1, 1/\alpha\}$. As corollaries of the large-system "ergodization" of Lemma 5 we have that: 1) In the large-system limit the composite channel has constant instantaneous throughput (average and instantaneous throughputs coincide); 2) The long-term and the short-term constraint are equivalent; 3) The throughput is asymptotically linear in the number of users with slope $\rho^{\rm rki}$, i.e., $R^{\rm rki} \approx r\rho^{\rm rki}$ for large r.

The optimal $a(\nu)$ is given by

$$a(\nu) = \left[\xi - \frac{1}{1-\alpha\nu}\right]_+ \tag{60}$$

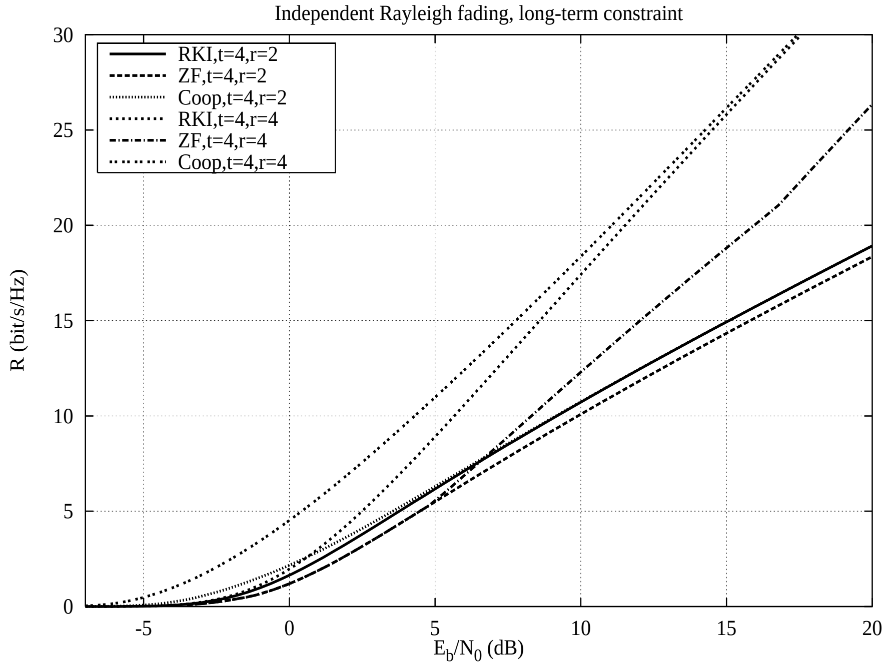

Figure 3: Throughput of RKI, ZF and cooperative strategies in the case $t = 4, r = 2$ and $t = 4, r = 4$, with independent Rayleigh fading.

where ξ is the solution of the waterfilling equation $\int_0^\mu [\xi - 1/(1-\alpha\nu)]_+ d\nu = A/\alpha$, and obviously $\xi \geq 1$. By letting

$$\mu' \triangleq \min\{\mu, (1/\alpha)(1 - 1/\xi)\} = \min\{1, (1/\alpha)(1 - 1/\xi)\} \tag{61}$$

we obtain ξ as the solution of

$$\mu'\xi + \frac{1}{\alpha}\log(1 - \alpha\mu') = A/\alpha \tag{62}$$

Explicitly, this yields the following solutions:

1. For $\alpha < 1$ and $A \geq \frac{\alpha}{1-\alpha} + \log(1-\alpha)$,

$$\xi = \frac{1}{\alpha}(A - \log(1-\alpha)) \tag{63}$$

2. For $\alpha < 1$ and $A < \frac{\alpha}{1-\alpha} + \log(1-\alpha)$, or for $\alpha \geq 1$ and any A, ξ is the solution of

$$\xi - \log\xi - 1 = A \tag{64}$$

By substituting the above solution in the expression of the normalized throughput we obtain,

$$\rho^{\text{rki}-\max} = \begin{cases} \log\left[\frac{1}{\alpha}(A - \log(1-\alpha))\right] - (1/\alpha - 1)\log(1-\alpha) - 1 & \text{(case 1)} \\ \frac{1}{\alpha}\left(\log\xi + \frac{1}{\xi} - 1\right) & \text{(case 2)} \end{cases} \tag{65}$$

where "case 1" and "case 2" refer to the cases 1 and 2 given above for ξ.

For the sake of comparison, we calculate also the normalized throughput with ZF beamforming and with cooperative receivers in the large-system regime. In the case of ZF, it can be shown that [41, 40] as $r \to \infty$ and $r/t = \alpha$, then $\frac{1}{t}b_i \to [1-\alpha\kappa]_+$ with probability 1, where $\kappa \in [0,\mu]$ is the fraction of active users. Then, the normalized ZF throughput can be written as $\rho^{\rm zf} = \int_0^\kappa \log(1+(1-\kappa\alpha)a(\nu))d\nu$ with the constraint $\int_0^\kappa a(\nu)d\nu = A/\alpha$. Since the logarithm is concave, it is immediate to see that this is maximized by $a(\nu)$ =constant. The result can be further maximized with respect to the fraction of active users κ, so that we obtain

$$\rho^{\rm zf} = \max_{\kappa\in[0,\mu]} \kappa \log(1+[1-\kappa\alpha]A/(\kappa\alpha)) \tag{66}$$

For the cooperative throughput, we use the following result (see [3, 41] and references therein). The limiting distribution of the non-zero eigenvalues of the normalized Wishart matrix $\frac{1}{m}\mathbf{W}$, where $\mathbf{W}$ is given in (56), is given by

$$f_\lambda(z) \triangleq \frac{1}{2\pi}\sqrt{[1-\lambda_-/z]_+ \, [\lambda_+/z-1]_+} \tag{67}$$

with

$$\lambda_- = \left(\sqrt{\max\{\alpha,1/\alpha\}}-1\right)^2, \quad \lambda_+ = \left(\sqrt{\max\{\alpha,1/\alpha\}}+1\right)^2$$

Hence, the limiting normalized throughput $\rho^{\rm coop}$ is given by

$$\left\{\begin{array}{l} \max \quad \mu\int_{\lambda_-}^{\lambda_+} \log(1+za(z))f_\lambda(z)dz \\ \text{subject to} \quad \int_{\lambda_-}^{\lambda_+} a(z)f_\lambda(z)dz \leq A \end{array}\right. \tag{68}$$

This is maximized by the waterfilling power allocation $a(\lambda) = [\xi - 1/\lambda]_+$, leading to

$$\rho = \mu \int_{\max\{1/\xi,\lambda_-\}}^{\max\{1/\xi,\lambda_+\}} \log(\xi z) f_\lambda(z)\, dz \tag{69}$$

where ξ is the solution of

$$\int_{\max\{1/\xi,\lambda_-\}}^{\max\{1/\xi,\lambda_+\}} \left(\xi - \frac{1}{z}\right) f_\lambda(z)\, dz = A \tag{70}$$

Fig. 4 shows the normalized throughputs of the RKI, ZF and cooperative schemes for $\alpha = 0.5, 1.0$ and 2.0. For $\alpha \leq 1$ the basic RKI strategy considered in this section, i.e., without optimization with respect to the user ordering, is asymptotically optimal for large SNR since the channel matrix has rank r with probability 1. On the contrary, we cannot invoke Theorem 3 to claim the asymptotic optimality of the basic RKI strategy in the case $\alpha > 1$, since in this case the channel matrix has rank $t < r$ with probability 1.

