Variable-rate Coding for Slowly-fading Gaussian Multiple-access Channels

Giuseppe Caire, Daniela Tuninetti and Sergio Verdú

September 16, 2003

Abstract

We consider a non-ergodic multiple access Gaussian block-fading channel where a fixed number of independent identically distributed fading coefficients affect each codeword. Variable-rate coding with input power constraint enforced on a per-codeword basis is examined. A centralized power and rate allocation policy is determined as a function of the previous and present fading coefficients. The power control policy that optimizes the expected rates is obtained through dynamic programming and the average capacity region and the average capacity region per unit energy are characterized. Moreover, we study the slope of spectral efficiency curve vs. (E_b/N_0)dB, and we quantify the penalty incurred by TDMA over superposition coding in the low power regime.

Keywords: Channel Capacity, Fading Channels, Multiple Access Channels, Power Control, Low Power Regime, Causal Channel State Information.

1   Introduction and motivations

In this paper we present results on fading multiple access systems employing variable-rate coding. Our setting captures the characteristic, common to several real-time services, of requiring *reliable* transmission in any channel conditions. Examples of this class of applications are real-time video streaming, for which the reception of erroneous packets may cause discontinuity in the service, and sensor networks, where measurements must be delivered at regular time intervals at possibly different rates.

We assume the channel to be *frequency non-selective and slowly varying* (i.e., the channel coherence bandwidth and the channel coherence time are larger than, respectively,

the bandwidth and time duration of the transmit signals). We use the popular *block-fading* channel model [13] in which the time axis is divided into equal-length slots and each slot is affected by one fading coefficient. The fading coefficient, or channel state, remains constant over the whole slot and varies independently from slot to slot. In practical systems, the independence assumption is motivated by time and/or by frequency hopping. Moreover, we assume that each slot has large enough bandwidth time-duration product so as to guarantee a certain desired level of reliability against the additive noise.

We also assume that codewords span a fixed number of slots. At the end of a block of N slots, decoding must be performed. The system parameter N, common to all the users, can be used to model the time-sensitivity of information that becomes "useless" if received after a given time since its generation. It can also be used to model the expansion factor in degrees of freedom that the designer is willing to incur in order to combat fading variations. Notice that since each slot of each user is affected by a single fading coefficient the channel is non-ergodic: the fading statistics are not revealed within the span of each codeword for any finite N.

The information-theoretic literature on fading channels has adopted various ways to characterize power constraints, foremost among those:

A Power constraint on a per-symbol basis.

B Power constraint on a per-codeword basis.

C Power constraint on an arbitrarily long sequence of codewords.

In the above cases, power is typically averaged over of the codebook. Constraint B is preferable over C in many practical systems in which the transmit power cannot fluctuate arbitrarily from codeword to codeword whenever power cannot be amortized over a long horizon.

Basic information theoretic results [19, 7] have shown that the constraints B and C offer no advantage in either unfaded channels or in fading channels where the transmitter does not know the channel. However, when the transmitter has instantaneous knowledge of the channel fading coefficients, constraints B and C lead to strictly larger capacities than A because they enable the use of "power control" which avoids wasting power at symbols where the channel undergoes deep fades. In ergodic settings, constraints B and C result in the same power control policy, e.g., in the single-user scalar (ergodic) Gaussian-noise fading channels, the optimal power allocation strategy is water-filling in time [1]. Under A the optimal power policy is constant power allocation. Although the different

constraints lead to different optimum transmission strategies, in the high spectral efficiency regime they achieve very similar single-user ergodic capacity. Only in conjunction with multiaccess and multiuser detection do optimum power control strategies lead to noticeable advantages in the high SNR (Signal to Noise Ratio) regime [14, 5]. On the other hand, in the low spectral efficiency regime, constraint B enables (for fading distributions with infinite support) reliable communication with energy per bit as small as desired [18]. This is in stark contrast to constraint A, which requires a minimum transmitted energy per bit equal to -1.59 dB [18].

In non-ergodic channels, constraints B and C lead to different power allocation strategies. In [16] the concept of "delay-limited" capacity region for multiaccess fading channels is introduced. In this setting, each codeword spans a single fading state and the input power constraint enforced is C. The reliable decoded rates are fixed while the transmit power fluctuates from codeword to codeword. The delay-limited capacity region is the set of rates which can be achieved *for all* fading states (up to a set of measure zero). In the single-user scalar case, the optimal power policy is "channel inversion", i.e., the SNR at the receiver is maintained constant by appropriate compensation at the transmitter. If, instead, the power constraint enforced were B, then only the rate corresponding to the least-favorable fading state could be guaranteed. For example, for Rayleigh fading no positive rate can be guaranteed with finite power under constraint C (and *a fortiori* under B).

Another way to characterize the performance of non-ergodic channels is by means of the ϵ-capacity [3, 8]. This approach, also referred to as capacity vs. outage [13, 11, 12], allows decoding failure with non-negligible probability. The power allocation policy has the objective to maximize the transmission rate for a given outage probability ϵ. As in the delay-limited setting, the transmit power responds to the fading fluctuations but the transmission rates remain constant. In the single-user scalar case with codewords spanning a single fading state, the optimal policy under B is constant power allocation while under C is truncated channel inversion, i.e., the fading is compensated for only if it is not too severe [8].

In this paper we take a best-effort approach that complements the delay-limited and outage approaches: we allow the transmit coding rates to vary according to the channel conditions while enforcing arbitrarily reliable communication. The goal of the encoder/decoder is to maximize the expected rate of reliable information transfer within each codeword subject to an average power constraint on a per-codeword basis (constraint B). A centralized controller that knows the previous and current fading realizations affecting all users (e.g. the receiver) determines the rate and power to be used by each user

at each slot. The resulting transmission rates vary from codeword to codeword and are function of the actual realization of N fading coefficients. The single-user version of this problem has been considered in [15]. If the N fading coefficients were known non-causally at the sender side, the best-effort variable-rate policy under constraint B would be the allocation of the whole available power on the slot with largest gain. However, the causal channel knowledge prevents this simple strategy.

Notice that the maximization of the average rate under constraint C, with or without causal channel knowledge, results in the optimal ergodic power allocation policy derived in [5].

As shown recently in [18], the minimum energy per bit, on which traditionally information theoretic analysis of the low spectral efficiency regime has focused, fails to capture the fundamental power-bandwidth tradeoff. To study that tradeoff it is necessary to analyze not only the minimum energy per bit, but also the "slope" of spectral efficiency vs. (E_b/N_0) (expressed in b/s/Hz/3 dB) at the point of minimum energy per bit. Accordingly, our analysis focuses on both fundamental limits: we make use of the framework developed in [17] for the capacity-per-unit-cost region for multiaccess channels as well as results on the wideband slope region, following the approach of [20].

We show that a "one-shot" power allocation policy, that concentrates the whole transmit energy over one out of N slots, yields both optimal minimum energy per bit and optimal wideband slope. Since such slot must be chosen on the basis of causal feedback, the transmitter cannot simply choose the most favorable slot in the codeword. Rather, the solution obtained through dynamic programming has the structure of a comparison of the instantaneous fading amplitude with a decreasing threshold function. The determination of the optimal power policy with causal channel knowledge requires the solution of a dynamic program whose closed-form solution is not known for arbitrary SNR even in the single-user case [15].

Interestingly, we show that TDMA in conjunction with the one-shot power policy suffices to achieve the capacity region per unit energy but is strictly suboptimal in terms of wideband slope for any non-degenerate fading distribution. On the contrary, superposition coding with successive interference cancellation at the receiver, in conjunction with the one-shot power policy, achieves both the capacity region per unit energy and the optimal wideband slope.

The paper is organized as follows: Section 2 gives a description of the system model and defines the variable rate coding scheme; Section 3 characterizes the average capacity region. As a byproduct of our results we show that by placing the additional constraint that reliable decisions be made at the end of each slot there is no loss in maximal ex-

pected rate. Section 4 specifies the average capacity region per unit energy including the asymptotic form of the dynamic programming power allocation strategy. The asymptotic optimality (in terms of wideband slope) is proved in Section 5, which also considers the performance of TDMA in the low-power regime. As a baseline of comparison a low-SNR analysis of the optimal non-causal policy is given in Section 6. Section 7 particularizes the results for the Rayleigh fading case.

2 System model and basic definitions

We consider a Gaussian Multiple Access Channel (MAC) where K transmitters must deliver their message to a central receiver by spending a fixed maximum energy per codeword. The propagation channel is modeled as frequency-flat block-fading. The fading gain of each user remains constant for a time slot of duration T seconds and changes independently in the next slot. The number of complex dimensions per slot is $L = \lfloor WT \rfloor$ where W is the channel bandwidth in Hz. For the block-fading assumption to be valid, T and W must be smaller than, respectively, the fading coherence time and the fading coherence bandwidth [10]. The baseband complex received vector in slot n is

$$\boldsymbol{y}_n = \sum_{k=1}^{K} c_{k,n} \boldsymbol{x}_{k,n} + \boldsymbol{z}_n \tag{1}$$

where $\boldsymbol{z}_n$ is a proper complex Gaussian random vector of dimension L with i.i.d. (independent and identically distributed) components of zero mean and unit variance, $\boldsymbol{x}_{k,n}$ is the length-L complex vector of symbols sent by user k in slot n and $c_{k,n}$ is the scalar complex fading coefficient affecting the transmission of user k in slot n. The cdf (cumulative distribution function) of the instantaneous fading powers $\alpha_{k,n} \triangleq |c_{k,n}|^2$, $F_\alpha^{(k)}(x) = \Pr[\alpha_{k,n} \leq x]$ is assumed to be a *continuous* function.

The codewords of all K users are synchronized and span a fixed number N of slots. Each codeword of length N slots is subject to the input constraint

$$\frac{1}{NL} \sum_{n=1}^{N} \|\boldsymbol{x}_{k,n}\|^2 \leq \gamma_k \tag{2}$$

where γ_k is the average transmit energy per coded symbol. Because of the noise variance normalization adopted here, γ_k has the meaning of average *transmit* Signal-to-Noise Ratio (SNR). [1]

---

[1]Note that the actual transmitted SNR is equal to γ_k/N_0 as the noise power is $P_n = N_0 W$ [Watt] and

The receiver has perfect Channel State Information (CSI)[2] and determines the rate and power allocated to each user at slot n on the basis of the history of the channel state up to time n, $\mathcal{F}_n$, defined as

$$\mathcal{F}_n \triangleq \{c_{k,i} : k = 1, \cdots, K, \ i = 1, \cdots, n\} \tag{3}$$

. Due to the causality constraint, the *instantaneous* transmit SNR of user k in slot n, indicated by $\beta_{k,n}$,

$$\beta_{k,n} \triangleq \frac{1}{L} \|\boldsymbol{x}_{k,n}\|^2, \tag{4}$$

can depend only on $\mathcal{F}_n$. Therefore, the input constraint in (2) can be re-written as

$$\frac{1}{N} \sum_{n=1}^{N} \beta_{k,n}(\mathcal{F}_n) \leq \gamma_k \tag{5}$$

For finite N and L no positive rate is achievable with arbitrary reliability. However, we can consider a sequence of channels indexed by the slot length L and study the achievable rates in the limit for $L \rightarrow \infty$ and fixed N. This is a standard mathematical abstraction in the study of the limit performance of block-fading channels [13] and it is motivated by the fact that, in many practical applications, the product $L = WT$ is large and sufficient to average out the additive noise. Note that for the power (joules/sec) and rate (bits/sec) not to grow without bound as the number of degrees of freedom grows, T must be allowed to be sufficiently large.

Even in the limit of large L, the rate K-tuple at which reliable communication is possible over a codeword of N slots is a random vector, because only a finite number N of fading coefficients affects each codeword. This means that, for fading processes with non-vanishing cdf in an interval around the origin, the counterpart of the delay-limited capacity obtained enforcing (2) would be zero.

We assume that transmitters have infinite "bit-reservoirs" and transmit variable numbers of bits per codeword, which depend on the KN fading coefficients affecting the K codewords. Therefore, at the end of each transmission the number of bits delivered to the receiver is a random variable. Because the transmission rates are chosen so that reliable decoding is always possible, the system is never in outage.

The largest *average rate region* achievable with variable-rate coding when each codeword is subject to the power constraint in (5) is the subject of the next section.

---

the user k signal power is $P_k = \frac{NL\gamma_k}{NT}$ [Watt], where N_0 [Watt/Hz] is the power spectral density of the additive noise, W [Hz] is the channel bandwidth, $NL\gamma_k$ [Joule] is total energy of the k-th user codeword and TN [seconds] is the total codeword duration.

[2]Because each slot contains a number of degrees of freedom that grows without bound, dropping this assumption has no effect on the capacity [6].

3 The average capacity region

The average capacity region is the set of average achievable rates defined in Appendix A and admits the following characterization.

Theorem 1. The average capacity region achieved with $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)$ is given by

$$C_{K,N}(\boldsymbol{\gamma}) = \bigcup_{\boldsymbol{\beta} \in \Gamma_{K,N}(\boldsymbol{\gamma})} \left\{ \boldsymbol{R} \in \mathbb{R}_+^K \ : \ \forall \mathcal{A} \subseteq \{1, \cdots, K\} \right.$$
$$\left. \sum_{k \in \mathcal{A}} R_k \leq \mathrm{E} \left[\frac{1}{N} \sum_{n=1}^N \log \left(1 + \sum_{k \in \mathcal{A}} \alpha_{k,n} \beta_{k,n}(\mathcal{F}_n) \right) \right] \right\} \tag{6}$$

where the expectation is taken with respect to $\mathcal{F}_N$ and where $\Gamma_{K,N}(\boldsymbol{\gamma})$ denotes the set of all power allocation policies satisfying the causality constraint in (5).

Proof. See Appendix B. $\qquad \square$

The region $C_{K,N}(\boldsymbol{\gamma})$ is convex in $\boldsymbol{\gamma}$. By applying Jensen's inequality it is straightforward to see that if $\boldsymbol{R}^{(a)} \in C_{K,N}(\boldsymbol{\gamma})$ and $\boldsymbol{R}^{(b)} \in C_{K,N}(\boldsymbol{\gamma})$ then, for every $\lambda \in [0,1]$ $\lambda \boldsymbol{R}^{(a)} + (1 - \lambda) \boldsymbol{R}^{(b)} \in C_{K,N}(\boldsymbol{\gamma})$. For this reason the convex hull operation is not needed in (6).

The boundary surface of the region $C_{K,N}(\boldsymbol{\gamma})$ is the convex closure of all K-tuples $\boldsymbol{R} \in \mathbb{R}_+^K$ that solve [5]

$$\max_{\boldsymbol{R} \in C_{K,N}(\boldsymbol{\gamma})} \sum_{k=1}^K \mu_k R_k \tag{7}$$

for some $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_K) \in \mathbb{R}_+^K$. It is easy to see that the set of average rates achievable by any fixed power policy $\boldsymbol{\beta} \in \Gamma_{K,N}(\boldsymbol{\gamma})$ is a polymatroid [5]. Hence, the optimization in (7) is equivalent to the optimization over $\boldsymbol{\beta} \in \Gamma_{K,N}(\boldsymbol{\gamma})$ of the functional

$$\sum_{k=1}^K \mu_{\pi_k} \mathrm{E} \left[\frac{1}{N} \sum_{n=1}^N \log \left(1 + \frac{\alpha_{\pi_k,n} \beta_{\pi_k,n}}{1 + \sum_{j<k} \alpha_{\pi_j,n} \beta_{\pi_j,n}} \right) \right] \tag{8}$$

where $\boldsymbol{\pi}$ is the permutation of $\{1, 2, \cdots, K\}$ such that $\mu_{\pi_1} > \cdots > \mu_{\pi_K}$ which corresponds to the decoding order $\pi_K, \pi_{K-1}, \cdots, \pi_1$. The optimization in (8) is a dynamic program solved by:

Theorem 2. The boundary surface of $C_{K,N}(\boldsymbol{\gamma})$ is the convex closure of the set

$$\left\{ \widehat{\boldsymbol{R}}_N(\boldsymbol{\mu}, \boldsymbol{\gamma}) \ : \quad \boldsymbol{\mu} \in \mathbb{R}_+^K, \ \sum_{k=1}^K \mu_k = 1 \right\} \tag{9}$$

where the k-th component of the rate K-tuple $\widehat{\boldsymbol{R}}_N(\boldsymbol{\mu}, \boldsymbol{\gamma})$ is given by

$$\widehat{R}_{k,N}(\boldsymbol{\mu}, \boldsymbol{\gamma}) = \mathrm{E}\left[\frac{1}{N}\sum_{n=1}^{N}\log\left(1 + \frac{\alpha_{k,n}\,\widehat{\beta}_{k,n}(\mathcal{F}_n; \boldsymbol{\mu}, \boldsymbol{\gamma})}{1 + \sum_{j < \pi^{-1}(k)}\alpha_{\pi_j,n}\,\widehat{\beta}_{\pi_j,n}(\mathcal{F}_n; \boldsymbol{\mu}, \boldsymbol{\gamma})}\right)\right] \quad (10)$$

($\pi^{-1}(k)$ gives the position of index k in the permuted vector $\boldsymbol{\pi}$) and where $\widehat{\beta}_{k,n}(\mathcal{F}_n; \boldsymbol{\mu}, \boldsymbol{\gamma})$, for all n and k, is given by the following dynamic programming recursion:

Let $\mathbf{P} = (P_1, \ldots, P_K)$ denote the users' energy (per L-symbols) available at any given slot. For $n = 1, \ldots, N$, define recursively the functions $S_n(\mathbf{P}; \boldsymbol{\mu})$ by

$$S_n(\mathbf{P}; \boldsymbol{\mu}) = \mathrm{E}\left[\max_{\substack{\boldsymbol{p}\in\mathbb{R}_+^K \\ p_k \leq P_k, \forall k}}\left\{\sum_{k=1}^{K}\mu_{\pi_k}\log\left(1 + \frac{\alpha_{\pi_k}\,p_{\pi_k}}{1 + \sum_{j<k}\alpha_{\pi_j}\,p_{\pi_j}}\right) + S_{n-1}(\mathbf{P} - \boldsymbol{p}; \boldsymbol{\mu})\right\}\right]$$
$$(11)$$

with $S_0(\mathbf{P}; \boldsymbol{\mu}) = 0$, where the expectation is with respect to $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_K)$. Let $(\widehat{p}_{1,n}(\boldsymbol{\alpha}, \boldsymbol{P}; \boldsymbol{\mu}), \cdots, \widehat{p}_{K,n}(\boldsymbol{\alpha}, \boldsymbol{P}; \boldsymbol{\mu}))$ be the vector $\boldsymbol{p}$ achieving the maximum in (11). Then, the optimal power policy is given by

$$\widehat{\beta}_{k,n}(\mathcal{F}_n; \boldsymbol{\mu}, \boldsymbol{\gamma}) = \widehat{p}_{k,N-n+1}\left(\boldsymbol{\alpha}_n, N\boldsymbol{\gamma} - \sum_{j=1}^{n-1}\widehat{\boldsymbol{\beta}}_j(\mathcal{F}_j; \boldsymbol{\mu}, \boldsymbol{\gamma}); \boldsymbol{\mu}\right) \quad (12)$$

where $\boldsymbol{\alpha}_n = (\alpha_{1,n}, \ldots, \alpha_{K,n})$ denotes the fading power vector in slot n.