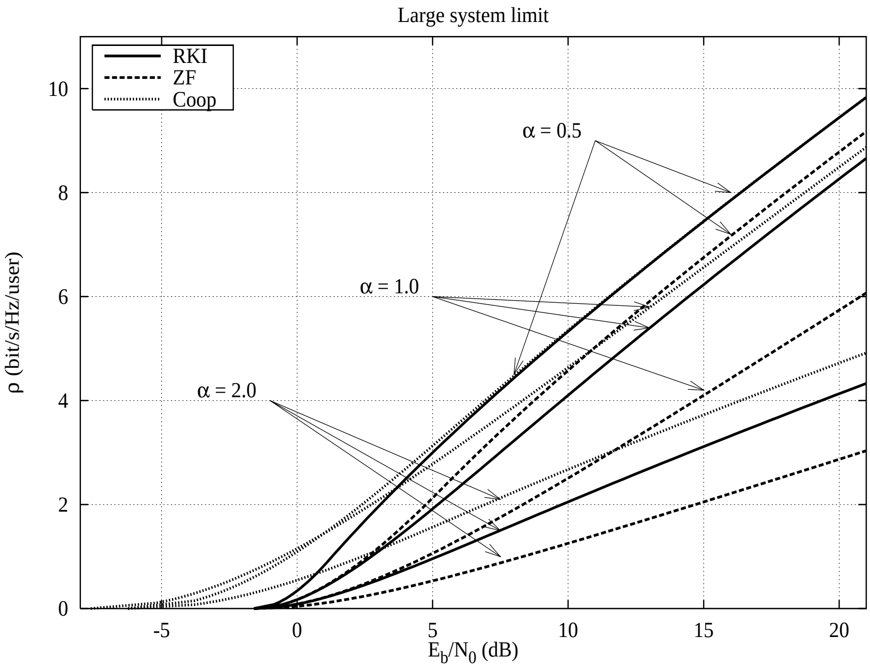

Figure 4: Normalized throughput in the large-system limit of RKI, ZF and cooperative strategies for $\alpha = 0.5, 1.0$ and 2.0.

Calculation of $(E_b/N_0)_{\min}$**.** For RKI (see (65)) as $A \downarrow 0$ we are always in "case 2", therefore,

$$\begin{aligned}\lim_{A\to 0}\frac{\alpha\rho}{A} &= \lim_{A\to 0}\frac{\log\xi + 1/\xi - 1}{A}\\ &= \lim_{A\to 0}\frac{\xi - 1}{\xi}\frac{d\xi}{dA}\\ &= 1\end{aligned}$$

where the last line follows from (64) which implies $\frac{d\xi}{dA} = \frac{\xi}{\xi - 1}$. By using this into (49) we obtain $(E_b/N_0)_{\min} = \log 2$. ZF beamforming achieves obviously the same value. For the cooperative scheme, by using (69) and (70) we can write

$$\begin{aligned}
\lim_{A\to 0} \frac{\rho}{A} &= \lim_{A\to 0} \frac{\mu}{A} \int_{1/\xi}^{\lambda_+} \log(\xi z) f_\lambda(z) dz \\
&= \mu \lim_{A\to 0} \left\{ \frac{d}{d\xi} \int_{1/\xi}^{\lambda_+} \log(\xi z) f_\lambda(z) dz \right\} \frac{d\xi}{dA} \\
&= \mu \lim_{\xi \to 1/\lambda_+} \frac{\frac{d}{d\xi} \int_{1/\xi}^{\lambda_+} \log(\xi z) f_\lambda(z) dz}{\frac{d}{d\xi} \int_{1/\xi}^{\lambda_+} (\xi - 1/z) f_\lambda(z) dz} \\
&= \mu \lim_{\nu \to \lambda_+} \frac{\frac{d}{d\nu} \int_{\nu}^{\lambda_+} \log(z/\nu) f_\lambda(z) dz}{\frac{d}{d\nu} \int_{\nu}^{\lambda_+} (1/\nu - 1/z) f_\lambda(z) dz} \\
&= \mu \lim_{\nu \to \lambda_+} \frac{-1/\nu \int_{\nu}^{\lambda_+} f_\lambda(z) dz}{-1/\nu^2 \int_{\nu}^{\lambda_+} f_\lambda(z) dz} \\
&= \mu \lambda_+ \\
&= (1 + 1/\sqrt{\alpha})^2
\end{aligned} \tag{71}$$

By using this into (50) we obtain $(E_b/N_0)_{\min} = \log 2/(1 + \sqrt{\min\{\alpha, 1/\alpha\}})^2$.

7 Conclusions

We investigated the achievable throughput of a generally non-degraded broadcast Gaussian channel where the transmitter has t antennas and the r receivers have one antenna each, subject to the assumption that the channel is perfectly known to all terminals ($t \times 1 : r$ GBC with perfect CSIT and CSIR). For this model, we proposed a new coding strategy where the t-input r-output channel is decomposed into $m = \text{rank}\{\mathbf{H}\}$ set of ranked interference channels, and where interference in channel k is due to signals transmitted to users $1, \ldots, k-1$. Since this interference signals are known non-causally by the transmitter, this can make use of known techniques for coding with known interference (e.g., the lattice-precoding strategy) in order to make interference in each channel harmless without power penalty.

We proved that the proposed RKI scheme is asymptotically optimal for high and low SNR, provided that $m = r$ (full row-rank channel matrix). For high SNR, the achievable throughput converges to the same throughput achievable by a $t \times r$ multiple-antenna single-user system obtained by letting all receivers to cooperate. For low SNR, MRC beamforming to the user enjoying the largest individual channel capacity is optimal.

For the special case of two users (arbitrary t) we found a closed-form expression for the optimal throughput at any SNR. The optimal throughput is achieved by a modification of the basic RKI strategy, optimized with respect to some design parameters.

Driven by the above results, we conclude by pointing out some considerations for the downlink throughput optimization in a wireless communication system. As an example, consider Fig. 5 showing the throughput of a system with $r = 4$ users and $t = 4$ transmit antennas, independent Rayleigh fading and long-term power constraint, for the basic RKI strategy, the ZF beamforming strategy and the MRC beamforming strategy transmitting to the best user only, the optimal throughput for the simple degraded GBC with $t = 1$, also obtained by transmitting to the best user only, and the optimal throughput of the degraded GBC with $t = 4$ but no CSIT, given by (5). [4]

For relatively large SNR, the throughput gain due to $t = 4$ over $t = 1$ antennas at the transmitter is modest if the system is constrained to serve a single user per slot. On the contrary, the throughput gain provided by the asymptotically optimal RKI strategy can be very large even for moderate SNR, and increases with SNR. ZF beamforming yields the same optimal throughput slope for high SNR, but it pays a fairly large throughput penalty with respect to RKI. Moreover, this penalty increases with $\min\{r, t\}$, as illustrated in Fig. 6.

If CSIT is absent, the throughput slope is independent of t, i.e., the channel "degrees of freedom" depend critically on the availability of CSIT. This is a rare example of a Gaussian channel where lack of CSIT costs not only in SNR degradation, but actually in the pre-log factor (i.e., degrees of freedom) in the throughput formula.