Proof. The recursion in (11) and the optimal power policy in (12) follow easily from the general theory of dynamic programming [4] when the cost function to be maximized is given by (8) and the system *state*, in the presence of a command $\boldsymbol{p}_n$, evolves from time n to time $n+1$ according to $(\boldsymbol{\alpha}_n, \boldsymbol{P}) \to (\boldsymbol{\alpha}_{n+1}, \boldsymbol{P} - \boldsymbol{p}_n)$. $\qquad\square$

It follows that the maximum of the rate weighted sum (8) is given by

$$\sum_{k=1}^{K}\mu_k\,\widehat{R}_{k,N}(\boldsymbol{\mu}, \boldsymbol{\gamma}) = \frac{1}{N}S_N(N\boldsymbol{\gamma}; \boldsymbol{\mu}) \quad (13)$$

Numerical results for the recursion in (11) in the case of Rayleigh fading and $K = 1$ are provided in [15].

Interestingly, in contrast with [5], the convex hull operation in the boundary characterization of Theorem 2 is needed since the rates $\widehat{\boldsymbol{R}}_N(\boldsymbol{\mu}, \boldsymbol{\gamma})$ might not be continuous functions of $\boldsymbol{\mu}$. Consider, as an example, the case for $N = 1$. The region $C_{K,1}(\boldsymbol{\gamma})$ coincides with the ergodic capacity region of a fading channel without CSI at the transmitters, the dominant face of which is an hyperplane in K dimensions. Due to the polymatroid structure of $C_{K,1}(\boldsymbol{\gamma})$, the solution in (10) is one of the (at most) $K!$ vertices of the dominant

face. Hence, as $\boldsymbol{\mu}$ varies in $\mathbb{R}_+^K$, the set of $\widehat{\boldsymbol{R}}_N(\boldsymbol{\mu}, \boldsymbol{\gamma})$ contains at most $K!$ points. It is clear that the convex hull operation is needed here.

Although for finite N a closed form solution of (11) seems infeasible, for large N we can prove:

Theorem 3. In the limit for large N, the average capacity region $C_{K,N}(\boldsymbol{\gamma})$ tends to the ergodic capacity region [5]

$$
C_K^{(\mathrm{erg})}(\boldsymbol{\gamma}) \;=\; \bigcup_{\boldsymbol{\beta} \in \Gamma_K^{(\mathrm{erg})}(\boldsymbol{\gamma})} \Big\{ \boldsymbol{R} \in \mathbb{R}_+^K \;:\; \forall \mathcal{A} \subseteq \{1, \cdots, K\}
$$
$$
\sum_{k \in \mathcal{A}} R_k \le \mathrm{E}\left[\log\left(1 + \sum_{k \in \mathcal{A}} \alpha_k \beta_k(\boldsymbol{\alpha}) \right) \right] \Big\} \tag{14}
$$

where the expectation is taken with respect to the instantaneous channel state $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_K)$ and $\Gamma_K^{(\mathrm{erg})}(\boldsymbol{\gamma})$ is the set of feasible memoryless and stationary power allocation policies $\boldsymbol{\beta} = \{\beta_k \;:\; k = 1, \cdots, K\}$ defined by

$$
\Gamma_K^{(\mathrm{erg})}(\boldsymbol{\gamma}) \triangleq \big\{ \boldsymbol{\beta} \in \mathbb{R}_+^K \;:\; \mathrm{E}[\beta_k(\boldsymbol{\alpha})] \le \gamma_k \big\} \tag{15}
$$

Proof. See Appendix C. $\qquad\square$

Theorem 3 shows that for large N the penalty incurred by the use of a causal power allocation policy with respect to the ergodic power allocation policy vanishes. In other words, as N increases, the past information becomes irrelevant and the power policy becomes time-invariant and memoryless. An interesting open question is the characterization of the rate of convergence of the average capacity region $C_{K,N}(\boldsymbol{\gamma})$ to the ergodic capacity region $C_K^{(\mathrm{erg})}(\boldsymbol{\gamma})$.

4   The average capacity region per unit energy

For multiaccess channels the fundamental limit that determines the optimum use of the energy is the capacity region per unit energy [17]. In the variable-rate coding setting the average capacity region per unit energy is defined in Appendix A and admits the following characterization.

Theorem 4. The average capacity region per unit energy is

$$
U_{K,N} = \bigcup_{\boldsymbol{\gamma} \in \mathbb{R}_+^K} \big\{ \boldsymbol{r} \in \mathbb{R}_+^K : (\gamma_1 r_1, \cdots, \gamma_K r_K) \in C_{K,N}(\boldsymbol{\gamma}) \big\} \tag{16}
$$

Proof. The proof follows immediately from [17, Theorem 5]. $\qquad\square$

In analogy with [17], we also have:

Theorem 5. The average capacity region per unit energy is the hyper-rectangle

$$U_{K,N} = \left\{ \boldsymbol{r} \in \mathbb{R}_+^K : r_k \leq s_N^{(k)} \right\} \tag{17}$$

where $s_N^{(k)}$, given by

$$s_N^{(k)} = \lim_{\gamma_k \to 0} \frac{1}{\gamma_k} \sup_{\boldsymbol{\beta} \in \Gamma_{1,N}(\gamma_k)} \mathrm{E}\left[\frac{1}{N} \sum_{n=1}^{N} \alpha_{k,n} \beta_{k,n}(\mathcal{F}_n) \right], \tag{18}$$

is the k-th user single-user average capacity per unit energy.

Proof. See Appendix D. $\qquad\square$

The explicit solution of (18) was found originally in [15] for the single-user case. We report it here in our notation for later use:

Theorem 6. The k-th user single-user average capacity per unit energy $s_N^{(k)}$ is given by the dynamic programming recursion

$$s_n^{(k)} = \mathrm{E}[\max\{s_{n-1}^{(k)}, \alpha_k\}] \tag{19}$$

for $n = 1, \ldots, N$ with initial condition $s_0^{(k)} = 0$ and where the expectation is taken with respect to $\alpha_k \sim F_\alpha^{(k)}(x)$. Furthermore, $s_N^{(k)}$ is achieved by the "one-shot" power allocation policy defined by

$$\beta_{k,n}^\star = \begin{cases} N\gamma_k & \text{if } n = n_k^\star \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

where the random variable $n_k^\star$, function of $(\alpha_{k,1}, \cdots, \alpha_{k,N})$, is defined as

$$n_k^\star = \min\left\{ n \in \{1, \ldots, N\} : \alpha_{k,n} \geq s_{N-n}^{(k)} \right\} \tag{21}$$

Proof. See the proof given in [15]. $\qquad\square$

We refer to the optimal policy $\beta^\star$ as "one-shot" because the whole available energy $N\gamma_k$ is spent in a single slot. In fact, in each slot $n \in \{1, \cdots, N\}$, the transmitter compares the instantaneous fading gain $\alpha_{k,n}$ with the threshold $s_{N-n}^{(k)}$. If $\alpha_{k,n} \geq s_{N-n}^{(k)}$, then all the available energy is transmitted in slot n. Since the threshold for $n = N$ is zero ($s_0^{(k)} = 0$), the available energy is used with probability 1 within the codeword of N slots. The intuitive explanation of why the optimal *power* policy is decentralized in
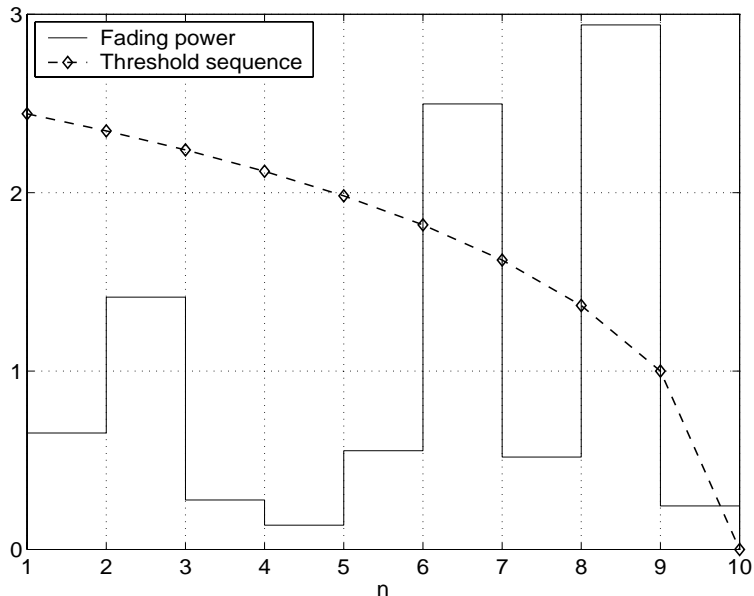
10

Figure 1: Rayleigh fading realization over a codeword of $N = 10$ slots and the corresponding thresholds for the "one-shot" policy.

the low power regime comes from the observation that, when the transmit powers are very small compared to the power of the additive noise, the presence of competing, and potentially interfering, users is not the primarily cause of performance degradation. In this case, the power allocation policy solely depends on the user fading process, however the *rate* allocation policy must be centralized. In fact, the users must coordinate their transmit rates so that reliably joint decoding at the central receiver is possible.

Fig. 1 shows a snapshot of a Rayleigh fading realization over a window of $N = 10$ slots and the corresponding thresholds for the "one-shot" policy. In this case transmission takes place in slot $n^\star = 6$. Notice that the optimal non-causal power policy, would have chosen for transmission the slot $(n = 8)$ with largest fading gain.

The threshold sequence $\{s_n^{(k)}\}_{n=0}^\infty$ is non-decreasing and depends only on the fading distribution $F_\alpha^{(k)}(\cdot)$ and not on the actual fading realization. Hence, it can be precomputed and stored in memory. When varying the delay requirements from N_1 to N_2 for the same fading statistics, the threshold sequence needs not be re-computed from scratch: only an extended segment $\{s_n^{(k)}\}_{n=0}^{N_2-1}$, instead of $\{s_n^{(k)}\}_{n=0}^{N_1-1}$, has to be used. Notice also that the number of active users K does not affect the value of the thresholds.

The behavior of $s_N^{(k)}$ when N grows to infinity is given by:

Theorem 7. For large N, the k-th user single-user average capacity per unit energy

11

$s_N^{(k)}$ tends to the k-th user single-user ergodic capacity per unit energy, given explicitly by

$$\lim_{N \to \infty} s_N^{(k)} = \sup\{\alpha_k\} \triangleq \inf\{x \geq 0 \,:\, F_\alpha^{(k)}(x) = 1\} \tag{22}$$

Proof. See Appendix E. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Notice that $\sup\{\alpha_k\} = \infty$ for fading distribution with infinite support.

5 Performance in the low power regime

In Section 3 we gave a characterization of the boundary surface of the average capacity region for arbitrary numbers of users K and slots N. In Section 4 we proved that the average capacity region per unit energy is achieved by letting all users transmit at vanishing SNR. In this section we characterize the average capacity region in the regime of small (but non-zero) SNR by comparing the average performance of the "one shot" policy, optimal for vanishing SNR, with the average performance of the optimal policy in (12).

5.1 The single-user case: background

The optimality of a coding scheme in the low power regime is defined and studied for several input-constrained additive noise channels in [18]. Let $C(\mathsf{SNR})$ be the capacity expressed in nat/s/Hz as a function of the (transmit) SNR, and let $\mathsf{C}(E_b/N_0)$ denote the corresponding spectral efficiency in bit/s/Hz as a function of the energy per bit vs. noise power spectral density, E_b/N_0, given implicitly by the parametric equation

$$\begin{cases} \frac{E_b}{N_0} = \frac{\mathsf{SNR}}{C(\mathsf{SNR})} \log 2 \\ \mathsf{C}\left(\frac{E_b}{N_0}\right) = C(\mathsf{SNR})\frac{1}{\log 2} \end{cases} \tag{23}$$

The value $(E_b/N_0)_{\min}$ for which $\mathsf{C}(E_b/N_0) > 0 \Leftrightarrow E_b/N_0 > (E_b/N_0)_{\min}$, is given by [18]

$$\left(\frac{E_b}{N_0}\right)_{\min} = \lim_{\mathsf{SNR} \downarrow 0} \frac{\mathsf{SNR}}{C(\mathsf{SNR})} \log 2 = \frac{\log 2}{\dot{C}(0)} \tag{24}$$

where $\dot{C}(0)$ is the derivative of the capacity function at $\mathsf{SNR} = 0$. From [17] and from the proof of Theorem 5, we see immediately that the reciprocal of $(E_b/N_0)_{\min}$ is the capacity per unit energy (expressed in bit/joule) of the channel.

In the low power regime, the behavior of spectral efficiency for energy per bit close to its minimum value is of great importance, as it is able to quantify, for example, the bandwidth requirement for a given desired data rate (see the detailed discussion in [18]). This

behavior is captured by the slope of spectral efficiency in bit/s/Hz/(3 dB), at $(E_b/N_0)_{\min}$, given by (see [18, Theorem 6])

$$\mathcal{S}_0 = \frac{2\left(\dot{C}(0)\right)^2}{-\ddot{C}(0)} \tag{25}$$

where $\ddot{C}(0)$ denotes the second derivative of the capacity function at $\textsf{SNR} = 0$. A signaling strategy is said to be *first-order optimal* if it achieves $(E_b/N_0)_{\min}$ and *second-order optimal* if it achieves both $(E_b/N_0)_{\min}$ and $\mathcal{S}_0$.

5.2 First and second-order optimality of $\beta^\star$ in the single user case

We deal first with the single user case, i.e., $K = 1$. For simplicity of the notation we drop the user index, we indicate the single-user average capacity given in Theorem 2 (with a slight abuse of notation) as

$$C_{1,N}(\gamma) = \frac{S_N(N\gamma)}{N} \tag{26}$$

and we re-write the recursion in (11) for $\mu = 1$ as

$$S_n(P) = \mathbb{E}\left[\max_{p \in [0,P]} \{\log(1 + \alpha\, p) + S_{n-1}(P - p)\}\right] \tag{27}$$

for $n = 1, \cdots, N$ with initial condition $S_0(P) = 0$. It is understood that, when considering user k, the mean value in (27) is computed with respect to $\alpha \sim F_\alpha^{(k)}(x)$ and the SNR in (26) is $\gamma = \gamma_k$.

Even if we cannot give a closed form expression for $S_N(P)$ and for $\widehat{\boldsymbol{\beta}}$, the low power characterization of the single-user average capacity and the second-order optimality of the "one-shot" policy $\beta^\star$ are given by:

Theorem 8. $(E_b/N_0)_{\min}$ and $\mathcal{S}_0$ for the single-user block-fading channel with causal transmitter CSI and frame length N are given by

$$\left(\frac{E_b}{N_0}\right)_{\min} = \frac{\log 2}{\dot{S}_N(0)} \tag{28}$$

$$\mathcal{S}_0 = \frac{2\left(\dot{S}_N(0)\right)^2}{-N\,\ddot{S}_N(0)} \tag{29}$$

where $\dot{S}_N(0)$ and $\ddot{S}_N(0)$ are, respectively, the first and the second derivative of $S_N(P)$ in (27) at $P = 0$. The first derivative is given by

$$\dot{S}_N(0) = s_N \tag{30}$$

where s_N is given by recursion (19), and the second derivative is given by the recursion

$$-\ddot{S}_n(0) = \mathrm{E}\left[\alpha^2 | \alpha \geq s_{n-1}\right] \Pr(\alpha \geq s_{n-1}) - \ddot{S}_{n-1}(0) \Pr(\alpha < s_{n-1}) \tag{31}$$

for $n = 1, \ldots, N$, with $\ddot{S}_n(0) = 0$.

Furthermore, the one-shot power allocation policy $\boldsymbol{\beta}^\star$ achieves $(E_b/N_0)_{\min}$ and slope $\mathcal{S}_0$, i.e., it is first and second-order optimal.

Proof. The expressions in (28) and in (29) follow by using (26) in (24) and (25). The statement in (30) follows immediately by noticing that $\dot{S}_N(0) = \dot{C}_{1,N}(0)$, from (26), and that $\dot{C}_{1,N}(0) = s_N$, from the proof of Theorem 5. The proof of (31) and of the second-order optimality of $\boldsymbol{\beta}^\star$ are given in Appendix F. $\qquad\square$

5.3   The multiuser case: background

In a multiaccess channel, the individual user energy per bit over N_0 are defined by $E_k/N_0 \triangleq \log 2\, \gamma_k/R_k$, where γ_k is the transmit SNR (energy/symbol) and R_k is the rate (in nat/s/Hz) of user k. We indicate by $\mathcal{S}_0^{(k)}$ the k-th user single-user slope and by $\mathcal{S}_k$ the slope of user k in the multiuser case. Note that $\mathcal{S}_0^{(k)}$ is given by (29), where the superscript "(k)" stresses the fact that the mean values are computed using $F_\alpha^{(k)}(x)$. In general, R_k is the k-th component of an achievable rate K-tuple $\mathbf{R}$. Without loss of generality we can consider only points on the boundary surface of the capacity region defined by the input constraints $\gamma_1, \ldots, \gamma_K$. To stress the fact that these points are functions of $\gamma_1, \ldots, \gamma_K$, we shall write $R_k = R_k(\gamma_1, \ldots, \gamma_K)$.

In order to make use of the theory developed for the single user case, we fix a vector $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_K) \in \mathbb{R}_+^K$ and we let the user SNRs vanish with fixed ratio $\gamma_k/\gamma_j = \theta_k/\theta_j$, for all $i, j \in \{1, \cdots, K\}$. The fact that, from Theorem 5, the average capacity region per unit energy is an hyper-rectangle implies that for vanishing rates $R_k \approx s_N^{(k)} \gamma_k$. Hence, in the low power regime, imposing SNR ratios is equivalent to fix rate ratios

$$\frac{\gamma_k}{\gamma_j} = \frac{\theta_k}{\theta_j} \quad \Rightarrow \quad \frac{R_k}{R_j} = \frac{s_N^{(k)}\, \theta_k}{s_N^{(j)}\, \theta_j} \tag{32}$$

The user k rate can be expressed solely as function of γ_k as

$$R_k = R_k\left(\frac{\theta_1}{\theta_k}\gamma_k, \cdots, \frac{\theta_K}{\theta_k}\gamma_k\right) \tag{33}$$

and, by applying (29), we obtain

$$\mathcal{S}_k = \frac{2\left(\sum_{j=1}^{K} \theta_j \cdot \partial_j R_k(0, \cdots, 0)\right)^2}{-\sum_{j=1}^{K}\sum_{m=1}^{K} \theta_j \theta_m \cdot \partial_{j,m} R_k(0, \cdots, 0)} \tag{34}$$

14

where we define the shorthand notations

$$\partial_j R_k(0, \cdots, 0) \triangleq \lim_{\boldsymbol{\gamma} \to \mathbf{0}} \frac{\partial R_k(\gamma_1, \cdots, \gamma_K)}{\partial \gamma_j} \tag{35}$$

and

$$\partial_{j,m} R_k(0, \cdots, 0) \triangleq \lim_{\boldsymbol{\gamma} \to \mathbf{0}} \frac{\partial^2 R_k(\gamma_1, \cdots, \gamma_K)}{\partial \gamma_j \, \partial \gamma_m} \tag{36}$$

Notice that the user k slope is completely characterized by the gradient and the Hessian matrix of the rate function R_k computed for $\boldsymbol{\gamma} = \mathbf{0}$.

In [20], the slope region for the standard two-user Gaussian MAC is studied and its boundary is explicitly parameterized with respect to the ratio $\theta = \gamma_1/\gamma_2$.

5.4   Slope region achieved by TDMA

Before carrying on the characterization of the slope region for the general multiuser case, we investigate the slope region achievable by TDMA. In this case, every slot is divided into K sub-slots each of which is assigned to a different user. Each user sees a single-user channel on its sub-slot, and applies a suitable (single-user) causal power policy satisfying its individual power constraint.

In Section 4 we have shown that the one-shot power allocation $\boldsymbol{\beta}^\star$ (in conjunction with Gaussian variable-rate coding) is optimal in the sense of achieving the average capacity region per unit energy, i.e., achieves $(E_k/N_0)_{\min}$ for all users. Then, we conclude that the one-shot policy is first-order optimal for any number of users K. From the proof of Theorem 5 it follows that first-order optimality can be obtained either by using superposition coding or by using TDMA inside each slot. As an immediate consequence of the second-order optimality of $\boldsymbol{\beta}^\star$ in the single-user case, stated by Theorem 8, we have:

Theorem 9.   For any arbitrary SNR ratios γ_k/γ_j, the slope region achievable by TDMA is given by

$$\left\{ \mathcal{S}_{k,\mathrm{tdma}} \geq 0 \quad \forall k = 1, \cdots, K \; : \; \sum_{k=1}^{K} \frac{\mathcal{S}_{k,\mathrm{tdma}}}{\mathcal{S}_0^{(k)}} \leq 1 \right\} \tag{37}$$

Furthermore, this is achieved by applying the one-shot power policy $\boldsymbol{\beta}^\star$.

Proof. For $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_K) \in \mathbb{R}_+^K$ such that $\sum_{k=1}^{K} \tau_k = 1$ the maximum achievable rates under TDMA are $R_k = (\tau_k/N) \cdot S_N(\gamma_k N/\tau_k)$. By straightforward application of (34), we have $\mathcal{S}_{k,\mathrm{tdma}} = \tau_k \, \mathcal{S}_0^{(k)}$ hence, by considering the union over all possible choice of $\boldsymbol{\tau}$, we get (37). $\qquad\square$

5.5 Second-order optimality of $\beta^\star$ in the multiuser case

The optimal slope region under the causal power constraint is given by:

Theorem 10. For any arbitrary SNR ratios $\gamma_k/\gamma_j = \theta_k/\theta_j$ (with $\theta_k > 0$ for all $k = 1, \ldots, K$) the optimal slope region is given by

$$\bigcup_{\boldsymbol{\lambda}} \left\{ \mathcal{S}_k \geq 0 \quad \forall k = 1, \cdots, K \ : \ \mathcal{S}_k \leq \frac{\mathcal{S}_0^{(k)}}{1 + \sum_{\boldsymbol{\pi}} \lambda_{\boldsymbol{\pi}} \sum_{j < \pi^{-1}(k)} \frac{\theta_{\pi_j}}{\theta_k} \mathcal{K}_{k,\pi_j}} \quad k = 1, \cdots, K \right\} (38)$$

where

$$\mathcal{K}_{k,j} = \frac{2 \sum_{n=1}^{N} \mathrm{E}[\alpha_{k,n} \mathbb{1}\{n_k^\star = n\}] \, \mathrm{E}[\alpha_{j,n} \mathbb{1}\{n_j^\star = n\}]}{\sum_{n=1}^{N} \mathrm{E}\left[\alpha_{k,n}^2 \mathbb{1}\{n_k^\star = n\}\right]} \tag{39}$$

where $\sum_{\boldsymbol{\pi}}$ denotes the sum over all permutations of $\{1, \ldots, K\}$ and where $\boldsymbol{\lambda} = \{\lambda_{\boldsymbol{\pi}}\}$ are $K!$ non-negative "time-sharing" coefficients (indexed by the permutations $\boldsymbol{\pi}$) such that $\sum_{\boldsymbol{\pi}} \lambda_{\boldsymbol{\pi}} = 1$. Furthermore, the one-shot policy $\boldsymbol{\beta}^\star$ achieves the optimal slope region, i.e., it is second-order optimal in the multiuser case.

Proof. See Appendix G. □

6 The optimal non-causal policy achieving the average capacity per unit energy

Before proceeding with numerical examples in which we compare the performance of the optimal power policy with the one-shot power policy in the low SNR regime and the performance of the (second order optimal) one-shot power policy with the (first order optimal) TDMA strategy, we briefly report the power policy that maximizes the average capacity region per unit energy with *non-causal* feedback, i.e., where the whole fading realization $\mathcal{F}_N$ is revealed to the transmitters at the beginning of each codeword. We limit ourselves to the single user case, since we saw that in the multiuser case the average capacity region per unit energy is the Cartesian product of the single-user average capacities per unit energy. If we allow the input to depend on the whole CSI $\mathcal{F}_N$ in a non-causal way, it is immediate to show that the optimal policy maximizing the average capacity per unit energy is "uniform maximum selection"

$$\beta_{k,n}^{\star \,(\mathrm{nc})}(\mathcal{F}_N) = \begin{cases} \frac{N\gamma_k}{|M_k|} & \text{if } n \in M_k \\ 0 & \text{otherwise} \end{cases} \tag{40}$$

where

$$M_k = \{n \; : \; \alpha_{k,n} = \max\{\alpha_{k,1}, \cdots, \alpha_{k,N}\}\} \tag{41}$$

The power policy in (40) allocates uniformly the available energy to the slots whose fading is equal to the maximum. Notice that with a continuous fading distribution $\Pr[|M_k| > 1] = 0$, therefore the whole available energy is concentrated in one slot almost surely. However, the selected slot might be different from the slot selected by the causal one-shot policy $\beta^\star$ in (20). For example, in the snapshot realization of Fig. 1 $\beta^{\star\,(\mathrm{nc})}$ would select slot 8 instead of slot 6 selected by $\beta^\star$.

The following results are straightforward extensions of the theory developed for the case of causal CSI.

Theorem 11. $(E_b/N_0)_{\min}^{(\mathrm{nc})}$ and $\mathcal{S}_0^{(\mathrm{nc})}$ for the single-user block fading channel with non-causal transmitter CSI and frame length of N slots are given by

$$\left(\frac{E_b}{N_0}\right)_{\min}^{(\mathrm{nc})} = \frac{\log 2}{\mathrm{E}[\max\{\alpha_1, \cdots, \alpha_N\}]} \tag{42}$$

$$\mathcal{S}_0^{(\mathrm{nc})} = \frac{2\,(\mathrm{E}[\max\{\alpha_1, \cdots, \alpha_N\}])^2}{N\;\mathrm{E}[(\max\{\alpha_1, \cdots, \alpha_N\})^2]} \tag{43}$$

Furthermore, the uniform maximum-selection power policy $\beta^{\star\,(\mathrm{nc})}$ achieves both $(E_b/N_0)_{\min}^{(\mathrm{nc})}$ and $\mathcal{S}_0^{(\mathrm{nc})}$. $\qquad\square$

With TDMA, because of the second-order optimality of $\beta^{\star\,(\mathrm{nc})}$ in the single-user case, we have:

Theorem 12. For any arbitrary SNR ratios γ_k/γ_j, the slope region achievable by TDMA is given by

$$\left\{ \mathcal{S}_{k,\mathrm{tdma}}^{(\mathrm{nc})} \geq 0 \quad \forall k = 1, \cdots, K \; : \; \sum_{k=1}^{K} \frac{\mathcal{S}_{k,\mathrm{tdma}}^{(\mathrm{nc})}}{\mathcal{S}_0^{(k)\,(\mathrm{nc})}} \leq 1 \right\} \tag{44}$$

$$\square$$

Finally, the optimal slope region is given by:

Theorem 13. The optimal slope region with non-causal CSI is given by (38) with the coefficients $\mathcal{K}_{k,j}$ given by

$$\mathcal{K}_{k,j} = \frac{2}{N} \frac{\mathrm{E}[\max\{\alpha_{k,1}, \cdots, \alpha_{k,N}\}]\,\mathrm{E}[\max\{\alpha_{j,1}, \cdots, \alpha_{j,N}\}]}{\mathrm{E}[(\max\{\alpha_{k,1}, \cdots, \alpha_{k,N}\})^2]} \tag{45}$$

Furthermore, $\boldsymbol{\beta}^{\star\,(\mathrm{nc})}$ is first and second-order optimal for any number of users K and any delay N. $\qquad\square$

Proof. See Appendix H.

7 Example: the Rayleigh fading case

In order to illustrate the results of previous sections we consider the case of i.i.d. Rayleigh fading, where the channel gain law is $F_\alpha(x) = 1 - \mathrm{e}^{-x}$ for $x \geq 0$ for all users.

Comparison between causal and non-causal power policy. The one-shot policy $\boldsymbol{\beta}^\star$ is completely determined by the thresholds given by the recursion in (19) and explicitly computable as

$$s_n = s_{n-1} + \mathrm{e}^{-s_{n-1}} \tag{46}$$

for $n = 1, 2, \cdots$ with $s_0 = 0$. The first and second-order derivatives of the average capacity region $C_{1,N}(\gamma)$ are given by $\dot{C}_{1,N}(0) = s_N$ and by $\ddot{C}_{1,N}(0) = N\ddot{S}_N(0)$ where $\ddot{S}_n(0)$ is given by the recursion in (31), that can be written explicitly as

$$-\ddot{S}_n(0) = \mathrm{e}^{-s_{n-1}}(2 + 2s_{n-1} + s_{n-1}^2) - \ddot{S}_{n-1}(0)(1 - \mathrm{e}^{-s_{n-1}}); \tag{47}$$

for $n = 1, 2, \cdots$ with $\ddot{S}_0(0) = 0$.

For the case of non-causal CSI, the minimum energy per bit and the slope are given by (42) and (43) respectively, with

$$\mathrm{E}[\max\{\alpha_1, \cdots, \alpha_N\}] \;\; = \;\; \sum_{n=1}^{N} \binom{N}{n} (-1)^{n+1} \frac{1}{n} \tag{48}$$

$$\mathrm{E}[(\max\{\alpha_1, \cdots, \alpha_N\})^2] \;\; = \;\; \sum_{n=1}^{N} \binom{N}{n} (-1)^{n+1} \frac{2!}{n^2} \tag{49}$$

Figs. 2 and 3 show $(E_b/N_0)_{\min}$ and $\mathcal{S}_0$ versus the frame length N and for both the causal and the non-causal knowledge of the channel state.

For a given delay N, the curves of spectral efficiency vs. E_b/N_0 for the causal system and for the non-causal system start at different $(E_b/N_0)_{\min}$, smaller for the non-causal system, with almost equal slope. The gain due to causal vs. non-causal transmit CSI is large, and increasing with N, as far as $(E_b/N_0)_{\min}$ is concerned. On the contrary, the slopes in the two cases are very similar. Notice that, in general, the slope with causal CSI need not be smaller than the slope with non-causal CSI since the corresponding values of $(E_b/N_0)_{\min}$ are different.

Comparison between TDMA and superposition coding. For a desired user rate R_b (in bit/s) common to all users, and assuming that all users transmit with equal power,

i.e., they have the same E_b/N_0 such that $(E_b/N_0)_{\text{dB}} - ((E_b/N_0)_{\text{min}})_{\text{dB}} = \epsilon$, the system bandwidth is given approximately by [18]

$$W \approx \frac{R_b}{\min_k S_0^{(k)} \epsilon} \tag{50}$$

We quantify the bandwidth expansion incurred by TDMA with respect to superposition coding for a given delay N.

Since (50) is determined by the minimum slope, in order to minimize the system bandwidth we have to maximize the minimum slope. From Theorems 9 and 10 we can find the max-min slope of an equal-rate system. For equal rates, $\theta_j/\theta_k = 1$ for all k, j, and the denominator of (38) becomes

$$
\begin{aligned}
1 + \mathcal{K}_0 \sum_{\boldsymbol{\pi}} \lambda_{\boldsymbol{\pi}} \sum_{j < \pi^{-1}(k)} 1 &= 1 + \mathcal{K}_0 \sum_{\boldsymbol{\pi}} \left(\lambda_{\boldsymbol{\pi}} \pi^{-1}(k) - 1 \right) \\
&= 1 - \mathcal{K}_0 + \mathcal{K}_0 \sum_{\boldsymbol{\pi}} \lambda_{\boldsymbol{\pi}} \pi^{-1}(k) \tag{51}
\end{aligned}
$$

where, for i.i.d. fading, $\mathcal{K}_{k,j}$ in (39) are all equal to $\mathcal{K}_0$ given by

$$\mathcal{K}_0 = 2 \frac{\sum_{n=1}^{N} \left(\mathrm{E}\left[\alpha_n 1\{n^\star = n\} \right] \right)^2}{\sum_{n=1}^{N} \mathrm{E}\left[\alpha_n^2 1\{n^\star = n\} \right]} \tag{52}$$

As $\boldsymbol{\pi}$ varies over all $K!$ permutations, $\pi^{-1}(k)$ takes on each value $1, \ldots, K$ exactly $(K-1)!$ times. Because of symmetry, the max-min slope is achieved by letting $S_0^{(k)} = \text{const.}$, i.e., $\lambda_{\boldsymbol{\pi}} = 1/K!$ for all $\boldsymbol{\pi}$. This yields

$$\max_{\boldsymbol{\lambda}} \min_k S_k = \frac{S_0}{1 + \mathcal{K}_0(K-1)/2} \tag{53}$$

For TDMA, the max-min slope is obtained by letting $\tau_k = 1/K$, which yields

$$\max_{\boldsymbol{\tau}} \min_k S_{k,\text{tdma}} = S_0/K \tag{54}$$

Therefore, the bandwidth expansion factor of TDMA with respect to superposition coding is given by

$$\eta = \frac{K}{1 + \mathcal{K}_0(K-1)/2} < \frac{2}{\mathcal{K}_0} \tag{55}$$

From (52) we have immediately that $\mathcal{K}_0 < 2$, which means that TDMA is strictly sub-optimal for any non-degenerate fading distribution. Notice also that the case of equal

E_b/N_0 for all users is the most favorable for TDMA [20]. For a very imbalanced system the bandwidth expansion factor can be much larger than (55).