We conclude by stressing the importance of exploiting transmitter channel knowledge in the multiple-antenna broadcast setting (downlink). For a system with a large number of users ($r \gg 1$) and fixed (large) transmit SNR A, a virtually arbitrarily large downlink throughput can be achieved by simply increasing the number of transmit antennas t and serving t users simultaneously. This, of course, depends on the ability of estimating reliably the channel matrix at the transmitter. In this respect, systems exploiting time-division duplexing (TDD) might be preferable, since the channel can be estimated from the uplink signals (see for example [44] and references therein).

[4] The cross-over of the RKI and ZF curves with the MRC curve for low SNR is due to the fact that for the former strategies no best user subset selection is made, while MRC selects always the best individual user.

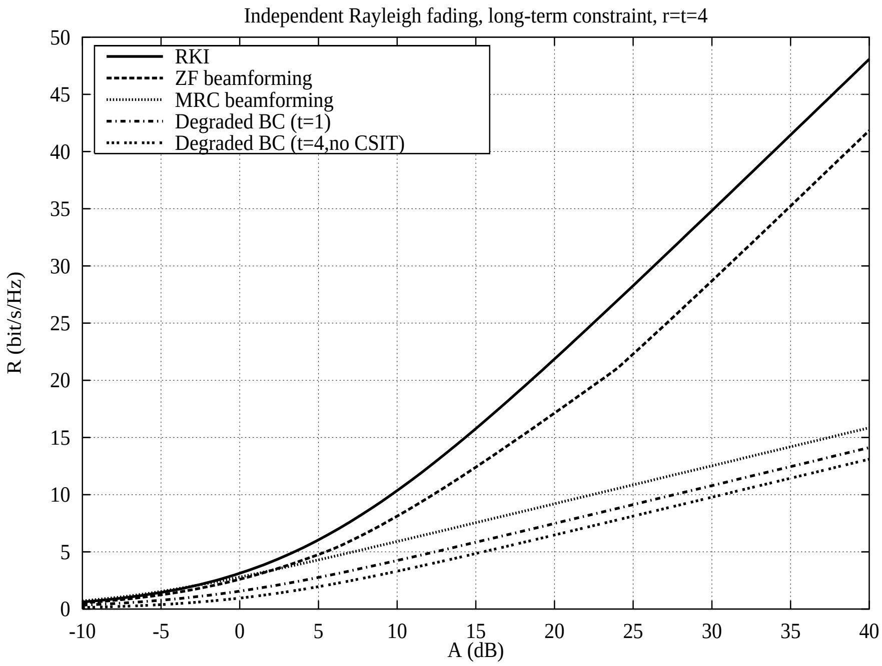

Figure 5: Throughput vs. SNR comparison for a system with independent Rayleigh fading and long-term input constraint, RKI, ZF, MRC with $t = 4, r = 4$, the degraded GBC with perfect CSIT, $t = 1, r = 4$ and the degraded GBC without CSIT, $t = 4$ and arbitrary r.

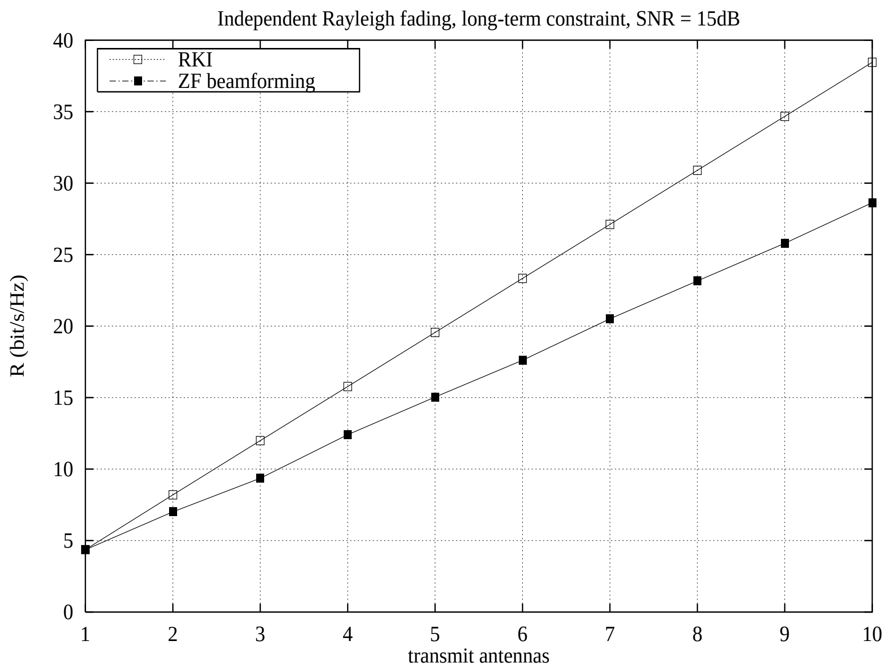

Figure 6: Throughput vs. the number of transmit antennas t for transmit SNR $A = 15$ dB in a system with independent Rayleigh fading and $r > t$ users, for RKI and ZF beamforming.

APPENDIX

A Extension to time-invariant frequency-selective channels

We consider a discrete-time ISI channel with finite memory of the form

$$\mathbf{y}_i = \sum_{\ell=0}^{L} \mathbf{H}_\ell \mathbf{x}_{i-\ell} + \mathbf{z}_i, \quad i = 1, \ldots, n \tag{72}$$

where L is finite, where $\mathbf{H}_\ell \in \mathbb{C}^{r \times t}$ and where the matrix impulse response $\mathbf{H}(D) = \sum_{\ell=0}^{L} \mathbf{H}_\ell D^\ell$ can be either deterministic or random but fixed. As in the frequency-flat case, the latter situation models the composite channel (block fading channel with arbitrarily large blocks) but, differently form the frequency-flat case, the resulting average throughput is not equivalent to that of a time-varying information-stable channel with CSIT causally known to the transmitter. In this case, we would have a time-varying random ergodic channel impulse response $\mathbf{H}(D, i) = \sum_{\ell=0}^{L} \mathbf{H}_\ell(i) D^\ell$ known to the transmitter for all $j \leq i$, which is a situation where not even the simple single-user capacity for the 1×1 channel is fully known [24].

For the deterministic and for the composite channel case we can apply [22, Theorem 1] (see also [45]), which holds for any multiterminal Gaussian network with finite-memory time-invariant channel impulse responses, and obtain that the capacity region (and therefore the throughput) of the frequency-selective $t \times 1 : r$ GBC defined by (72) is equivalent to that of a n-block circular channel [46] in the limit for large n. The n-block circular channel, which can be implemented in practice by adding a cyclic prefix of length L, is defined by $\text{vec}(\mathbf{Y}) = \underline{\mathbf{H}}\text{vec}(\mathbf{X}) + \text{vec}(\mathbf{Z})$ where $\text{vec}(\mathbf{A}) = [\mathbf{a}_1^T, \mathbf{a}_2^T, \ldots, \mathbf{a}_n^T]^T$ and where $\underline{\mathbf{H}}$ is a $nr \times nt$ block-circulant matrix with first row $[\mathbf{H}_0, \mathbf{0}, \cdots, \mathbf{0}, \mathbf{H}_L, \cdots, \mathbf{H}_1]$. The n-block circular channel is easily turned into a set of parallel $t \times 1 : r$ GBCs by applying discrete-Fourier transform, and in the limit for $n \to \infty$ we obtain the frequency-domain parallel channels

$$\mathbf{y}(\omega) = \mathbf{H}(\omega)\mathbf{x}(\omega) + \mathbf{z}(\omega) \ , \quad \omega \in [-1/2, 1/2]$$

where $\mathbf{H}(\omega) = \sum_{\ell=0}^{L} \mathbf{H}_\ell e^{-j2\pi\omega\ell}$.