Fig. 4 shows the asymptotic expansion factor $2/\mathcal{K}_0$ for large number of users ($K \to \infty$) versus the delay N for different fading statistics. Fig. 5 shows the bandwidth expansion factor η versus the number of users K and different values of N for the Rayleigh fading case. For example, at $N = 2$ and $K = 4$, the TDMA requires more than twice the bandwidth required by a system with superposition coding (Fig. 5) and, asymptotically for a large K, the TDMA requires more than three times the bandwidth required by a system with superposition coding (Fig. 4).

By increasing either the frame length N and/or the number of users K, TDMA becomes increasingly suboptimal.

Slope region for the two user case. We study in more detail the case $K = 2$. For superposition coding, by letting $\theta = \theta_1/\theta_2$, we have

$$
\begin{cases}
\mathcal{S}_1 & \leq \quad \frac{\mathcal{S}_0}{1+\mathcal{K}_0(1-\lambda)\frac{1}{\theta}} \\
\mathcal{S}_2 & \leq \quad \frac{\mathcal{S}_0}{1+\mathcal{K}_0\lambda\theta}
\end{cases}
\tag{56}
$$

By eliminating the time-sharing parameter λ we obtain explicitly the slope region boundary as

$$
\left(\frac{\mathcal{S}_0}{\mathcal{S}_1} - 1\right)\theta + \left(\frac{\mathcal{S}_0}{\mathcal{S}_2} - 1\right)\frac{1}{\theta} = \mathcal{K}_0, \quad 0 \leq \mathcal{S}_k \leq \mathcal{S}_0
\tag{57}
$$

With TDMA we obtain the boundary $\mathcal{S}_{1,\text{tdma}} + \mathcal{S}_{2,\text{tdma}} = \mathcal{S}_0$.

Fig. 6 shows the two-user slope region for different rate ratios. The slope region achievable by TDMA is shown for comparison. This figure clearly illustrates that even though TDMA achieves the capacity per unit energy, it is actually suboptimal in the low power regime, especially in a fading scenario.

8   Conclusions

In this paper we have analyzed an idealized slotted multiaccess Gaussian channel characterized by block-fading, where each codeword must be transmitted and decoded within N slots and undergoes N independently drawn fading states. At each slot, the rate and power allocated to each user is computed on the basis of the history of all the fading coefficients encountered up to and including that slot.

Much of our analysis has focused in the low spectral efficiency regime, which is where the major benefits of transmitter feedback occurs. We have analyzed not only the rates
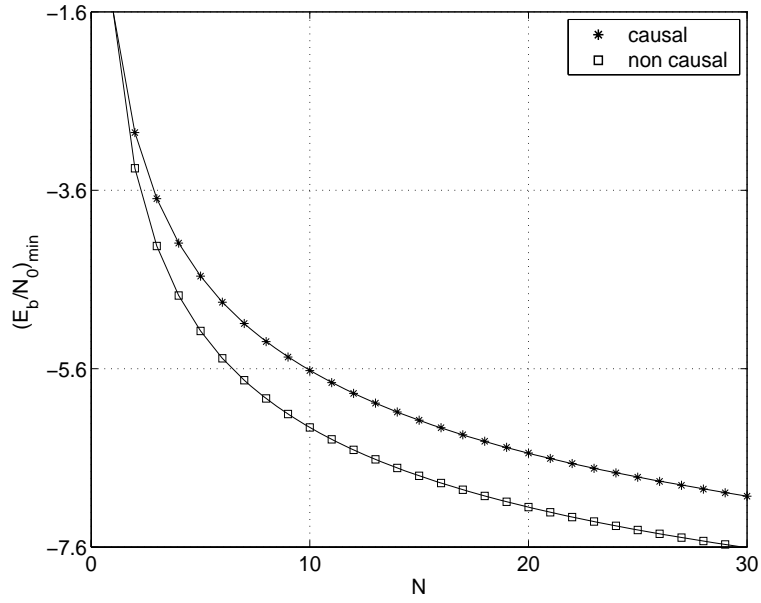
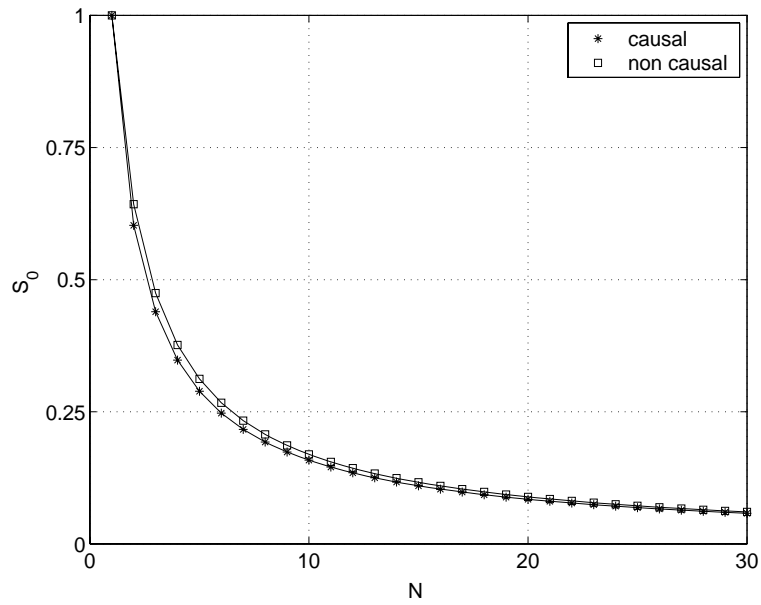Figure 2: $(E_b/N_0)_{\min}$ in dB vs N for the Rayleigh fading case.

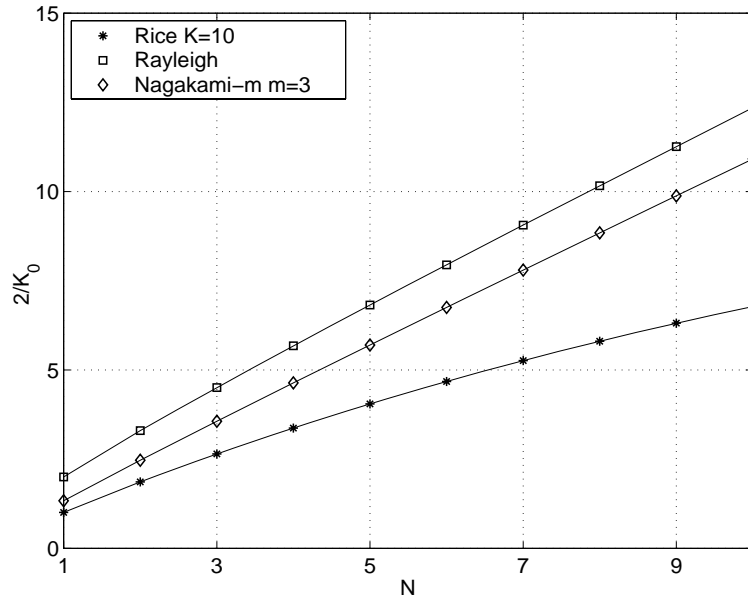Figure 3: S_0 vs N for the Rayleigh fading case.

21

Figure 4: Limiting expansion factor of TDMA over superposition coding vs N for different fading distributions.
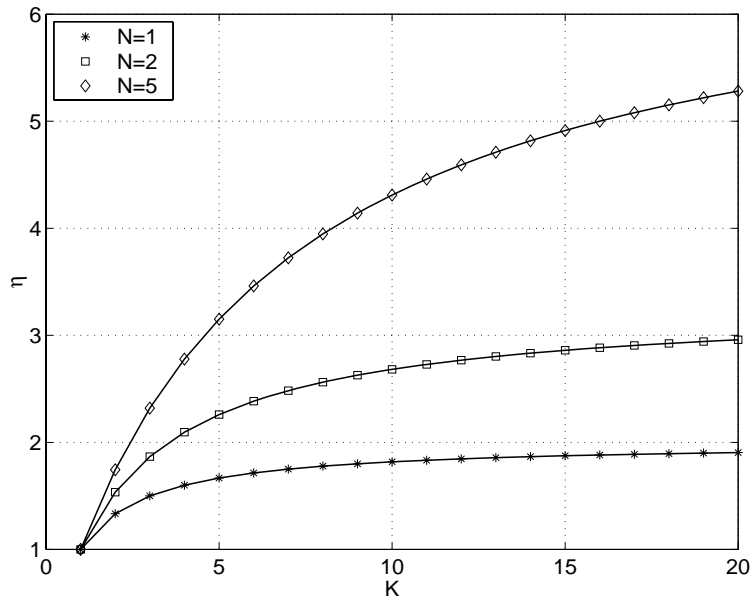

Figure 5: Bandwidth expansion factor of TDMA over superposition coding vs. the number of users K for the Rayleigh fading case.
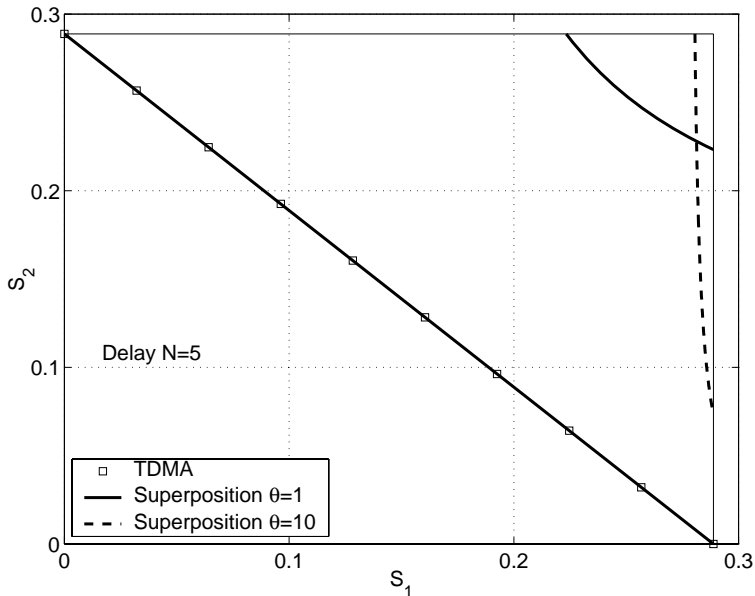
Figure 6: Slope region for $K = 2$ for the Rayleigh fading case.

achievable in the vanishing SNR regime (capacity region per unit energy, or equivalently the minimum value of E_b/N_0), but also the slopes of the users individual spectral efficiencies at the point $(E_b/N_0)_{\min}$.

In particular, we have shown that the optimal transmission scheme in the low power regime is based on Gaussian variable-rate coding whose power (and rate) is allocated according to a one-shot policy, that concentrates all transmitter available energy in the first slot whose fading power is above a time-varying threshold function. The threshold function can be explicitly computed by a simple recursive formula and depends only on the fading statistics. Interestingly, the power allocation policy of user k depends only on the k-th fading state sequence. However, even for the one-shot power allocation policy, the rate allocation is, in general, centralized. A notable exception is when the one-shot power policy is used in conjunction with TDMA inside each slot. This is a simple and decentralized scheme where each user allocates its power and rate based on the (causal) observation of its own fading only. This scheme is first-order optimal, in the sense that it achieves the capacity region per unit energy (equivalently, it achieves $(E_b/N_0)_{\min}$ for all users). However, this scheme is not second-order optimal, i.e., its slope region is strictly inside the optimal slope region, for any non-degenerate fading distribution. The penalty incurred by TDMA is rather substantial and depends on the fading statistics and grows with both the number of fading states N and the number of users K.

We have shown that the optimal slope region is achieved by the same one-shot policy in conjunction with superposition coding (and successive interference cancellation decoding at the receiver). Fully decentralized schemes (with uncoordinated rates) cannot achieve the optimal slope region, since superposition coding requires the users to coordinate their transmission rates. The investigation of the achievable performance in the low power regime under fully decentralized schemes is left as an interesting problem for future research.

Appendix

A   Definitions

We model the variable rate coding scenario by letting the message set size depend on the fading state. For user k, let $\mathcal{W}_{k,n} = \{W_{k,n}(\mathcal{F}_n) : \mathcal{F}_n \in \mathbb{C}^{nK}\}$ be a collection of message sets indexed by the channel state $\mathcal{F}_n$, each with cardinality $|W_{k,n}(\mathcal{F}_n)| = M_{k,n}(\mathcal{F}_n)$.

Definition 1.   A variable-rate coding system is defined by:
a) An assignment $\mathcal{W}_{k,n}$ of message sets to the fading states;
b) KN encoding functions $\phi_{k,n} : W_{k,n}(\mathcal{F}_n) \times \mathbb{C}^{nK} \to \mathbb{C}^L$ for $n = 1, \ldots, N$ such that $\phi_{k,n}(w, \mathcal{F}_n) = \boldsymbol{x}_{k,n}$, where $w \in W_{k,n}(\mathcal{F}_n)$, and such that the resulting codewords satisfy (2);
c) For each channel state sequence $\mathcal{F}_N$, a decoding function

$$\psi_{\mathcal{F}_N} : \mathbb{C}^{NL} \to W_{1,1}(\mathcal{F}_1) \times \cdots \times W_{K,N}(\mathcal{F}_N)$$

such that $\psi_{\mathcal{F}_N}(\{\boldsymbol{y}_n : n = 1, \cdots, N\}) = (w_{1,1}, \cdots, w_{K,N})$ where $w_{k,n} \in W_{k,n}(\mathcal{F}_n)$. $\Diamond$

For given $\mathcal{F}_N$, the coding rate of user k is given by

$$R_k(\mathcal{F}_N) = \frac{1}{NL} \sum_{n=1}^{N} \log\left(M_{k,n}(\mathcal{F}_n)\right) \tag{58}$$

and the error probability is given by

$$
\begin{aligned}
\mathrm{P_e}(\mathcal{F}_N) \;\; &= \;\; \frac{1}{\prod_{k=1}^{K} \prod_{n=1}^{N} M_{k,n}(\mathcal{F}_n)} \cdot \\
&\sum_{w_{1,1}, \cdots, w_{K,N}} \mathrm{Pr}\left(\psi_{\mathcal{F}_N}(\{\boldsymbol{y}_n\}) \neq (w_{1,1}, \ldots, w_{K,N}) | (w_{1,1}, \ldots, w_{K,N})\right)
\end{aligned}
\tag{59}
$$

Definition 2.   A variable-rate coding scheme for frame length N, slot length L, with average rate K-tuple

$$\mathbf{R} = (R_1, \ldots, R_K),$$

where

$$R_k = E\left[R_k(\mathcal{F}_N)\right],$$

with power constraint defined by the K-tuple $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)$, and attaining error probability

$$\mathrm{P_e}(\mathcal{F}_N) \leq \epsilon \;\; \forall \mathcal{F}_N$$

is said to be an $(N, L, \mathbf{R}, \boldsymbol{\gamma}, \epsilon)$-code.

The operative definitions of average capacity region and of average capacity region per unit-energy mimic, the standard definitions for input constrained channels in [19] and [17], respectively.

Definition 3. A rate K-tuple $\boldsymbol{R}^\star \in \mathbb{R}_+^K$ is average ϵ-achievable if for all $\lambda > 0$ there exist $\overline{L}$ such that for $L \geq \overline{L}$ variable-rate $(N, L, \boldsymbol{R}, \boldsymbol{\gamma}, \epsilon)$-codes can be found with $R_k > R_k^\star - \lambda$ for $k = 1, \cdots, K$. A rate K-tuple is achievable if it is ϵ-achievable for all $0 < \epsilon < 1$. The average capacity region $C_{K,N}(\boldsymbol{\gamma})$ is the closure of the convex hull of all achievable rate K-tuples. $\diamondsuit$

Definition 4. A K-tuple $\boldsymbol{r}^\star \in \mathbb{R}_+^K$ is a average ϵ-achievable rate per unit energy if for all $\lambda > 0$ there exist an energy vector $\bar{\boldsymbol{\nu}} = (\bar{\nu}_1, \cdots, \bar{\nu}_K)$ such that for $\boldsymbol{\nu} \geq \bar{\boldsymbol{\nu}}$ [3] variable-rate $(N, L, \boldsymbol{R}, \boldsymbol{\nu}/(NL), \epsilon)$-codes can be found with $(LNR_k)/\nu_k > r_k^\star - \lambda$ for $k = 1, \cdots, K$. A rate K-tuple is achievable if it is ϵ-achievable for all $0 < \epsilon < 1$. The average capacity region per unit-energy $U_{K,N}$ is the set of all achievable rate K-tuples per unit-energy. $\diamondsuit$

B Proof of Theorem 1

Achievability is easily obtained by considering a particular variable-rate coding system that encodes and decodes independently over the N slots. For each channel state [4] $\mathcal{F}_N$, the users construct a sequence of Gaussian codebooks of length L with i.i.d. entries of zero mean and unit variance and sizes $M_{k,n}(\mathcal{F}_n)$, satisfying the set of inequalities

$$\frac{1}{L} \sum_{k \in \mathcal{A}} \log\left(M_{k,n}(\mathcal{F}_n)\right) < \log\left(1 + \sum_{k \in \mathcal{A}} \alpha_{k,n} \beta_{k,n}(\mathcal{F}_n)\right) \tag{60}$$

for all $\mathcal{A} \subseteq \{1, \cdots, K\}$, where $\boldsymbol{\beta} \in \Gamma_{K,N}(\boldsymbol{\gamma})$. Each transmitter k, during slot n, after observing $\mathcal{F}_n$, selects a message uniformly on $W_{k,n}(\mathcal{F}_n) = \{1, \ldots, M_{k,n}(\mathcal{F}_n)\}$ and independently of the other transmitters, and sends the corresponding codeword amplified by the transmit power level $\beta_{k,n}(\mathcal{F}_n)$. The receiver perform decoding on a slot-by-slot basis (even though it is allowed to wait until the end of the frame of N slots). From the standard Gaussian MAC [19], any rate K-tuple satisfying the set of inequalities (60) is achievable,

---

[3]For two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, the notation $\boldsymbol{a} \geq \boldsymbol{b}$ means that the difference $\boldsymbol{a} - \boldsymbol{b}$ has non-negative components.

[4]For a rigorous treatment in the case where the fading is a continuous random vector we should use the argument of [5] based on the discretization of the fading state. For the sake of brevity, we cut short and we assume that we can define a codebook *for each* channel state.

i.e., the conditional decoding error probability given the channel state $\mathcal{F}_n$ vanishes as $L \to \infty$. By summing over N slots we get

$$\frac{1}{NL} \sum_{k \in \mathcal{A}} \log \left(\prod_{n=1}^N M_{k,n}(\mathcal{F}_n) \right) < \frac{1}{N} \sum_{n=1}^N \log \left(1 + \sum_{k \in \mathcal{A}} \alpha_{k,n} \beta_{k,n}(\mathcal{F}_n) \right) \tag{61}$$

with conditional (w.r.t. $\mathcal{F}_N$) error probability not larger than N times the maximum of the conditional error probabilities over the N slots. Finally, by taking expectation with respect to the channel state of both sides in (61) we get that the set of rates defined in (6) is achievable.

For the converse part, we consider the N-slot extension of our channel, with input $\mathbf{X}_k = \{ \boldsymbol{x}_{k,n} : n = 1, \cdots, N \}$ and output $\mathbf{Y} = \{ \boldsymbol{y}_n : n = 1, \cdots, N \}$, where the input constraint is given "frame-wise" by (2). [5] One frame of the original channel corresponds to a channel use of the new channel. Moreover, we relax the definition of achievable rates by constraining the average error probability.

The new channel is block-wise memoryless and its input constraint is imposed on a per-symbol basis (averaged over the codebook). We consider a sequence of such channels indexed by increasing L, and define the capacity region of the N-slot extension channel as the closure of the union of all regions for $L = 1, 2, \cdots$. The ergodic capacity region of the N-slot extension channel provides an outer bound to the average capacity region of the original channel.

Let $\mathbf{X} = \{ \mathbf{X}_k : k = 1, \ldots, K \}$ and, for any $\mathcal{A} \subseteq \{ 1, \ldots, K \}$, let $\mathbf{X}(\mathcal{A}) \triangleq \{ \mathbf{X}_k : k \in \mathcal{A} \}$ and $R(\mathcal{A}) \triangleq \sum_{k \in \mathcal{A}} R_k$. From standard results on memoryless MAC [19], the capacity region of the N-slot extension channel is given by

$$\bigcup_{\Pr(\mathbf{X}, V, \mathcal{F}_N)} \left\{ \boldsymbol{R} \in \mathbb{R}_+^K \ : \ R(\mathcal{A}) \leq \frac{1}{LN} I \left(\mathbf{X}(\mathcal{A}); \mathbf{Y} | \mathbf{X}(\overline{\mathcal{A}}), \mathcal{F}_N, V \right) \ \forall \mathcal{A} \subseteq \{ 1, \cdots, K \} \right\} \tag{62}$$

where the joint probability of $(\mathbf{X}, V, \mathcal{F}_N)$ satisfies

$$\Pr(\mathbf{X}, V, \mathcal{F}_N) = \left(\prod_{k=1}^K \prod_{n=1}^N \Pr(\boldsymbol{x}_{k,n} | \mathcal{F}_n, V, \boldsymbol{x}_{k,1}, \cdots, \boldsymbol{x}_{k,n-1}) \right) \Pr(V | \mathcal{F}_N) \Pr(\mathcal{F}_N) \tag{63}$$

and each factor $\prod_{n=1}^N \Pr(\boldsymbol{x}_{k,n} | \mathcal{F}_n, V, \boldsymbol{x}_{k,1}, \cdots, \boldsymbol{x}_{k,n-1})$ puts zero probability outside the sphere $\frac{1}{NL} \sum_{n=1}^N |\boldsymbol{x}_{k,n}|^2 \leq \gamma_k$. The input probability in the form (63) expresses the fact that encoding is independent for all transmitters, when conditioned with respect to the common CSI $\mathcal{F}_N$ and the time-sharing variable V, and that the common CSI is causal,

---

[5]Similar "blocking" techniques have been used to prove coding theorems for channels with ISI [2, 9].

i.e., that $\boldsymbol{x}_{k,n}$ depends only on $\mathcal{F}_n$ and not on the whole $\mathcal{F}_N$. Notice that we allow the time-sharing variable V to depend on the whole CSI $\mathcal{F}_N$, even if the CSI is only revealed causally to the transmitters (again, this can only increase the capacity region).

Fix an input probability distribution $P(\mathbf{X}, V, \mathcal{F}_N)$ in the form (63) with conditional componentwise second-order moments

$$\beta_{k,n}^{(\ell)}(\mathcal{F}_n, V) = \mathrm{E}[|x_{k,n}^{(\ell)}|^2 | \mathcal{F}_n, V] \tag{64}$$

where $x_{k,n}^{(\ell)}$ denotes the ℓ-th component of $\boldsymbol{x}_{k,n}$. Since the channel is additive and the input second-order moment is constrained, the boundary of the region (62) is clearly achieved only if $P(\mathbf{X}, V, \mathcal{F}_N)$ satisfies $\mathrm{E}[\mathbf{X}|\mathcal{F}_N, V] = \mathbf{0}$. Then, we shall restrict to this case. Let $P(\mathbf{Y}, \mathbf{X}, \mathcal{F}_N, V)$ be the joint input-output probability corresponding to $P(\mathbf{X}, V, \mathcal{F}_N)$ and to the transition probability of the channel. Let $\Phi(\mathbf{Y}, \mathbf{X}, \mathcal{F}_N, V)$ be the joint input-output probability for input $\mathbf{X}$ conditionally Gaussian with independent components of zero conditional mean and conditional variance as in (64). Notice that such input distribution is valid, in the sense that it is in the form (63).

For every subset $\mathcal{A}$ we have

$$
\begin{aligned}
& I(\mathbf{X}(\mathcal{A}); \mathbf{Y} | \mathbf{X}(\overline{\mathcal{A}}), \mathcal{F}_N, V) \\
& \quad = \; D\left(\Pr(\mathbf{Y}|\mathbf{X}, \mathcal{F}_N, V) \| \Phi(\mathbf{Y}|\mathbf{X}(\overline{\mathcal{A}}), \mathcal{F}_N, V) | \Pr(\mathbf{X}, \mathcal{F}_N, V)\right) \\
& \quad\quad - D\left(\Pr(\mathbf{Y}|\mathbf{X}(\overline{\mathcal{A}}), \mathcal{F}_N, V) \| \Phi(\mathbf{Y}|\mathbf{X}(\overline{\mathcal{A}}), \mathcal{F}_N, V) | \Pr(\mathbf{X}, \mathcal{F}_N, V)\right) \\
& \quad \overset{(a)}{\leq} \; D\left(\Pr(\mathbf{Y}|\mathbf{X}, \mathcal{F}_N, V) \| \Phi(\mathbf{Y}|\mathbf{X}(\overline{\mathcal{A}}), \mathcal{F}_N, V) | \Pr(\mathbf{X}, \mathcal{F}_N, V)\right) \\
& \quad = \; D\left(\mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \mathbf{I}) \| \mathcal{N}_{\mathbb{C}}(\boldsymbol{\nu}, \boldsymbol{\Lambda}) | \Pr(\mathbf{X}, \mathcal{F}_N, V)\right) \tag{65}
\end{aligned}
$$

where (a) follows from the non-negativity of divergence [19] and where we defined the conditional mean vectors of dimension $NL \times 1$ as

$$
\boldsymbol{\mu} = \begin{bmatrix} \sum_{k=1}^{K} c_{k,1} \boldsymbol{x}_{k,1} \\ \vdots \\ \sum_{k=1}^{K} c_{k,N} \boldsymbol{x}_{k,N} \end{bmatrix}, \quad \boldsymbol{\nu} = \begin{bmatrix} \sum_{k \notin \mathcal{A}} c_{k,1} \boldsymbol{x}_{k,1} \\ \vdots \\ \sum_{k \notin \mathcal{A}} c_{k,N} \boldsymbol{x}_{k,N} \end{bmatrix} \tag{66}
$$

and the conditional covariance matrix of dimension $NL \times NL$ as

$$\boldsymbol{\Lambda} = \mathrm{diag}\left(1 + \sum_{k \in \mathcal{A}} \alpha_{k,1} \beta_{k,1}^{(1)}(\mathcal{F}_1, V), \ldots, 1 + \sum_{k \in \mathcal{A}} \alpha_{k,N} \beta_{k,N}^{(L)}(\mathcal{F}_N, V)\right) \tag{67}$$

By applying the general formula for the divergence of two Gaussian complex circularly

symmetric distributions [18] we obtain

$$
\begin{aligned}
D\left(\mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \mathbf{I}) \| \mathcal{N}_{\mathbb{C}}(\boldsymbol{\nu}, \boldsymbol{\Lambda}) | \Pr(\mathbf{X}, \mathcal{F}_N, V)\right) & \\
= \quad & \mathrm{E}\left[\log \prod_{n=1}^{N} \prod_{\ell=1}^{L}\left(1+\sum_{k \in \mathcal{A}} \alpha_{k,1} \beta_{k,n}^{(\ell)}(\mathcal{F}_n, V)\right)\right. \\
& \left.+\sum_{n=1}^{N} \sum_{\ell=1}^{L} \frac{\left|\sum_{k \in \mathcal{A}} c_{k,n} x_{k,n}^{(\ell)}\right|^2 - \sum_{k \in \mathcal{A}} \alpha_{k,n} \beta_{k,n}^{(\ell)}(\mathcal{F}_n, V)}{1+\sum_{k \in \mathcal{A}} \alpha_{k,1} \beta_{k,n}^{(\ell)}(\mathcal{F}_n, V)}\right] \\
\stackrel{(a)}{=} \quad & \mathrm{E}\left[\mathrm{E}\left[\sum_{n=1}^{N} \sum_{\ell=1}^{L} \log \left(1+\sum_{k \in \mathcal{A}} \alpha_{k,1} \beta_{k,n}^{(\ell)}(\mathcal{F}_n, V)\right) \middle| \mathcal{F}_N\right]\right] \\
\stackrel{(b)}{\leq} \quad & \mathrm{E}\left[\mathrm{E}\left[\sum_{n=1}^{N} L \log \left(1+\sum_{k \in \mathcal{A}} \alpha_{k,1} \beta_{k,n}(\mathcal{F}_n, V)\right) \middle| \mathcal{F}_N\right]\right] \\
\stackrel{(c)}{\leq} \quad & \mathrm{E}\left[\sum_{n=1}^{N} L \log \left(1+\sum_{k \in \mathcal{A}} \alpha_{k,1} \beta_{k,n}(\mathcal{F}_n)\right)\right]
\end{aligned}
\tag{68}
$$

where (a) follows by taking conditional expectation with respect to $\mathbf{X}$, given $\mathcal{F}_N$ and V, and by using the fact that, from (63) the $\mathbf{X}_k$ are mutually independent given $\mathcal{F}_N$ and V, (b) follows by defining $\beta_{k,n}(\mathcal{F}_n, V) \stackrel{\Delta}{=} \frac{1}{L} \sum_{\ell=1}^{L} \beta_{k,n}^{(\ell)}(\mathcal{F}_n, V)$ and from Jensen's inequality applied to the concave function $\log(1+x)$, and (c) follows by defining $\beta_{k,n}(\mathcal{F}_n) \stackrel{\Delta}{=} \mathrm{E}[\beta_{k,n}(\mathcal{F}_n, V) | \mathcal{F}_n]$ and again from Jensen's inequality.

From (65) and (68) we have that

$$
\frac{1}{NL} I(\mathbf{X}(\mathcal{A}); \mathbf{Y} | \mathbf{X}(\overline{\mathcal{A}}), \mathcal{F}_N, V) \leq \mathrm{E}\left[\frac{1}{N} \sum_{n=1}^{N} \log \left(1+\sum_{k \in \mathcal{A}} \alpha_{k,1} \beta_{k,n}(\mathcal{F}_n)\right)\right]
\tag{69}
$$

and that the LHS of (69) is achieved by degenerate V (i.e., a constant) and $P(\mathbf{X} | \mathcal{F}_N, V)$ Gaussian with conditionally independent elements $x_{k,n}^{(\ell)} \sim \mathcal{N}_{\mathbb{C}}(0, \beta_{k,n}(\mathcal{F}_n))$. Since this holds for arbitrary $\mathcal{A}$ and input distribution $P(\mathbf{X}, \mathcal{F}_N, V)$, we conclude that (62) coincides with (6), thus proving the converse.

C Proof of Theorem 3

In order to fix ideas, we treat first the single-user case ($K = 1$). The proof of Theorem 3 follows by applying the same technique in the slightly more involved multiuser case.

For notation simplicity we drop the user index k. The single-user ergodic capacity is

given by

$$C_1^{(\text{erg})}(\gamma) = \max_{\boldsymbol{\beta}} \mathrm{E}\left[\log\left(1 + \alpha\beta(\alpha)\right)\right]$$
$$\text{s.t.} \quad \beta(\alpha) \geq 0 \quad \text{and} \quad \mathrm{E}[\beta(\alpha)] \leq \gamma \tag{70}$$

The single-user average capacity with causal CSI, delay N and per-codeword power constraint is given by

$$C_{1,N}(\gamma) = \max_{\boldsymbol{\beta}} \mathrm{E}\left[\frac{1}{N}\sum_{n=1}^{N}\log\left(1 + \alpha_n\beta_n(\mathcal{F}_n)\right)\right]$$
$$\text{s.t.} \quad \beta_n(\mathcal{F}_n) \geq 0 \quad \text{and} \quad \sum_{n=1}^{N}\beta_n(\mathcal{F}_n) \leq N\gamma \tag{71}$$

while the single-user average capacity with non-causal CSI, delay N and "long-term" power constraint is given by

$$C_{1,N}^{(\text{LT}-\text{nc})}(\gamma) = \max_{\boldsymbol{\beta}} \mathrm{E}\left[\frac{1}{N}\sum_{n=1}^{N}\log\left(1 + \alpha_n\beta_n(\mathcal{F}_N)\right)\right]$$
$$\text{s.t.} \quad \beta_n(\mathcal{F}_N) \geq 0 \quad \text{and} \quad \mathrm{E}\left[\sum_{n=1}^{N}\beta_n(\mathcal{F}_N)\right] \leq N\gamma \tag{72}$$

When user k is considered, the mean values in (70), (71) and (72) are computed with respect to α_n i.i.d. $\sim F_\alpha^{(k)}(x)$ and for $\gamma = \gamma_k$.

Problem (70) has solution [1]

$$C_1^{(\text{erg})}(\gamma) = \mathrm{E}\left[\log\left(1 + \alpha\beta^{(\text{erg})}(\alpha;\gamma)\right)\right] = \mathrm{E}\left[\log\left(\frac{\alpha}{\lambda}\right)^+\right] \tag{73}$$

where $\beta^{(\text{erg})}(\alpha;\gamma)$ is the ergodic waterfilling power allocation

$$\beta^{(\text{erg})}(\alpha;\gamma) = \left[\frac{1}{\lambda} - \frac{1}{\alpha}\right]^+ \tag{74}$$

and the Lagrange multiplier λ satisfies

$$\mathrm{E}\left[\beta^{(\text{erg})}(\alpha;\gamma)\right] = \gamma \tag{75}$$

It is immediate to see that, for every N,

$$C_{1,N}(\gamma) \leq C_{1,N}^{(\text{LT}-\text{nc})}(\gamma) = C_1^{(\text{erg})}(\gamma) \tag{76}$$

where the inequality in (76) follows since the set of feasible causal power allocations is a subset of the set of feasible "long-term" non-causal power allocations, and the equality

30

in (76) follows straightforwardly. It is also easy to see that, since $C_1^{(\mathrm{erg})}(\gamma)$ is a non-decreasing continuous function of γ, for every $\epsilon > 0$ it exists $\delta > 0$ such that

$$C_1^{(\mathrm{erg})}(\gamma) + \epsilon = C_1^{(\mathrm{erg})}(\gamma + \delta) \qquad (77)$$

Next, we find a lower bound on $C_{1,N}(\gamma)$ by choosing a particular causal power allocation policy, and we show that, in the limit for $N \to \infty$, the lower bound can be made arbitrarily close to the upper bound $C_1^{(\mathrm{erg})}(\gamma)$. For every N and for $\delta \in [0, \gamma]$, consider the (suboptimal) power allocation $\widetilde{\boldsymbol{\beta}} \in \Gamma_{1,N}(\gamma)$ defined by

$$\widetilde{\beta}_n(\mathcal{F}_n) = \begin{cases} \beta^{(\mathrm{erg})}(\alpha_n; \gamma - \delta) & \text{if} \quad \sum_{i=1}^n \widetilde{\beta}_i(\mathcal{S}_i) \leq N\gamma \\ 0 & \text{otherwise} \end{cases} \qquad (78)$$

Hence, the desired lower bound is given by

$$\mathrm{E}\left[\left(\frac{1}{N}\sum_{n=1}^N \log\left(1 + \alpha_n \beta^{(\mathrm{erg})}(\alpha_n; \gamma - \delta)\right)\right) 1\left\{\frac{1}{N}\sum_{n=1}^N \beta^{(\mathrm{erg})}(\alpha_n; \gamma - \delta) \leq \gamma\right\}\right] \leq C_{1,N}(\gamma) \quad (79)$$