For $t > 1$ the resulting channel is a set of parallel generally non-degraded GBCs. Nevertheless, by a straightforward application of Theorem 1 the RKI scheme can be used directly in the frequency domain, yielding the throughput

$$R^{\text{rki}} = \int_{-1/2}^{1/2} \sum_{k=1}^{m} [\log(\xi d_k(\omega))]_+ d\omega$$

where ξ is the solution of $\int_{-1/2}^{1/2} \sum_{k=1}^{m} [\xi - 1/d_k(\omega)]_+ d\omega = A$, and where $d_k(\omega)$ is the squared magnitude of the k-th diagonal element of the matrix $\mathbf{G}(\omega)$ obtained by the QR decomposition $\mathbf{H}(\omega) = \mathbf{G}(\omega)\mathbf{Q}(\omega)$.

In the composite channel case, subject to a long-term constraint, if $\mathbf{H}(\omega)$ is identically distributed for all ω the average throughput expressions simplify and the integral with respect to frequency disappears. In fact, we have

$$\bar{R}^{\rm rki} = E\left[\int_{-1/2}^{1/2}\sum_{k=1}^{m}[\log(\xi d_k(\omega))]_+ d\omega\right] = \sum_{k=1}^{m} E\left[[\log(\xi d_k(\omega))]_+\right]$$

with waterfilling equation $\sum_{k=1}^{m} E[[\xi - 1/d_k(\omega)]_+]$. In particular, this holds if the channel is Rayleigh with uncorrelated scattering, i.e., if the matrices $\mathbf{H}_\ell$ are jointly Gaussian zero-mean and mutually independent (notice that the elements inside each matrix $\mathbf{H}_\ell$ can be correlated, i.e., there might be spatial correlation but there must be path independence). In particular, the optimal power allocation and throughput formula given in (47) and (48) for the $2 \times 1 : 2$ GBC in the composite channel case subject to the long-term power constraint holds verbatim in the case of frequency selective uncorrelated-scattering Rayleigh fading.

B Proof of Theorem 1 and coding strategies for RKI

The achievability of (8) in Theorem 1 follows as a corollary of some previously known results. For the sake of completeness, we provide here their statements extended (under mild regularity conditions) to case of the continuous additive-noise channel with interference

$$Y_i = X_i + S_i + Z_i, \quad i = 1, \dots, n \tag{73}$$

where Y_i, X_i, S_i, Z_i take on values in the complex plane, where $\{Z_i\}$ is an i.i.d. circularly-symmetric noise sequence with mean zero and variance σ_z^2, $\{S_i\}$ is an interference random signal, and $\{X_i\}$ is subject to the input constraint $E[|X_i|^2] \leq \sigma_x^2$. In the following, we assume also that $\{S_i\}$ is a power-limited process, i.e., that there exist Q such that $E[|S_i|^2] \leq Q < \infty$. Further assumptions about channel (73) are: i) $\{S_i\}$ is statistically independent of the noise $\{Z_i\}$ and of the input $\{X_i\}$; ii) The transmitter knows the whole individual realization $\mathbf{s} = (s_1, \dots, s_n)$ of the interference signal non-causally, i.e., before transmission of its code word, but does not necessarily know the probability law with which $\mathbf{s}$ is generated. iii) The receiver has no knowledge of the probability law of $\{S_i\}$ and (obviously) of its individual realization $\mathbf{s}$.

In [14] Gel'fand and Pinsker considered the case where $\{S_i\}$ is i.i.d., with single-letter probability distribution q. In this case we have

Lemma B.1 (Gel'fand and Pinsker). The capacity of (73) when $\{S_i\}$ is i.i.d. with single-letter probability q is given by

$$C_q = \sup_{V,S,X: p(x,v,s) = p(x|v,s)p(v|s)q(s)} \{I(V;Y) - I(V;S)\} \tag{74}$$

where $p(x|v,s) = \delta(x - f(v,s))$, i.e., the optimization over $p(x|v,s)$ can be restricted to deterministic mappings $f : v, s \mapsto x$ without loss of optimality. □

Notice that whether or not the transmitter has knowledge of q is irrelevant here, since it can estimate q by looking at the (typical) realization $\mathbf{s}$ for sufficiently large n. Moreover, since q is fixed, this can be done "off-line" in the system set-up, with no capacity loss.

In [16] Ahlswede considered the general case where $\{S_i\}$ is arbitrarily distributed. In this case we have

Lemma B.2 (Ahlswede). The capacity of (73) when $\{S_i\}$ is distributed arbitrarily is given by

$$C = \inf_q C_q \tag{75}$$

□

The above result "contains" the case where $\{S_i\} = \mathbf{s}_0$, an individual arbitrary interference signal, as a special case, since this corresponds to the case where the probability assignment of $\{S_i\}$ ranges over all single-mass point distributions.

Now we focus on the AWGN case, where $Z_i \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_z^2)$, and we use Lemmas B.1 and B.2 to prove our achievability result.

Proof of Theorem 1. The proof makes use of a derivation found in [18, Sect. 2.2.1]. Choose an arbitrary (finite-power) distribution q and assume that $\{S_i\}$ is i.i.d. with $S_i \sim q$. Since all the following expressions are "single-letter", we drop the time index i for the sake of simplicity. Let $X \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_x^2)$ independent of S and Z. Let $V = \alpha S + X$ with $\alpha = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_z^2}$. We have

$$V - \alpha Y = \alpha S + X - \alpha X - \alpha S - \alpha Z = X - \alpha(X + Z)$$

Notice that $V - \alpha Y$ is independent of $X + Z$, in fact, they are jointly Gaussian and

$$E[(V - \alpha Y)(X + Z)^*] = E[(X - \alpha(X + Z))(X + Z)^*] = \sigma_x^2 - \alpha(\sigma_x^2 + \sigma_z^2) = 0$$

Moreover, $V - \alpha Y$ and Y are also independent, since S is independent of X and Z. Then, we can write

$$\begin{aligned} h(V|Y) &= h(V - \alpha Y|Y) \\ &= h(V - \alpha Y) \\ &= h(V - \alpha Y|X + Z)) \\ &= h(X - \alpha(X + Z)|X + Z) \\ &= h(X|X + Z) \end{aligned} \tag{76}$$

and

$$\begin{aligned} h(V|S) &= h(\alpha S + X|S) \\ &= h(X|S) \\ &= h(X) \end{aligned} \tag{77}$$

By using (76)and (77) in the capacity formula of Lemma B.1 we have

$$\begin{aligned} I(V;Y) - I(V;S) &= h(V) - h(V|Y) - h(V) + h(V|S) \\ &= -h(X|X+Z) + h(X) \\ &= I(X;X+Z) \\ &= \log(1+\sigma_x^2/\sigma_z^2) \end{aligned} \tag{78}$$

Since the latter is equal to the capacity of the channel $Y = X + Z$, which is clearly an upperbound for any additive interference signal S independent of X and Z, we have that $C_q = \log(1 + \sigma_x^2/\sigma_z^2)$, and that this result is independent of q (which is arbitrary).