Notice that $\{\log\left(1 + \alpha_n \beta^{(\mathrm{erg})}(\alpha_n; \gamma - \delta)\right)\}_{n=1}^N$ and $\{\beta^{(\mathrm{erg})}(\alpha_n; \gamma - \delta)\}_{n=1}^N$ are i.i.d. sequences. Since $\mathrm{E}[\beta^{(\mathrm{erg})}(\alpha_n; \gamma - \delta)] = \gamma - \delta$ by definition (75) and because of the law of large numbers, the indicator function $1\left\{\frac{1}{N}\sum_{n=1}^N \beta^{(\mathrm{erg})}(\alpha_n; \gamma - \delta) \leq \gamma\right\}$ tends to the constant value 1 almost surely, for $N \to \infty$. For the same reasons, $\frac{1}{N}\sum_{n=1}^N \log\left(1 + \alpha_n \beta^{(\mathrm{erg})}(\alpha_n; \gamma - \delta)\right)$ tends to $\mathrm{E}[\log\left(1 + \alpha_n \beta^{(\mathrm{erg})}(\alpha_n; \gamma - \delta)\right)] = C_1^{(\mathrm{erg})}(\gamma - \delta)$ almost surely, for $N \to \infty$. Hence, because of (77), we have that the RHS of (79) converges almost surely to $C_1^{(\mathrm{erg})}(\gamma) - \epsilon$ for some $\epsilon > 0$. Finally, since

$$C_1^{(\mathrm{erg})}(\gamma) - \epsilon \leq \lim_{N\to\infty} C_{1,N}(\gamma) \leq C_1^{(\mathrm{erg})}(\gamma) \qquad (80)$$

holds for every $\epsilon > 0$, we have that

$$\lim_{N\to\infty} C_{1,N}(\gamma) = C_1^{(\mathrm{erg})}(\gamma) \qquad (81)$$

In order to extend this result to the multiuser case and prove the statement of Theorem 3, we consider the explicit characterization of the boundary of $C_K^{(\mathrm{erg})}(\boldsymbol{\gamma})$ given in [5]. A rate K-tuple $\boldsymbol{R} = (R_1, \dots, R_K)$ is on the boundary surface of $C_K^{(\mathrm{erg})}(\boldsymbol{\gamma})$ if it is the solution of

$$\max_{\boldsymbol{R} \in C_K^{(\mathrm{erg})}(\boldsymbol{\gamma})} \sum_{k=1}^K \mu_k R_k \qquad (82)$$

for some $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_K) \in \mathbb{R}_+^K$. A point $(R_1^{(\mathrm{erg})}(\boldsymbol{\mu}, \boldsymbol{\gamma}), \cdots, R_K^{(\mathrm{erg})}(\boldsymbol{\mu}, \boldsymbol{\gamma}))$ is solution of the above problem if it exists a vector of Lagrangian multipliers $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_K) \in \mathbb{R}_+^K$ such that

$$\mathrm{E}[\beta_k^{(\mathrm{erg})}(\boldsymbol{\alpha}; \boldsymbol{\mu}, \boldsymbol{\gamma})] = \gamma_k \tag{83}$$

$$\mathrm{E}[r_k^{(\mathrm{erg})}(\boldsymbol{\alpha}; \boldsymbol{\mu}, \boldsymbol{\gamma})] = R_k^{(\mathrm{erg})}(\boldsymbol{\mu}, \boldsymbol{\gamma}) \tag{84}$$

where the average is with respect to $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_K)$ and

$$\begin{cases} u_k(z) \triangleq \dfrac{\mu_k}{1+z} - \dfrac{\lambda_k}{\alpha_k} \quad \text{for} \quad z \in \mathbb{R}_+ \\ u^\star(z) \triangleq \max\limits_{k=1,\cdots,K} \left\{ [u_k(z)]^+ \right\} \\ \beta_k^{(\mathrm{erg})}(\boldsymbol{\alpha}; \boldsymbol{\mu}, \boldsymbol{\gamma}) \triangleq \dfrac{1}{\alpha_k} \displaystyle\int 1\{u_k(z) = u^\star(z)\}\mathrm{d}z \\ r_k^{(\mathrm{erg})}(\boldsymbol{\alpha}; \boldsymbol{\mu}, \boldsymbol{\gamma}) \triangleq \displaystyle\int \dfrac{1}{1+z} 1\{u_k(z) = u^\star(z)\}\mathrm{d}z \end{cases} \tag{85}$$

Note that $r_k^{(\mathrm{erg})}(\boldsymbol{\alpha}; \boldsymbol{\mu}, \boldsymbol{\gamma})$ and $\beta_k^{(\mathrm{erg})}(\boldsymbol{\alpha}; \boldsymbol{\mu}, \boldsymbol{\gamma})$ are, respectively, the instantaneous rate and instantaneous power allocated to user k in fading state $\boldsymbol{\alpha}$. It is clear that if $\boldsymbol{\gamma}_1 \leq \boldsymbol{\gamma}_2$ then $C_K^{(\mathrm{erg})}(\boldsymbol{\gamma}_1) \subseteq C_K^{(\mathrm{erg})}(\boldsymbol{\gamma}_2)$ and for any $\boldsymbol{\mu} \in \mathbb{R}_+^K$

$$\max_{\boldsymbol{R} \in C_{K,N}(\boldsymbol{\gamma}_1)} \sum_{k=1}^K \mu_k \, R_k \leq \max_{\boldsymbol{R} \in C_{K,N}(\boldsymbol{\gamma}_2)} \sum_{k=1}^K \mu_k \, R_k \tag{86}$$

Conversely, if (86) holds for any direction vector $\boldsymbol{\mu}$, then $C_K^{(\mathrm{erg})}(\boldsymbol{\gamma}_1) \subseteq C_K^{(\mathrm{erg})}(\boldsymbol{\gamma}_2)$ and $\boldsymbol{\gamma}_1 \leq \boldsymbol{\gamma}_2$.

With arguments analogous to the single-user case, we can show that the upper bound $C_{K,N}(\boldsymbol{\gamma}) \subseteq C_{K,N}^{(\mathrm{LT-nc})}(\boldsymbol{\gamma}) \equiv C_K^{(\mathrm{erg})}(\boldsymbol{\gamma})$ holds for every delay N. For an arbitrary direction $\boldsymbol{\mu} \in \mathbb{R}_+^K$, an inner bound to $C_{K,N}(\boldsymbol{\gamma})$ is obtained by fixing the allocation policy $\widetilde{\boldsymbol{\beta}}$ as follows: for given $\boldsymbol{\delta} \in \mathbb{R}_+^K$ such that $\boldsymbol{\gamma} - \boldsymbol{\delta} \geq 0$, we define

$$\widetilde{\beta}_{k,n}(\mathcal{F}_n) = \begin{cases} \beta_k^{(\mathrm{erg})}(\boldsymbol{\alpha}_n; \boldsymbol{\mu}, \boldsymbol{\gamma} - \boldsymbol{\delta}) & \text{if } \sum_{j=1}^n \widetilde{\beta}_{k,j}(\mathcal{S}_j) \leq N \gamma_k \\ 0 & \text{otherwise} \end{cases} \tag{87}$$

The inner bound implies that

$$\mathrm{E}\left[\left(\sum_{k=1}^K \mu_k \frac{1}{N} \sum_{n=1}^N r_k^{(\mathrm{erg})}(\boldsymbol{\alpha}_n; \boldsymbol{\mu}, \boldsymbol{\gamma} - \boldsymbol{\delta}) \right) \prod_{k=1}^K 1\left\{ \frac{1}{N} \sum_{n=1}^N \beta_k^{(\mathrm{erg})}(\boldsymbol{\alpha}_n; \boldsymbol{\mu}, \boldsymbol{\gamma} - \boldsymbol{\delta}) \leq \gamma_k \right\} \right]$$
$$\leq \sum_{k=1}^K \mu_k \widehat{R}_{k,N}(\boldsymbol{\mu}, \boldsymbol{\gamma}) \tag{88}$$

where $\widehat{R}_{k,N}(\boldsymbol{\mu}, \boldsymbol{\gamma})$ are the rates on the boundary surface of $C_{K,N}(\boldsymbol{\gamma})$, given in (10). Now, since both $\{\sum_{k=1}^K \mu_k r_k^{(\mathrm{erg})}(\boldsymbol{\alpha}_n; \boldsymbol{\mu}, \boldsymbol{\gamma} - \boldsymbol{\delta})\}_{n=1}^N$ and $\{\beta_k^{(\mathrm{erg})}(\boldsymbol{\alpha}_n; \boldsymbol{\mu}, \boldsymbol{\gamma} - \boldsymbol{\delta})\}_{n=1}^N$ are i.i.d. sequences, the indicator functions in the RHS of (88) tend to the constant value 1 almost

surely and the sum of instantaneous rates tends to $\sum_{k=1}^{K} \mu_k R_k^{(\text{erg})}(\boldsymbol{\mu}, \boldsymbol{\gamma} - \boldsymbol{\delta})$ almost surely, as $N \to \infty$. Again, the RHS of (88) converges almost surely to $\sum_{k=1}^{K} \mu_k R_k^{(\text{erg})}(\boldsymbol{\mu}, \boldsymbol{\gamma} - \boldsymbol{\delta})$ and hence

$$\sum_{k=1}^{K} \mu_k R_k^{(\text{erg})}(\boldsymbol{\mu}, \boldsymbol{\gamma} - \boldsymbol{\delta}) \leq \lim_{N \to \infty} \sum_{k=1}^{K} \mu_k \widehat{R}_{k,N}(\boldsymbol{\mu}, \boldsymbol{\gamma}) \leq \sum_{k=1}^{K} \mu_k R_k^{(\text{erg})}(\boldsymbol{\mu}, \boldsymbol{\gamma}) \tag{89}$$

Since $\boldsymbol{\delta}$ is arbitrary and (89) holds for any $\boldsymbol{\mu}$, we conclude that

$$\lim_{N \to \infty} C_{K,N}(\boldsymbol{\gamma}) \equiv C_K^{(\text{erg})}(\boldsymbol{\gamma}) \tag{90}$$

D   Proof of Theorem 5

In the following we indicate with $C_{1,N}^{(k)}(\gamma)$ the single-user average capacity for user k as defined in (71), where the extra superscript "(k)" stresses the fact that the mean value is computed using cdf $F_\alpha^{(k)}(x)$. Note that $C_{1,N}^{(k)}(\gamma) = \widehat{R}_{k,N}(\mathbf{1}_k, \boldsymbol{\gamma})$ for $\widehat{R}_{k,N}(\boldsymbol{\mu}, \boldsymbol{\gamma})$ defined in (10) and where $\mathbf{1}_k$ is the vector of length K of all zeros but a "1" in position k.

Consider the following inner and outer bounds for $C_{K,N}(\boldsymbol{\gamma})$

$$\left\{ \boldsymbol{R} \in \mathbb{R}_+^K : R_k \leq \frac{1}{K} C_{1,N}^{(k)}(K \gamma_k) \right\} \subseteq C_{K,N}(\boldsymbol{\gamma}) \subseteq \left\{ \boldsymbol{R} \in \mathbb{R}_+^K : R_k \leq C_{1,N}^{(k)}(\gamma_k) \right\} \tag{91}$$

where the inner bound is clearly achievable by TDMA, i.e., by letting each user transmit for a fraction $1/K$ of the slot time, and the outer bound is the Cartesian product of the single user average capacity regions. Theorem 4 implies the following inner and outer bounds for $U_{K,N}$

$$\left\{ \boldsymbol{r} \in \mathbb{R}_+^K : r_k \leq \sup_{\gamma_k > 0} \frac{1}{K \gamma_k} C_{1,N}^{(k)}(K \gamma_k) \right\} \subseteq U_{K,N} \subseteq \left\{ \boldsymbol{r} \in \mathbb{R}_+^K : r_k \leq \sup_{\gamma_k > 0} \frac{1}{\gamma_k} C_{1,N}^{(k)}(\gamma_k) \right\} \tag{92}$$

Define the feasible power allocation policy

$$(\beta_{k,1}^\star, \cdots, \beta_{k,N}^\star) = \arg \sup_{\boldsymbol{\beta} \in \Gamma_{1,N}(\gamma_k)} \mathrm{E} \left[\frac{1}{N} \sum_{n=1}^{N} \alpha_{k,n} \beta_{k,n}(\mathcal{F}_n) \right] \tag{93}$$

and indicate with $(\widehat{\beta}_{k,1}, \cdots, \widehat{\beta}_{k,N})$ the k-th user single-user average capacity achieving

policy. The boundary surface of the outer region in (92) is given by

$$
\begin{aligned}
\sup_{\gamma_k > 0} \frac{1}{\gamma_k} C_{1,N}^{(k)}(\gamma_k) &= \sup_{\gamma_k > 0} \frac{\mathrm{E}\left[\dfrac{1}{N} \displaystyle\sum_{n=1}^{N} \log\left(1 + \alpha_{k,n}\widehat{\beta}_{k,n}\right)\right]}{\gamma_k} \\
&\overset{(a)}{=} \lim_{\gamma_k \to 0} \frac{\mathrm{E}\left[\dfrac{1}{N} \displaystyle\sum_{n=1}^{N} \log\left(1 + \alpha_{k,n}\widehat{\beta}_{k,n}\right)\right]}{\gamma_k} \\
&\overset{(b)}{=} \lim_{\gamma_k \to 0} \frac{\mathrm{E}\left[\dfrac{1}{N} \displaystyle\sum_{n=1}^{N} \alpha_{k,n}\beta_{k,n}^{\star}\right]}{\gamma_k} \triangleq s_N^{(k)}
\end{aligned}
\tag{94}
$$

where: (a) follows since $C_{1,N}^{(k)}(\gamma_k)$ is concave in γ_k (see the Corollary to Lemma 1 at the end of this section) and (b) follows for Lemma 2 at the end of this section.

With similar steps, we find that the boundary surface of the inner region in (92) is also given by (94). We conclude that the K-user average capacity region per unit energy is the hyper-rectangle

$$
U_{K,N} = \left\{ \boldsymbol{r} \in \mathbb{R}_+^K : r_k \leq s_N^{(k)} \right\}.
\tag{95}
$$

for $s_N^{(k)}$ given in (94) and that $\boldsymbol{\beta}^{\star} = \{\boldsymbol{\beta}_{k,n}^{\star} \ : \ k = 1, \cdots, K, n = 1, \cdots, N\}$ is the optimal K-user average capacity region per unit energy achieving policy. $\qquad\square$

In the following we drop the superscript "(k)" since no confusion may arise.

Lemma 1. $C_{1,N}(\gamma)$ given in (71) is a concave function of γ.
Proof. Consider the single-user average capacity achieving power allocation that, for notation convenience, we re-write as follows

$$
\left(\widehat{\beta}_1(\mathcal{F}_1; \gamma), \cdots, \widehat{\beta}_N(\mathcal{F}_N; \gamma)\right) = \arg \sup_{\boldsymbol{\beta} \in \Gamma_{1,N}(\gamma)} \mathrm{E}\left[\frac{1}{N} \sum_{n=1}^{N} \log\left(1 + \alpha_n \beta_n(\mathcal{F}_n)\right)\right]
\tag{96}
$$

to explicitly denote the dependency on the constraint γ. For every $\lambda \in [0,1]$ and for every $\gamma_a, \gamma_b \geq 0$ consider the convex combination

$$
\begin{aligned}
\lambda C_{1,N}(\gamma_a) + (1-\lambda)C_{1,N}(\gamma_b) &= \lambda \, \mathrm{E}\left[\frac{1}{N}\sum_{n=1}^{N}\log\left(1 + \alpha_n \widehat{\beta}_n(\mathcal{F}_n; \gamma_a)\right)\right] \\
&\quad + (1-\lambda)\,\mathrm{E}\left[\frac{1}{N}\sum_{n=1}^{N}\log\left(1 + \alpha_n \widehat{\beta}_n(\mathcal{F}_n; \gamma_b)\right)\right] \\
&\stackrel{(a)}{\leq} \frac{1}{N}\sum_{n=1}^{N}\mathrm{E}\left[\log\left(1 + \alpha_n \lambda \widehat{\beta}_n(\mathcal{F}_n; \gamma_a) + \alpha_n(1-\lambda)\widehat{\beta}_n(\mathcal{F}_n; \gamma_b)\right)\right] \\
&\stackrel{(b)}{\leq} \frac{1}{N}\sum_{n=1}^{N}\mathrm{E}\left[\log\left(1 + \alpha_n \widehat{\beta}_n(\mathcal{F}_n; \lambda\gamma_a + (1-\lambda)\gamma_b)\right)\right] \\
&= C_{1,N}(\lambda\gamma_a + (1-\lambda)\gamma_b)
\end{aligned}
\tag{97}
$$

where: (a) follows from Jensen's inequality and (b) because the feasible power policy $\lambda\widehat{\boldsymbol{\beta}}(\cdot; \gamma_a) + (1-\lambda)\widehat{\boldsymbol{\beta}}(\cdot; \gamma_b)$ does not coincide in general with the optimal power allocation (96) for $\gamma = \lambda\gamma_a + (1-\lambda)\gamma_b$. $\qquad\square$

Corollary. Since $C_{1,N}(\gamma)$ is non-negative and concave we have

$$
\sup_{\gamma > 0} \frac{C_{1,N}(\gamma)}{\gamma} = \dot{C}_{1,N}(0)
\tag{98}
$$

where $\dot{C}_{1,N}(0)$ denotes the first derivative of $C_{1,N}(\gamma)$ at $\gamma = 0$.