Then, by Lemma B.2 we have that $C = \inf_q C_q = \log(1 + \sigma_x^2/\sigma_z^2)$ for any arbitrary interference signal distribution (not necessarily i.i.d.). Theorem 1 is finally proved by observing that each one of the interference channels (6) generated by the QR decomposition $\mathbf{H} = \mathbf{GQ}$ and ranked in the order $k = 1, \ldots, m$ behaves like a channel with interference non-causally known to the transmitter. In fact, the transmitter selects the code word $\mathbf{u}^1$ for user 1. Then, it treats $\mathbf{s}^2 = g_{2,1}\mathbf{u}^1$ as known interference and selects the code word $\mathbf{u}^2$ for user 2. Next, it treats $\mathbf{s}^3 = g_{3,1}\mathbf{u}^1 + g_{3,2}\mathbf{u}^2$ as known interference and selects the code word $\mathbf{u}^3$ for user 3, and so on, till user m. For what said above, the rate $\log(1 + |g_{k,k}|^2 a_k)$ is achievable in the k-th channel, where $a_k = E[|u_{k,i}|^2]$ is the input constraint in the k-th channel (recall that the noise variance is normalized to 1 for all users). Finally, the rate-sum $\sum_{k=1}^m \log(1 + |g_{k,k}|^2 a_k)$ can be maximized with the constraint $\sum_{k=1}^m a_k \leq A$, $a_k \geq 0$, yielding (8). □

The modulo-lattice precoding scheme. Remarkably, the above proof implies that for the interference channel (73) with Gaussian i.i.d. noise there exist *universal* coding schemes achieving the AWGN capacity independently of the probability assignment of the interfering signal. One of such schemes has been recently proposed by Erez, Shamai and Zamir in [17], and it is based on *modulo-lattice precoding*. Let $\Lambda_n \subset \mathbb{C}^n$ be an n-dimensional lattice with fundamental Voronoi cell $\mathcal{V}$ with second-order moment σ_x^2. The codebook consists of points $\mathbf{v}$ generated with uniform distribution over $\mathcal{V}$. The encoder is aware of the interfering signal realization $\mathbf{s}$, and in order to send the code word $\mathbf{v}$ it transmits the signal

$$\mathbf{x} = [\mathbf{v} - \alpha\mathbf{s}]_{\Lambda_n} \tag{79}$$

where α is a scalar constant and $[\cdot]_{\Lambda_n}$ indicates a modulo-Λ_n operation, i.e., if we define $Q_{\Lambda_n} : \mathbb{C}^n \to \Lambda_n$ to be the lattice quantizer based on Λ_n, then $[\mathbf{u}]_{\Lambda_n} = \mathbf{u} - Q_{\Lambda_n}(\mathbf{u})$. The receiver performs decoding on the modulo-Λ_n signal $\mathbf{y}' = [\alpha\mathbf{y}]_{\Lambda_n}$, where $\mathbf{y} = \mathbf{x} + \mathbf{s} + \mathbf{z}$ is the channel output.

In order to analyze this scheme, we notice that the whole system is equivalent to the interference-free modulo-Λ_n channel

$$\mathbf{y}' = [\mathbf{v} + \mathbf{z}']_{\Lambda_n} \tag{80}$$

where $\mathbf{z}'$ has the same statistics of $[(1-\alpha)\mathbf{u} + \alpha\mathbf{z}]_{\Lambda_n}$, with $\mathbf{u}$ uniformly distributed over $\mathcal{V}$ and independent of $\mathbf{z}$ (see Lemma 2 of [17]). In fact, by (79) since $\mathbf{v}$ is uniformly distributed over $\mathcal{V}$ also $\mathbf{x}$ is, for any individual $\mathbf{s}$. By definition of the modulo-Λ_n operation, we can write $\mathbf{v} = \alpha\mathbf{s} + \mathbf{x} + \boldsymbol{\lambda}$ for some $\boldsymbol{\lambda} \in \Lambda_n$. We have also that

$$\begin{aligned} \mathbf{v} - \alpha\mathbf{y} &= \alpha\mathbf{s} + \mathbf{x} + \boldsymbol{\lambda} - \alpha(\mathbf{x} + \mathbf{s} + \mathbf{z}) \\ &= \mathbf{x} - \alpha(\mathbf{x} + \mathbf{z}) + \boldsymbol{\lambda} \end{aligned} \tag{81}$$

Hence, by applying $[\cdot]_{\Lambda_n}$ to both sides of (81) we get

$$[\mathbf{v} - \alpha\mathbf{y}]_{\Lambda_n} = [(1-\alpha)\mathbf{x} - \alpha\mathbf{z}]_{\Lambda_n}$$

which yields (80).

The achievable rate for channel (80) is given by

$$\begin{aligned} \frac{1}{n}I(\mathbf{v};\mathbf{y}') &= \frac{1}{n}(h(\mathbf{y}') - h(\mathbf{z}')) \\ &\to \log \sigma_x^2 - \log((1-\alpha)^2\sigma_x^2 + \alpha^2\sigma_z^2) \end{aligned}$$

where we have assumed that Λ_n for $n \to \infty$ is a sequence of optimal lattices, with shaping gain $G_n \to 1/(\pi e)$ and we used the fact that for such optimal lattices $\mathbf{u}$ uniformly distributed over $\mathcal{V}$ tends to be Gaussian (see [17] and references therein). Finally, by choosing $\alpha = \frac{\sigma_x^2}{\sigma_x^2+\sigma_z^2}$ we obtain $\frac{1}{n}I(\mathbf{v};\mathbf{y}') \to \log(1 + \sigma_x^2/\sigma_z^2)$ as desired.

Since the lattice-precoding scheme applies to any individual interference signal, it can be immediately extended to the RKI case, by applying it in sequences to the known-interference subchannels $k = 1, 2, \ldots, m$ in (6). A suboptimal one-dimensional version of the above lattice-precoding RKI strategy was independently proposed in [20, 21] for the purpose of canceling far-end cross-talk (FEXT) in the downstream of a DSL system by joint processing the transmit signals at the (common) transmitter). In the one-dimensional case, the interference subtraction followed by the modulo-lattice operation is analogous to Tomlinson-Harashima precoding, commonly used for the ISI channel [19], and suffers from inherent shaping and modulo loss [47].

Costa's "writing on dirty paper" scheme. In [15], under the suggestive title "Writing on dirty paper", Costa showed the achievability of (78) in the special case where the interference signal is Gaussian i.i.d., $S_i \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_s^2)$. Costa's scheme can be outlined as follows: define an auxiliary random variable $V = \alpha S + X$, where $X \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_x^2)$ is independent of S. Then,

- Construct a codebook of size $\exp(n(I(V;Y) - \epsilon))$ and block length n by random generation of sequences $\mathbf{v}$ with i.i.d. components $\sim \mathcal{N}_{\mathbb{C}}(0, \sigma_x^2 + \alpha^2\sigma_s^2)$.
- Distribute randomly with uniform probability the sequences $\mathbf{v}$ over e^{nR} bins. The codebook and bin assignment is revealed to the receiver.
- When message w is to be transmitted, the encoder searches in the w-th bin a sequence $\mathbf{v}$ such that $\mathbf{v}$ and the current realization of the interference signal $\mathbf{s}$ (which

is known non-causally to the transmitter) are jointly typical. If there is no such sequence or if $\frac{1}{n}|\mathbf{v} - \alpha\mathbf{s}|^2 > \sigma_x^2$ an error is declared, otherwise the encoder sends the difference signal $\mathbf{x} = \mathbf{v} - \alpha\mathbf{s}$.