In fact, since $C_{1,N}(\gamma)$ is concave, its second derivative is non-positive, i.e., $\ddot{C}_{1,N}(\gamma) \leq 0$, and hence its first derivartive is non-increasing, i.e. $\dot{C}_{1,N}(\gamma) \leq \dot{C}_{1,N}(0)$. Since $C_{1,N}(\gamma)$ is non-negative, by integrating both sides of the inequality $\dot{C}_{1,N}(\gamma) \leq \dot{C}_{1,N}(0)$ and imposing the initial condition $C_{1,N}(0) = 0$ we get

$$
0 \leq C_{1,N}(\gamma) \leq \gamma \dot{C}_{1,N}(0)
\tag{99}
$$

hence (98) follows. $\qquad\square$

Lemma 2. Let $\widehat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \mathrm{E}\left[\frac{1}{N}\sum_{n=1}^{N}\log\left(1 + \alpha_n \beta_n(\mathcal{F}_n)\right)\right]$ and $\boldsymbol{\beta}^\star = \arg\max_{\boldsymbol{\beta}} \mathrm{E}\left[\frac{1}{N}\sum_{n=1}^{N}\alpha_n \beta_n(\mathcal{F}_n)\right]$, where in both case $\boldsymbol{\beta} \in \Gamma_{1,N}(\gamma)$. Then, the following relation holds

$$
\begin{aligned}
\mathrm{E}\left[\frac{1}{N}\sum_{n=1}^{N}\log\left(1 + \alpha_n \beta_n^\star\right)\right] &\stackrel{(a)}{\leq} \mathrm{E}\left[\frac{1}{N}\sum_{n=1}^{N}\log\left(1 + \alpha_n \widehat{\beta}_n\right)\right] \\
&\stackrel{(b)}{\leq} \mathrm{E}\left[\frac{1}{N}\sum_{n=1}^{N}\alpha_n \widehat{\beta}_n\right] \stackrel{(c)}{\leq} \mathrm{E}\left[\frac{1}{N}\sum_{n=1}^{N}\alpha_n \beta_n^\star\right]
\end{aligned}
\tag{100}
$$

where (a) and (c) follow by definition and (b) follows since $\log(1 + x) \leq x$ for $x \geq 0$. By recalling the definition of $\boldsymbol{\beta}^\star$ (see (20) and (21)), relation (100) implies

$$
\begin{aligned}
\mathrm{E}\left[\frac{1}{N\gamma}\sum_{n=1}^{N}\log\left(1 + N\gamma\alpha_n 1\{n^\star = n\}\right)\right] &\leq \mathrm{E}\left[\frac{1}{N\gamma}\sum_{n=1}^{N}\log\left(1 + \alpha_n\widehat{\beta}_n\right)\right] \\
&\leq \mathrm{E}\left[\sum_{n=1}^{N}\alpha_n 1\{n^\star = n\}\right]
\end{aligned}
\tag{101}
$$

and by letting $\gamma \to 0$, equality (94) follows.

E   Proof of Theorem 7

Relation (98) and definition (94) imply $\dot{C}_{1,N}^{(k)}(0) = s_N^{(k)}$ for every N. By using (81), we have

$$
\lim_{N\to\infty} s_N^{(k)} = \lim_{N\to\infty} \dot{C}_{1,N}^{(k)}(0) = \dot{C}_1^{(\mathrm{erg})}(0)
\tag{102}
$$

The single-user ergodic capacity is given by the waterfilling formula (73) parameterized by the Lagrangian multiplier λ satisfying (75). Hence, we have

$$
\begin{aligned}
\lim_{\gamma\to 0}\frac{\mathrm{d}C_1^{(\mathrm{erg})}(\gamma)}{\mathrm{d}\gamma} &= \lim_{\gamma\to 0}\left.\frac{\dfrac{\mathrm{d}\mathrm{E}\left[\left(\log\frac{\alpha}{\lambda}\right)^+\right]}{\mathrm{d}\lambda}}{\dfrac{\mathrm{d}\mathrm{E}\left[\left(\frac{1}{\lambda}-\frac{1}{\alpha}\right)^+\right]}{\mathrm{d}\lambda}}\right|_{\gamma=\mathrm{E}[(1/\lambda-1/\alpha)^+]} \\
&= \lim_{\lambda\to\sup\{\alpha\}}\lambda\,1\{\lambda \leq \sup\{\alpha\}\} = \sup\{\alpha\}
\end{aligned}
\tag{103}
$$

$\square$

F   Proof of Theorem 8

Let

$$
S_n(P) = \mathrm{E}\left[\max_{u\in[0,P]}\{\log\left(1 + \alpha\,u\right) + S_{n-1}(P - u)\}\right]
\tag{104}
$$

and

$$
\widehat{u}_n(\alpha, P) = \arg\max_{u\in[0,P]}\{\log\left(1 + \alpha\,u\right) + S_{n-1}(P - u)\}
\tag{105}
$$

36

for $n = 1, \cdots, N$ and initial condition $S_0(P) = 0$.

In order to prove statement (31) we need to analyze in detail expression (104). Because of the concavity of $S_n(P)$ (from Lemma 1 in Appendix D since $S_n(P) = nC_{1,n}(P/n)$ from Theorem 2), $\widehat{u}_n(\alpha, P)$ in (105) can be written as

$$\widehat{u}_n(\alpha, P) = \begin{cases} 0 & \text{if } \alpha < \dot{S}_{n-1}(P) \\ P & \text{if } \frac{\alpha}{1 + \alpha P} \geq \dot{S}_{n-1}(0) \\ u_n^* & \text{elsewhere} \end{cases} \tag{106}$$

with u_n^* the unique solution of

$$\frac{\alpha}{1 + \alpha u_n^*} = \dot{S}_{n-1}(P - u_n^*) \tag{107}$$

The first and second derivative of $S_n(P)$ are given by

$$\begin{aligned} \dot{S}_n(P) &= \text{E}\left[\dot{S}_{n-1}(P) \, 1\{\widehat{u}_n(\alpha, P) = 0\} \right] \\ &\quad + \text{E}\left[\frac{\alpha}{1 + \alpha P} \, 1\{\widehat{u}_n(\alpha, P) = P\} \right] \\ &\quad + \text{E}\left[\frac{\alpha}{1 + \alpha u_n^*} \, 1\{\widehat{u}_n(\alpha, P) = u_n^*\} \right] \end{aligned} \tag{108}$$

and by

$$\begin{aligned} -\ddot{S}_n(P) &= \text{E}\left[\left(-\ddot{S}_{n-1}(P) \right) 1\{\widehat{u}_n(\alpha, P) = 0\} \right] \\ &\quad + \text{E}\left[\left(\frac{\alpha}{1 + \alpha P} \right)^2 1\{\widehat{u}_n(\alpha, P) = P\} \right] \\ &\quad + \text{E}\left[\left(\frac{\alpha}{1 + \alpha u_n^*} \right)^2 \frac{\partial u_n^*}{\partial P} 1\{\widehat{u}_n(\alpha, P) = u_n^*\} \right] \end{aligned} \tag{109}$$

Now, as $P \to 0$ we have

$$\widehat{u}_n(\alpha, P) = \begin{cases} 0 & \text{if } \alpha < \dot{S}_{n-1}(0) - (-\ddot{S}_{n-1}(0)) P + o(P) \\ P & \text{if } \alpha \geq \dot{S}_{n-1}(0) + (\dot{S}_{n-1}(0))^2 P + o(P) \\ P \left. \frac{\partial u_n^*}{\partial P} \right|_{P=0} & \text{elsewhere} \end{cases} \tag{110}$$

Hence, by substituting (110) in (109) and by letting $P \to 0$ we obtain

$$-\ddot{S}_n(0) = \text{E}\left[\alpha^2 \, 1\{\alpha \geq \dot{S}_{n-1}(0)\} \right] - \ddot{S}_{n-1}(0) \, \text{E}\left[1\{\alpha < \dot{S}_{n-1}(0)\} \right] \tag{111}$$

which coincides with (31).

Next, in order to prove the second-order optimality of the policy $\boldsymbol{\beta}^{\star}$, we show that the rate function $C_{1,N}^{\star}(\gamma)$, defined as

$$C_{1,N}^{\star}(\gamma) = \mathrm{E}\left[\frac{1}{N}\sum_{n=1}^{N}\log\left(1 + N\gamma\alpha_n\,1\{n^{\star} = n\}\right)\right] \tag{112}$$

obtained by applying $\boldsymbol{\beta}^{\star}$, has the first and second derivative at $\gamma = 0$ equal to those of $C_{1,N}(\gamma)$.

It follows immediately that the first and second derivative of (112) w.r.t. γ computed for $\gamma = 0$ are

$$\dot{C}_{1,N}^{\star}(0) = \mathrm{E}\left[\sum_{n=1}^{N}\alpha_n 1\{n^{\star} = n\}\right] \tag{113}$$

$$-\frac{1}{N}\ddot{C}_{1,N}^{\star}(0) = \mathrm{E}\left[\sum_{n=1}^{N}\alpha_n^2 1\{n^{\star} = n\}\right] \tag{114}$$

From the proof of Theorem 5 it follows that $\dot{C}_{1,N}^{\star}(0) = \dot{C}_{1,N}(0)$, i.e., $\boldsymbol{\beta}^{\star}$ achieves $(E_b/N_0)_{\min}$. Next we show that (114) is equal to $-\ddot{S}_N(0)$ which implies $\ddot{C}_{1,N}^{\star}(0) = \ddot{C}_{1,N}(0)$. In order to show the identity of the second order derivatives we shall show that (114) can be computed by a recursion identical to (111).

The probability that transmission occurs in slot n is

$$\Pr(n^{\star} = n) = \Pr\left(\alpha_n \geq s_{N-n}\right)\prod_{j=1}^{n-1}\Pr\left(\alpha_j < s_{N-j}\right) \tag{115}$$

Obviously, $\sum_{n=1}^{N}\Pr(n^{\star} = n) = 1$. For every $n = 1,\cdots,N$, the cdf of $\alpha_n 1\{n^{\star} = n\}$ is given by

$$
\begin{aligned}
&\Pr(\alpha_n 1\{n^{\star} = n\} \leq x) \\
&= \Pr(\alpha_n 1\{n^{\star} = n\} \leq x | n^{\star} = n)\Pr(n^{\star} = n) \\
&\quad + \Pr(\alpha_n 1\{n^{\star} = n\} \leq x | n^{\star} \neq n)\Pr(n^{\star} \neq n) \\
&= \Pr(\alpha_n \leq x | n^{\star} = n)\Pr(n^{\star} = n) + \Pr(0 \leq x | n^{\star} \neq n)\Pr(n^{\star} \neq n) \\
&= \Pr(\alpha_n \leq x | n^{\star} = n)\Pr(n^{\star} = n) + \Pr(n^{\star} \neq n) \quad \text{for} \quad x \geq 0 \tag{116}
\end{aligned}
$$

By recalling the expression of $\Pr(n^{\star} = n)$ in (115) we finally get

$$
\begin{aligned}
\Pr(\alpha_n 1\{n^{\star} = n\} \leq x) =\ & \prod_{j=1}^{n-1}\Pr(\alpha_j < s_{N-j})\Pr(\alpha_n \leq x, \alpha_n \geq s_{N-n}) \\
& + \left(1 - \Pr(\alpha_n \geq s_{N-n})\prod_{j=1}^{n-1}\Pr(\alpha_j < s_{N-j})\right) \tag{117}
\end{aligned}
$$

and hence, for every $r \in \mathbb{N}$ such that the r-th moment of $F_\alpha(x)$ exists, we have

$$
\mathrm{E}\left[\alpha_n^r 1\{n^\star = n\}\right] = \prod_{j=1}^{n-1} F_\alpha(s_{N-j}) \int_{s_{N-n}}^\infty x^r \mathrm{d}F_\alpha(x) \tag{118}
$$

By summing the terms in (118) over $n = 1, \cdots, N$ for $r = 1$ and $r = 2$ we get respectively (113) and (114). Let $\mu_N(r) \triangleq \sum_{n=1}^N \mathrm{E}\left[\alpha_n^r 1\{n^\star = n\}\right]$, then by using (118) we have

$$
\begin{aligned}
\mu_N(r) &= \sum_{n=1}^N \prod_{j=1}^{n-1} F_\alpha(s_{N-j}) \int_{s_{N-n}}^\infty x^r \, \mathrm{d}F_\alpha(x) \\
&= \int_{s_{N-1}}^\infty x^r \, \mathrm{d}F_\alpha(x) + \sum_{n=2}^N \prod_{j=1}^{n-1} F_\alpha(s_{N-j}) \int_{s_{N-n}}^\infty x^r \, \mathrm{d}F_\alpha(x) \\
&= \int_{s_{N-1}}^\infty x^r \, \mathrm{d}F_\alpha(x) + F_\alpha(s_{N-1}) \, \mu_{N-1}(r) \\
&= \mathrm{E}\left[\alpha_N^r \, 1\{\alpha_N \geq s_{N-1}\}\right] + \mathrm{E}\left[1\{\alpha < s_{N-1}\}\right] \mu_{N-1}(r) \tag{119}
\end{aligned}
$$

Since $\dot{S}_n(0) = s_n$ for all n and that $-\ddot{S}_N(0)$ and $\mu_N(2)$ satisfy the same recursion and have the same initial condition for $N = 0$, they coincide for all N. This concludes the proof.

Remark. The cdf (116) can be used to compute recursively $C_{1,N}^\star(\gamma)$ as defined in (112) for all γ. In fact, with initial condition $S_0^\star(P) = 0$, we have

$$
\begin{aligned}
S_N^\star(P) &\triangleq \mathrm{E}\left[\sum_{n=1}^N \log\left(1 + P\alpha_n \, 1\{n^\star = n\}\right)\right] \\
&= \sum_{n=1}^N \prod_{j=1}^{n-1} \Pr[\alpha_j < s_{N-j}] \int_{s_{N-n}}^\infty \log\left(1 + Px\right) \mathrm{d}F_\alpha(x) \\
&= \int_{s_{N-1}}^\infty \log\left(1 + Px\right) \mathrm{d}F_\alpha(x) + \Pr[\alpha_j < s_{N-1}] S_{N-1}^\star(P) \tag{120}
\end{aligned}
$$

and $C_{1,N}^\star(\gamma) = \frac{1}{N} S_N^\star(N\gamma)$.

G   Proof of Theorem 10

Consider the following inner and outer bound to the average capacity region

$$
\left\{\boldsymbol{R} \in \mathbb{R}_+^K : \forall \mathcal{A} \quad \sum_{j \in \mathcal{A}} R_j \leq g^{(\mathcal{A})}\right\} \subseteq C_{K,N}(\boldsymbol{\gamma}) \subseteq \left\{\boldsymbol{R} \in \mathbb{R}_+^K : \forall \mathcal{A} \quad \sum_{j \in \mathcal{A}} R_j \leq f^{(\mathcal{A})}\right\} \tag{121}
$$

where we define the set functions

$$f^{(\mathcal{A})} \triangleq \mathrm{E}\left[\frac{1}{N}\sum_{n=1}^{N}\log\left(1+\sum_{j\in\mathcal{A}}\alpha_{j,n}\beta_{j,n}^{(\mathcal{A})}\right)\right] \tag{122}$$

where

$$\{\beta_{j,n}^{(\mathcal{A})} : j\in\mathcal{A}, n=1,\cdots,N\} \triangleq \arg\max_{\boldsymbol{\beta}\in\Gamma_{K,N}(\boldsymbol{\gamma})}\mathrm{E}\left[\frac{1}{N}\sum_{n=1}^{N}\log\left(1+\sum_{j\in\mathcal{A}}\alpha_{j,n}\beta_{j,n}\right)\right] \tag{123}$$

and

$$g^{(\mathcal{A})} \triangleq \mathrm{E}\left[\frac{1}{N}\sum_{n=1}^{N}\log\left(1+\sum_{j\in\mathcal{A}}\alpha_{j,n}\beta_{j,n}^{\star}\right)\right] \tag{124}$$

for all $\mathcal{A}\subseteq\{1,\cdots,K\}$. The inner bound in (121) is the average achievable region when users apply the one-shot policy $\boldsymbol{\beta}^{\star}$ and the outer bound in (121) is obtained by applying the "max-flow-min-cut" theorem for multiterminal networks [19, Theorem 14.10.1] to our system.

Before proceeding, we point out some characteristics of the set functions $g^{(\mathcal{A})}$ and $f^{(\mathcal{A})}$. First, they do not depend on the whole SNR vector $\boldsymbol{\gamma}=(\gamma_1,\cdots,\gamma_K)$ but only on $\{\gamma_j\}_{j\in\mathcal{A}}$. Second, by recalling that $\beta_{j,n}^{\star} = N\,\gamma_j\,1\{n_j^{\star}=n\}$ for $n_j^{\star}$ defined in (21), it easy to see that, in the limit for $\boldsymbol{\gamma}\to\mathbf{0}$, the first-order partial derivative of $g^{(\mathcal{A})}(\{\gamma_j\}_{j\in\mathcal{A}})$ w.r.t. γ_ℓ for all $\ell\in\mathcal{A}$ is given by

$$\partial_\ell g^{(\mathcal{A})}(\mathbf{0}) = \sum_{n=1}^{N}\mathrm{E}\left[\alpha_{\ell,n}1\{n_\ell^{\star}=n\}\right] \tag{125}$$

and that the second-order partial derivative of $g^{(\mathcal{A})}(\{\gamma_j\}_{j\in\mathcal{A}})$ w.r.t. γ_ℓ and γ_m for all $\ell, m\in\mathcal{A}$ is given by

$$\partial_{\ell,m}g^{(\mathcal{A})}(\mathbf{0}) = -N\sum_{n=1}^{N}\mathrm{E}\left[\alpha_{\ell,n}1\{n_\ell^{\star}=n\}\alpha_{m,n}1\{n_m^{\star}=n\}\right] \tag{126}$$

Notice that, since $n_\ell^{\star}$ only depends on the fading sequence of user ℓ, in equation (126) the mean value factors when $\ell\neq m$. From Theorem 8 we have

$$\partial_\ell g^{(\mathcal{A})}(\mathbf{0}) = s_N^{(\ell)} = \dot{C}_{1,N}^{(\ell)}(0) \tag{127}$$

$$\partial_{\ell,\ell}g^{(\mathcal{A})}(\mathbf{0}) = \ddot{C}_{1,N}^{(\ell)}(0) \tag{128}$$

where $C_{1,N}^{(\ell)}(\gamma_\ell)$ is the ℓ-th user single-user average capacity. Hence, we can write the single user slope as $S_0^{(\ell)} = -2(\partial_\ell g^{(\mathcal{A})}(\mathbf{0}))^2 / \partial_{\ell,\ell} g^{(\mathcal{A})}(\mathbf{0})$.