- On the receiver side, the decoder looks for a sequence $\widehat{\mathbf{v}}$ jointly typical with the received sequence $\mathbf{y} = \mathbf{x} + \mathbf{s} + \mathbf{z}$. If this sequence is found and it is unique, the decoded message $\widehat{w}$ is given by the index of the bin that contains $\widehat{\mathbf{v}}$. If such sequence is not found or it is not unique, an error is declared.

If the number of sequences in each bin is larger than $\exp(n(I(V;S) + \epsilon))$, the probability of finding $\mathbf{v}$ jointly typical with $\mathbf{s}$ tends exponentially to 1 as $n \to \infty$. Then, if $R \leq I(V;Y) - I(V;S) - 2\epsilon$ the error probability of this scheme can be made arbitrarily small. Finally, by letting $\alpha = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_z^2}$ and by direct calculation it is easy to show that $I(V;Y) - I(V;S) = \log(1 + \sigma_x^2/\sigma_z^2)$.

In order to apply Costa's scheme to the RKI channels, we have to check that for each k-th channel the interference signal (generated by the random choice of the codewords in the previous channels $j = 1, \ldots, k-1$) is i.i.d. Gaussian with known variance. In fact, we should notice that even though the capacity formula is independent of the interference variance σ_s^2, the encoder must know it in order to generate the code book and check typicality. In the following, we show that Costa's scheme can indeed by applied to the RKI channels, thus achieving the same throughput of the lattice-precoding scheme. For two vectors $\mathbf{a}$ and $\mathbf{b}$ of length n we let $\langle \mathbf{a}, \mathbf{b} \rangle \stackrel{\Delta}{=} \frac{1}{n}\mathbf{a}^H\mathbf{b}$. In the Gaussian additive case it is worthwhile to notice that $\mathbf{v}$ and $\mathbf{s}$ being jointly typical is equivalent to the condition [5]

$$\begin{aligned} \langle \mathbf{v}, \mathbf{v} \rangle &\to \sigma_x^2 + \alpha^2\sigma_s^2 \\ \langle \mathbf{s}, \mathbf{s} \rangle &\to \sigma_s^2 \\ \langle \mathbf{v} - \alpha\mathbf{s}, \mathbf{s} \rangle &\to 0 \end{aligned} \tag{82}$$

This also implies that $\langle \mathbf{x}, \mathbf{x} \rangle \to \sigma_x^2$, so that the input constraint is satisfied with arbitrarily large probability.

Now, consider the channel $Y = X + S_1 + S_2 + N$ where S_1 and S_2 are independent Gaussian interference signals known non-causally to the transmitter, with powers σ_{s1}^2 and σ_{s2}^2 such that $\sigma_{s1}^2 + \sigma_{s2}^2 = \sigma_s^2$. This is clearly equivalent to Costa's channel by letting $S = S_1 + S_2$. We repeat the code construction of above, by defining the auxiliary random variable $V = X + \alpha(S_1 + S_2)$ where $X \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_x^2)$ is independent of S_1 and S_2. The transmitter now looks for a sequence $\mathbf{v}$ in the w-th bin such that $\mathbf{v}, \mathbf{s}_1, \mathbf{s}_2$ are jointly typical and transmits $\mathbf{x} = \mathbf{v} - \alpha(\mathbf{s}_1 + \mathbf{s}_2)$. Since

$$I(V; S_1, S_2) = I(V; S_1 + S_2) = I(V; S)$$

this can be found with arbitrarily large probability provided that $R \leq I(V;Y) - I(V;S) - 2\epsilon$, i.e., there is no loss in coding rate with respect to original Costa's case. The joint

[5] Limits hold with probability 1, although convergence in probability is sufficient to prove achievability [26].

typicality implies that $\langle \mathbf{v} - \alpha \mathbf{s}_1, \mathbf{s}_1 \rangle \to 0$, $\langle \mathbf{v} - \alpha \mathbf{s}_2, \mathbf{s}_2 \rangle \to 0$ and $\langle \mathbf{s}_1, \mathbf{s}_2 \rangle \to 0$, which yield $\langle \mathbf{x}, \mathbf{s}_1 \rangle \to 0$ and $\langle \mathbf{x}, \mathbf{s}_2 \rangle \to 0$. Then, as a byproduct, we get that $\mathbf{x}$ is asymptotically uncorrelated with any linear combination (with bounded coefficients) of the signals $\mathbf{s}_1$ and $\mathbf{s}_2$. This simple observation is the key to extend Costa's scheme to the RKI case.

Consider the RKI channels given in (6). Without loss of generality each receiver k can divide its received signal by $g_{k,k}$ (assumed non-zero) so that we obtain

$$\mathbf{y}^k = \mathbf{u}^k + \sum_{j<k} \mathbf{s}_k^j + \mathbf{z}^k, \quad k = 1, \ldots, m \tag{83}$$

where we let $\mathbf{s}_k^j = (g_{k,j}/g_{k,k})\mathbf{u}^j$. For all $k = 1, \ldots, m$, we define the mutually independent random variables $U_k \sim \mathcal{N}_{\mathbb{C}}(0, a_k)$, the interference random variables $S_{k,j} = (g_{k,j}/g_{k,k})U_j$ for $j < k$, the interference channels $Y_k = U_k + \sum_{j<k} S_{k,j} + Z_k'$, with $Z_k' \sim \mathcal{N}_{\mathbb{C}}(0, 1/|g_{k,k}|^2)$ and the auxiliary random variables $V_k = U_k + \alpha_k \sum_{j<k} S_{k,j}$.

The coding scheme is obtained as follows. For all $k = 1, \ldots, m$, a set of $\exp(n(I(V_k; Y_k) - \epsilon))$ Gaussian sequences are generated with i.i.d. components $\sim \mathcal{N}_{\mathbb{C}}\left(0, a_k + \alpha_k^2 \sum_{j<k} |g_{k,j}/g_{k,k}|^2 a_j\right)$. Each k-th set of code words is distributed with uniform probability over $\exp(nR_k)$ bins and is revealed to user k receiver (users do not need to know the other users code books). In order to transmit an m-tuple of messages $(w_1, \ldots, w_m)$, the encoder consider the users in sequence. For each user k, is looks for a code word $\mathbf{v}^k$ in the w_k-th bin such that $\mathbf{v}^k, \mathbf{s}_k^1, \ldots, \mathbf{s}_k^{k-1}$ are jointly typical. Then, it sends the difference signal $\mathbf{u}^k = \mathbf{v}^k - \alpha_k \sum_{j<k} \mathbf{s}_k^j$. If any of these code words is not found, or if any of the interfering signals is not typical, an error is declared. The signals $\mathbf{u}^1, \ldots, \mathbf{u}^m$ are sent to the unitary precoder which transmits the code word

$$\mathbf{X} = \mathbf{Q}^H \begin{bmatrix} \mathbf{u}^1 \\ \mathbf{u}^2 \\ \vdots \\ \mathbf{u}^m \end{bmatrix}$$

Each k-th receiver looks for the code word $\widehat{\mathbf{v}}^k$ jointly typical with its own received signal $\mathbf{y}^k$, and outputs the index of the bin that contains $\widehat{\mathbf{v}}^k$.

Now, from the observation made before, every $\mathbf{u}^k$ is asymptotically uncorrelated with each of the $\mathbf{s}_k^j : j < k$. This implies that for any choice of the coefficients $\beta_1, \ldots, \beta_k$ (such that $\sum_{j=1}^k |\beta_j|^2 < \infty$),

$$\frac{1}{n} \left| \sum_{j \le k} \beta_j \mathbf{u}^j \right|^2 \to \sum_{j \le k} |\beta_j|^2 a_j$$

In particular, since the interference signal "seen" at level $k+1$ is a linear combination of the signals $\mathbf{u}^1, \ldots, \mathbf{u}^k$, its variance is arbitrarily close to $\sum_{j \le k} |g_{k+1,j}/g_{k+1,k+1}|^2 a_j$ with arbitrarily high probability for all (typical) choice of the code words $\mathbf{v}^1, \ldots, \mathbf{v}^k$. Since at each level k the interference signal is typical, the probability of finding $\mathbf{v}^k$ in bin w_k jointly typical with the k-th interference signal can be made arbitrarily large provided that $R_k \leq I(V_k; Y_k) - I(V_k; \sum_{j<k} S_{k,j}) - 2\epsilon$. The choice $\alpha_k = \frac{d_k a_i}{1+d_k a_i}$ (where $d_k = |g_{k,k}|^2$ as defined in Section 3) yields the maximum achievable user rates $R_k = \log(1 + d_k a_k)$, and the rate-sum can be maximized subject to the constraint $\sum_{k=1}^{m} a_k \leq A$, $a_k \geq 0$, thus obtained the throughput of Theorem 1.

C Proof of Lemma 3

Without loss of generality, we can reorder the rows of $\mathbf{H}_1$ and $\mathbf{H}_2$ so that the rows to be exchanged are the first (say) ℓ. Then,

$$\begin{aligned} \mathbf{H}'_1 &= \mathbf{M}_1\mathbf{H}_1 + \mathbf{M}_2\mathbf{H}_2 \\ \mathbf{H}'_2 &= \mathbf{N}_1\mathbf{H}_2 + \mathbf{N}_2\mathbf{H}_1 \end{aligned}$$

with

$$\mathbf{M}_1 = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{m-\ell} \end{bmatrix}, \quad \mathbf{M}_2 = \begin{bmatrix} \mathbf{I}_\ell & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$
$$\mathbf{N}_1 = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{r-m-\ell} \end{bmatrix}, \quad \mathbf{N}_2 = \begin{bmatrix} \mathbf{I}_\ell & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

and where the dimensions of all submatrices are appropriately chosen (clearly, $\ell \leq \min\{m, r-m\}$). By substituting $\mathbf{H}_2 = \mathbf{B}\mathbf{H}_1$ with $\mathbf{B} = \mathbf{H}_2\mathbf{H}_1^+$ into the above equations we obtain

$$\mathbf{B}' = (\mathbf{N}_1\mathbf{B} + \mathbf{N}_2)(\mathbf{M}_1 + \mathbf{M}_2\mathbf{B})^{-1}$$

where the inverse exists by the assumption that $\mathbf{H}'_1$ has rank m, i.e., $\mathbf{H}'_1 = (\mathbf{M}_1 + \mathbf{M}_2\mathbf{B})\mathbf{H}_1$ is just a change of coordinates in the m-dimensional space spanned by the rows of $\mathbf{H}_1$. Now, we partition $\mathbf{B}$ in the 2×2 block form

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}$$

where $\mathbf{B}_{11} \in \mathbb{C}^{\ell \times \ell}$ is invertible, since

$$(\mathbf{M}_1 + \mathbf{M}_2\mathbf{B})^{-1} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{0} & \mathbf{I}_{m-\ell} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{B}_{11}^{-1} & -\mathbf{B}_{11}^{-1}\mathbf{B}_{12} \\ \mathbf{0} & \mathbf{I}_{m-\ell} \end{bmatrix}$$

After some algebra, we obtain the upper-left $\ell \times \ell$ block of the matrix $\mathbf{B}'(\mathbf{B}')^H$ in the form

$$\mathbf{B}_{11}^{-1}\left(\mathbf{I}_\ell + \mathbf{B}_{12}\mathbf{B}_{12}^H\right)\mathbf{B}_{11}^{-H}$$

Since for any Hermitian symmetric matrix we have that the maximum eigenvalue is not smaller than the maximum diagonal element [48], we have that

$$\rho\left(\mathbf{I}_\ell + \mathbf{B}_{12}\mathbf{B}_{12}^H\right) \geq 1 \tag{84}$$

By using the submultiplicative property of the matrix 2-norm [12] we can write

$$\|\mathbf{B}_{11}\|_2^2 \|\mathbf{B}_{11}^{-1}\left(\mathbf{I}_\ell + \mathbf{B}_{12}\mathbf{B}_{12}^H\right)\mathbf{B}_{11}^{-H}\|_2 \geq \|\mathbf{I}_\ell + \mathbf{B}_{12}\mathbf{B}_{12}^H\|_2 \geq 1$$

where the last inequality follows from (84). From the fact that $\|\mathbf{B}\|_2 \leq 1$ it is immediate to see that also $\|\mathbf{B}_{11}\|_2 \leq 1$. This yields

$$\left\|\mathbf{B}_{11}^{-1}\left(\mathbf{I}_\ell + \mathbf{B}_{12}\mathbf{B}_{12}^H\right)\mathbf{B}_{11}^{-H}\right\| \geq \frac{1}{\|\mathbf{B}_{11}\|_2^2} \geq 1$$

which in turns implies that $\|\mathbf{B}'\|_2 \geq 1$, as desired. □

References

[1] S. Shamai and A. Wyner, "Information theoretic considerations for symmetric cellular, multiple-access fading channels – Part I and II," *IEEE Trans. on Inform. Theory*, vol. 43, no. 6, pp. 1877–1911, November 1997.

[2] S. Shamai and B. Zaidel, "Enhancing the cellular downlink capacity via co-processing at the transmitter end," in *IEEE Semiannual Veh. Tech. Conf. VTC, spring*, Rhodes, Greece, May 2001.

[3] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. on Telecomm. ETT*, vol. 10, no. 6, pp. 585–596, November 1999.

[4] A. Barreto and G. Fettweis, "On the downlink capacity of TDD CDMA systems using a pre-RAKE," in *GLOBECOM '99*, Rio de Janeiro, Brazil, December 1999.

[5] J. Proakis, *Digital communications, 3rd Ed.*, McGraw-Hill, New York, 1995.

[6] B. Vojcic and W. Mee Jang, "Transmitter precoding in synchronous multiuser communications," *IEEE Trans. on Commun.*, vol. 46, no. 10, pp. 1346–1355, October 1998.

[7] M. Brandt-Pearce and A. Dharap, "Transmitter-based multiuser interference rejection for the down-link of a wireless CDMA system in a multipath environment," *IEEE J. Select. Areas Commun.*, vol. 18, no. 3, pp. 407–417, March 2000.

[8] P. Baier, M. Meurer, T. Weber, and H. Troeger, "Joint transmission (JT), an alternative rationale for the downlink of time division CDMA using multi-element transmit antennas," in *IEEE 6th Int. Symp. on Spread-Spectrum Tech. & Appl. (ISSSTA)*, NJIT, New Jersey, USA, September 2000, pp. 1–5.

[9] F. Kowalewski and P. Mangold, "Joint predistortion and transmit diversity," in *GLOBECOM '00*, San Francisco, CA, USA, December 2000.

[10] T. Luo, E. Sousa, and S. Pasupathy, "Transmit beamforming and power control in downlink channels of a multi-rate CDMA communication system," in *ICC 2000*, New Orleans, Louisiana, USA, June 2000.

[11] A. Barreto and G. Fettweis, "Capacity increase in the downlink of spread spectrum systems through joint signal precoding," in *ICC 2001*, Helsinki, Finland, June 2001.

[12] R. A. Horn and C. R. Johnson, *Matrix analysis*, Cambridge University Press, New York, 1985.

[13] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance for block-fading channels with multiple antennas," *IEEE Trans. on Inform. Theory*, vol. 47, no. 4, pp. 1273–1289, May 2001.

[14] S. Gelfand and M. Pinsker, "Coding for channel with random parameters," *Problems of Control and Information Theory*, vol. 9, no. 1, pp. 19–31, January 1980.

[15] M. Costa, "Writing on dirty paper," *IEEE Trans. on Inform. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[16] R. Ahlswede, "Arbitrarily varying channels with states sequence known to the sender," *IEEE Trans. on Inform. Theory*, vol. 32, no. 5, pp. 621–629, September 1986.

[17] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice-strategies for cancelling known interference," in *ISITA 2000*, Honolulu, Hawaii, USA, November 2000.

[18] A. Cohen and A. Lapidoth, "The Gaussian watermarking game - Part I," to appear on IEEE Trans. on Inform. Theory, 2001.

[19] V. Eyuboglu and D. Forney, "Combined equalization and coding using precoding," *IEEE Comm. Magazine*, December 1991.

[20] G. Ginis and J. Cioffi, "Vectored-DMT: A FEXT canceling modulation scheme for coordinating users," in *ICC 2001*, Helsinki, Finland, June 2001.

[21] G. Ginis and J. Cioffi, "A multi-user precoding scheme achieving crosstalk cancellation with application to DSL systems," in *34-th Asilomar Conference on Signals, Systems and Computers*, Pacific-Groove, CA, USA, November 2000.

[22] A. Goldsmith and M. Effros, "The capacity region of broadcast channels with intersymbol interference and colored Gaussian noise," *IEEE Trans. on Inform. Theory*, vol. 47, no. 1, pp. 219–240, January 2001.

[23] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channel," *IEEE Trans. on Inform. Theory*, vol. 45, no. 5, pp. 1468–1489, July 1999.

[24] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: information-theoretic and communications aspects," *IEEE Trans. on Inform. Theory*, vol. 44, no. 6, pp. 2619–2692, October 1998.

[25] S. Verdú and T.S. Han, "A general formula for channel capacity," *IEEE Trans. on Inform. Theory*, vol. 40, no. 4, pp. 1147–1157, July 1994.

[26] T. Cover and J. Thomas, *Elements of information theory*, Wiley, New York, 1991.

[27] D. Tse, "Optimal power allocation over parallel Gaussian broadcast channels," submitted to IEEE Trans. on Inform. Theory, October 1998.

[28] L. Li and A. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels.I. Ergodic capacity," *IEEE Trans. on Inform. Theory*, vol. 47, no. 3, pp. 1083–1102, March 2001.

[29] G. Poltyrev, "Capacity for a sum of certain broadcast channels," *Probl. Pered. Inform.*, vol. 15, no. 2, pp. 40–44, 1979.

[30] A. El Gamal, "Capacity for the product and sum of two unmatched broadcast channels," *Probl. Pered. Inform.*, vol. 16, no. 1, pp. 3–23, 1980.

[31] D. Hughes-Hartogs, "The capacity of a degraded spectral Gaussian broadcast channel," Ph.D. Thesis, Stanform Univ., USA, July 1975.

[32] R. Gallager, *Information theory and reliable communication*, Wiley, New York, 1968.

[33] H. Sato, "An outer bound on the capacity region of broadcast channels," *IEEE Trans. on Inform. Theory*, vol. 24, no. 3, pp. 374–377, May 1978.

[34] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, and A. Viterbi, "CDMA/HDR: A bandwith-efficient high-speed wireless data service for nomadic users," *IEEE Commun. Magazine*, vol. 38, no. 7, pp. 70–77, July 2000.

[35] S. Borst and P. Whiting, "Dynamic rate control algorithms for HDR throughput optimization," in *INFOCOM 2001*, 2000, vol. 2, pp. 976–985.

[36] A. Lapidoth, "Nearest neighbor decoding for additive non-Gaussian noise channels," *IEEE Trans. on Inform. Theory*, vol. 42, no. 5, pp. 1520–1529, September 1996.

[37] S. Verdu, "Spectral efficiency in the wideband regime," to appear on IEEE Trans. on Inform. Theory, 2001.

[38] S. Verdu and S. Shamai, "The impact of frequency-flat fading on the spectral efficiency of CDMA," *IEEE Trans. on Inform. Theory*, vol. 47, no. 4, pp. 1302–1327, May 2001.

[39] A. Edelman, "Eigenvalues and condition numbers of random matrices," Ph.D. Thesis, M.I.T., Cambridge MA, USA, May 1989.

[40] D. Tse and S. Hanly, "Linear multiuser receivers: Effective interference, effective bandwidth and capacity," *IEEE Trans. on Inform. Theory*, vol. 45, no. 2, pp. 641–675, March 1999.

[41] S. Verdu and S. Shamai, "Spectral efficiency of CDMA with random spreading," *IEEE Trans. on Inform. Theory*, vol. 45, no. 2, pp. 622–640, March 1999.

[42] R. Mueller, "Multiuser receivers for randomly spread signals: fundamental limits with and without decision-feedback," *IEEE Trans. on Inform. Theory*, vol. 47, no. 1, pp. 268–283, January 2001.

[43] M.-S. Alouini and A. Goldsmith, "Capacity of Rayleigh fading channels under different adaptive transmission and diversity-combining techniques," *IEEE Trans. on Vehic. Tech.*, vol. 48, no. 4, pp. 1165–1181, July 1999.

[44] R. Knopp and G. Caire, "Power control and beamforming for systems with multiple transmit and receive antennas," to appear on IEEE J. Sel. Areas on Commun., 2001.

[45] L. Brandenburg and A. Wyner, "Capacity of the Gaussian channel with memory: the multivariate case," *Bell Syst. Tech. J.*, vol. 53, pp. 745–779, May 1974.

[46] W. Hirt and J. Massey, "Capacity of the discrete-time Gaussian channel with intersymbol interference," *IEEE Trans. on Inform. Theory*, vol. 34, no. 3, pp. 380–388, May 1988.

[47] R. Wesel and J. Cioffi, "Achievable rates for Tomlinson-Harashima precoding," *IEEE Trans. on Inform. Theory*, vol. 44, no. 5, pp. 825–831, September 1998.

[48] A. Marshall and I. Olkin, *Inequalities: Theory of majorization and its applications*, Academic Press, London, UK, 1979.