Now we derive an achievable slope region based on the inner bound in (121). For a given permutation $\boldsymbol{\pi} = (\pi_1 \cdots, \pi_K)$ of $\{1, \ldots, K\}$, corresponding to the decoding order $\pi_K, \pi_{K-1}, \cdots, \pi_1$, we have the following vertex of the inner bound region

$$R_{\pi_k}(\boldsymbol{\pi}) = g^{\{\pi_1 \cdots, \pi_k\}} - g^{\{\pi_1 \cdots, \pi_{k-1}\}} \tag{129}$$

Every point on the dominant face of the inner bound region can be expressed as a convex combination of the $K!$ vertices, of coordinates (129), as follow

$$R_k = \sum_{\boldsymbol{\pi}} \lambda_{\boldsymbol{\pi}} R_{\pi_{\pi^{-1}(k)}}(\boldsymbol{\pi}) \tag{130}$$

where $\pi^{-1}(k)$ gives the position of the integer k in the permuted vector $\boldsymbol{\pi}$, where $\sum_{\boldsymbol{\pi}}$ denotes the sum over the $K!$ permutations of $\{1, \ldots, K\}$ and where $\boldsymbol{\lambda} = \{\lambda_{\boldsymbol{\pi}}\}$ are non-negative "time-sharing" coefficients (indexed by the permutations $\boldsymbol{\pi}$) such that $\sum_{\boldsymbol{\pi}} \lambda_{\boldsymbol{\pi}} = 1$.

For fixed $(\theta_1, \cdots, \theta_K) \in \mathbb{R}_+^K$ we let $\gamma_k/\gamma_j = \theta_k/\theta_j$ for all $i, j \in \{1, \cdots, K\}$ and we compute the derivatives of $R_{\pi_k}(\boldsymbol{\pi})$ in (129), expressed as a function of γ_{π_k}, that for notation simplicity we indicate with x. The rate is given by

$$R_{\pi_k}(\boldsymbol{\pi}) = g^{\{\pi_1 \cdots, \pi_k\}} \left(\frac{\theta_{\pi_1}}{\theta_{\pi_k}} x, \cdots, \frac{\theta_{\pi_{k-1}}}{\theta_{\pi_k}} x, x \right) - g^{\{\pi_1 \cdots, \pi_{k-1}\}} \left(\frac{\theta_{\pi_1}}{\theta_{\pi_k}} x, \cdots, \frac{\theta_{\pi_{k-1}}}{\theta_{\pi_k}} x \right) \tag{131}$$

Its first derivative is

$$\dot{R}_{\pi_k}(\boldsymbol{\pi}) = \sum_{j=1}^{k-1} \frac{\theta_{\pi_j}}{\theta_{\pi_k}} \partial_{\pi_j} \left[g^{\{\pi_1 \cdots, \pi_k\}} - g^{\{\pi_1 \cdots, \pi_{k-1}\}} \right] + \partial_{\pi_k} g^{\{\pi_1 \cdots, \pi_k\}} \tag{132}$$

and its second derivative is

$$\ddot{R}_{\pi_k}(\boldsymbol{\pi}) = \sum_{j=1}^{k-1} \sum_{\ell=1}^{k-1} \frac{\theta_{\pi_j}}{\theta_{\pi_k}} \frac{\theta_{\pi_\ell}}{\theta_{\pi_k}} \partial_{\pi_j, \pi_\ell} \left[g^{\{\pi_1 \cdots, \pi_k\}} - g^{\{\pi_1 \cdots, \pi_{k-1}\}} \right]$$

$$+ 2 \sum_{j=1}^{k-1} \frac{\theta_{\pi_j}}{\theta_{\pi_k}} \partial_{\pi_j, \pi_k} g^{\{\pi_1 \cdots, \pi_k\}} + \partial_{\pi_k, \pi_k} g^{\{\pi_1 \cdots, \pi_k\}} \tag{133}$$

In the limit for $x \to 0$ we get

$$\lim_{x \to 0} \dot{R}_{\pi_k}(\boldsymbol{\pi}) = \partial_{\pi_k} g^{\{\pi_1 \cdots, \pi_k\}}(\mathbf{0}) \tag{134}$$

$$\lim_{x \to 0} \ddot{R}_{\pi_k}(\boldsymbol{\pi}) = \partial_{\pi_k, \pi_k} g^{\{\pi_1 \cdots, \pi_k\}}(\mathbf{0}) + 2 \sum_{j=1}^{k-1} \frac{\theta_{\pi_j}}{\theta_{\pi_k}} \partial_{\pi_j, \pi_k} g^{\{\pi_1 \cdots, \pi_k\}}(\mathbf{0}) \tag{135}$$

Note that the summation in (135) accounts for the users not decoded yet according to the decoding order $\pi_K, \cdots, \pi_1$. Finally, by substituting (134) and (135) in (130) we get

$$
\begin{aligned}
\lim_{x \to 0} \dot{R}_k &= \sum_{\boldsymbol{\pi}} \lambda_{\boldsymbol{\pi}} \partial_{\pi_k} g^{\{\pi_1 \cdots, \pi_k\}}(\mathbf{0}) \\
&= \partial_k g^{\{\pi_1 \cdots, \pi_k\}}(\mathbf{0}) \quad\quad (136) \\
\lim_{x \to 0} \ddot{R}_k &= \sum_{\boldsymbol{\pi}} \lambda_{\boldsymbol{\pi}} \left(\partial_{\pi_k, \pi_k} g^{\{\pi_1 \cdots, \pi_k\}}(\mathbf{0}) + 2 \sum_{j=1}^{k-1} \frac{\theta_{\pi_j}}{\theta_{\pi_k}} \partial_{\pi_j, \pi_k} g^{\{\pi_1 \cdots, \pi_k\}}(\mathbf{0}) \right) \\
&= \partial_{k,k} g^{\{\pi_1 \cdots, \pi_k\}}(\mathbf{0}) + 2 \sum_{\boldsymbol{\pi}} \lambda_{\boldsymbol{\pi}} \sum_{j < \pi^{-1}(k)} \frac{\theta_{\pi_j}}{\theta_k} \partial_{\pi_j, k} g^{\{\pi_1 \cdots, \pi_k\}}(\mathbf{0}) \quad\quad (137)
\end{aligned}
$$

By recalling (127) and (128), and from expression (126), we get

$$
\mathcal{S}_k = \frac{\mathcal{S}_0^{(k)}}{1 + 2 \sum_{\boldsymbol{\pi}} \lambda_{\boldsymbol{\pi}} \sum_{j < \pi^{-1}(k)} \dfrac{\theta_{\pi_j}}{\theta_k} \dfrac{\sum_{n=1}^N \mathrm{E}[\alpha_{\pi_j, n} \mathbb{1}\{n_{\pi_j}^\star = n\}] \, \mathrm{E}[\alpha_{k,n} \mathbb{1}\{n_k^\star = n\}]}{\sum_{n=1}^N \mathrm{E}[\alpha_{k,n}^2 \mathbb{1}\{n_k^\star = n\}]}}
$$

$$(138)$$

The slope region obtained as union over all $\boldsymbol{\lambda}$ of (138) for all k is in general an inner bound to the optimal slope region. Similarly, the slope region obtained considering the outer bound (121) is in general an outer bound to the optimal slope region. Next we prove that those two bounds coincide, thus proving that policy $\boldsymbol{\beta}^\star$ in conjunction with superposition-coding is second-order optimal for any number of users K and any delay N.

In order to express a general point on the dominant face of the outer bound in (121) we follow the same steps that led to (138). In particular we need the gradient and Hessian matrix of $f^{(\mathcal{A})}$, computed in $\boldsymbol{\gamma} = \mathbf{0}$, for all subsets $\mathcal{A}$. The proof that the outer bound yields the same slope region of the inner bound is hence complete if we show that $\partial_{\ell,m} g^{(\mathcal{A})}(\mathbf{0}) = \partial_{\ell,m} f^{(\mathcal{A})}(\mathbf{0})$ for all $\ell, m \in \mathcal{A}$ and $\ell \neq m$ and for all subsets $\mathcal{A}$. In fact it is obvious that $\partial_\ell g^{(\mathcal{A})}(\mathbf{0}) = \partial_\ell f^{(\mathcal{A})}(\mathbf{0})$, otherwise the points on the outer-bound region would achieve higher minimum energy per bit than the points on the inner-bound region, and that $\partial_{\ell,\ell} g^{(\mathcal{A})}(\mathbf{0}) = \partial_{\ell,\ell} f^{(\mathcal{A})}(\mathbf{0})$, otherwise the numerator of the equivalent of (138) for the outer bound region would be different from the optimal ℓ-th single-user slope $\mathcal{S}_0^{(\ell)}$.

For every subset $\mathcal{A}$, for all $n = 1, \cdots, N$ let

$$
S_n\left(\{P_j\}_{j \in \mathcal{A}}; \mathcal{A}\right) = \mathrm{E}\left[\max_{\forall j \in \mathcal{A}: u_j \in [0, P_j]} \left\{ \log\left(1 + \sum_{j \in \mathcal{A}} \alpha_j \, u_j\right) + S_{n-1}\left(\{P_j - u_j\}_{j \in \mathcal{A}}; \mathcal{A}\right) \right\} \right] \quad (139)
$$

with initial condition $S_0\left(\mathbf{0};\mathcal{A}\right)=0$, then

$$f^{(\mathcal{A})}(\{\gamma_j\}_{j\in\mathcal{A}}) = \frac{1}{N}S_N\left(\{N\,\gamma_j\}_{j\in\mathcal{A}};\mathcal{A}\right) \tag{140}$$

Let $\boldsymbol{b}\in\{0,1\}^{|\mathcal{A}|}$, then a necessary condition for $\{u_j = P_j\,b_j\}_{j\in\mathcal{A}}$ to be solution of (139) is

$$\frac{\alpha_\ell}{1+\sum_{j\in\mathcal{A}}\alpha_j\,P_j\,b_j} - \partial_\ell S_{n-1}\left(\{P_j(1-b_j)\}_{j\in\mathcal{A}};\mathcal{A}\right) \begin{cases} < 0 & \text{if} \quad b_\ell = 0 \\ \geq 0 & \text{if} \quad b_\ell = 1 \end{cases} \tag{141}$$

Then it follows easily that in the limit for small $\{P_j\}_{j\in\mathcal{A}}$ we have $u_\ell = 0$ if $\alpha_\ell < \partial_\ell S_{n-1}\left(\mathbf{0};\mathcal{A}\right)$ and $u_\ell = P_\ell$ if $\alpha_\ell \geq \partial_\ell S_{n-1}\left(\mathbf{0};\mathcal{A}\right)$. Then we can write

$$\begin{aligned} S_n\left(\{P_j\}_{j\in\mathcal{A}};\mathcal{A}\right) &= \sum_{\boldsymbol{b}} \mathrm{E}\left[\left(\log\left(1+\sum_{j\in\mathcal{A}}\alpha_j\,P_j\,b_j\right) + S_{n-1}\left(\{P_j(1-b_j)\}_{j\in\mathcal{A}};\mathcal{A}\right)\right)\cdot \right. \\ &\qquad \left. \cdot\prod_{j\in\mathcal{A}}\mathbb{1}\{u_j = P_j\,b_j\}\right] + \text{vanishing terms with} \quad \{P_j\}_{j\in\mathcal{A}} \end{aligned} \tag{142}$$

Finally, in the limit for vanishing $\{P_j\}_{j\in\mathcal{A}}$ the second-order partial derivative of $S_n\left(\{P_j\}_{j\in\mathcal{A}};\mathcal{A}\right)$ w.r.t. P_ℓ and P_m is

$$\begin{aligned} \partial_{\ell,m}S_n\left(\mathbf{0};\mathcal{A}\right) &= \sum_{\boldsymbol{b}} \mathrm{E}\left[\left(-b_\ell b_m\alpha_\ell\alpha_m + (1-b_\ell)(1-b_m)\partial_{\ell,m}S_{n-1}(\mathbf{0};\mathcal{A})\right)\cdot \right. \\ &\qquad \left. \cdot\prod_{j\in\mathcal{A}:b_j=0}\mathbb{1}\{\alpha_j < \partial_\ell S_{n-1}\left(\mathbf{0};\mathcal{A}\right)\}\prod_{j\in\mathcal{A}:b_j=1}\mathbb{1}\{\alpha_j \geq \partial_\ell S_{n-1}\left(\mathbf{0};\mathcal{A}\right)\}\right] \\ &= \mathrm{E}\left[-\alpha_\ell\alpha_m\mathbb{1}\{\alpha_\ell \geq \partial_\ell S_{n-1}\left(\mathbf{0};\mathcal{A}\right)\}\mathbb{1}\{\alpha_m \geq \partial_m S_{n-1}\left(\mathbf{0};\mathcal{A}\right)\} \right. \\ &\qquad \left. +\partial_{\ell,m}S_{n-1}(\mathbf{0};\mathcal{A})\mathbb{1}\{\alpha_\ell < \partial_\ell S_{n-1}\left(\mathbf{0};\mathcal{A}\right)\}\mathbb{1}\{\alpha_m < \partial_m S_{n-1}\left(\mathbf{0};\mathcal{A}\right)\}\right] \\ &= \frac{1}{N}\partial_{\ell,m}f^{(\mathcal{A})}(\mathbf{0}) \end{aligned} \tag{143}$$

In order to prove that $N\partial_{\ell,m}S_N\left(\mathbf{0};\mathcal{A}\right)$ indeed coincides with (126) we must show that (143) is the recursion to compute (126). In fact, by recalling (118), we can write

$$\begin{aligned} \mu_N(\ell,m) &\triangleq \sum_{n=1}^{N}\mathrm{E}\left[\alpha_{\ell,n}\mathbb{1}\{n_\ell^\star = n\}\right]\mathrm{E}\left[\alpha_{m,n}\mathbb{1}\{n_m^\star = n\}\right] \tag{144} \\ &= \sum_{n=1}^{N}\prod_{j=1}^{n-1}F_\alpha^{(\ell)}(s_{N-j}^{(\ell)})\int_{s_{N-n}^{(\ell)}}^{\infty}x\mathrm{d}F_\alpha^{(\ell)}(x)\cdot\prod_{j=1}^{n-1}F_\alpha^{(m)}(s_{N-j}^{(m)})\int_{s_{N-n}^{(m)}}^{\infty}x\mathrm{d}F_\alpha^{(m)}(x) \end{aligned}$$

now, by separating the term for $n = 1$ in the summation, we can write

$$
\begin{aligned}
\mu_N(\ell, m) &= \int_{s_{N-1}^{(\ell)}}^{\infty} x \mathrm{d}F_\alpha^{(\ell)}(x) \cdot \int_{s_{N-1}^{(m)}}^{\infty} x \mathrm{d}F_\alpha^{(m)}(x) \\
&+ \sum_{n=2}^{N} \prod_{j=1}^{n-1} F_\alpha^{(\ell)}(s_{N-j}^{(\ell)}) \int_{s_{N-n}^{(\ell)}}^{\infty} x \mathrm{d}F_\alpha^{(\ell)}(x) \cdot \prod_{j=1}^{n-1} F_\alpha^{(m)}(s_{N-j}^{(m)}) \int_{s_{N-n}^{(m)}}^{\infty} x \mathrm{d}F_\alpha^{(m)}(x) \\
&= \mathrm{E}\left[\alpha_\ell \mathbf{1}\{\alpha_\ell \geq s_{N-1}^{(\ell)}\}\right] \cdot \mathrm{E}\left[\alpha_m \mathbf{1}\{\alpha_m \geq s_{N-1}^{(m)}\}\right] \\
&+ \mathrm{E}\left[\mathbf{1}\{\alpha_\ell < s_{N-1}^{(\ell)}\}\right] \cdot \mathrm{E}\left[\mathbf{1}\{\alpha_m < s_{N-1}^{(m)}\}\right] \cdot \mu_{N-1}(\ell, m) \quad (145)
\end{aligned}
$$

which, by recalling that $\partial_\ell S_n(\mathbf{0}; \mathcal{A}) = s_n^{(\ell)}$ for all $n = 1, 2, \cdots$ and all $\ell \in \{1, \cdots, K\}$, coincides with (143) for $n = N$. This concludes the proof that $\partial_{\ell,m} g^{(\mathcal{A})}(\mathbf{0}) = \partial_{\ell,m} f^{(\mathcal{A})}(\mathbf{0})$ for all $\ell, m \in \mathcal{A}$ and for all subsets $\mathcal{A}$, thus proving that the optimal slope region, parameterized by $\boldsymbol{\theta}$ can be written as in (38)

$$
\bigcup_{\boldsymbol{\lambda}} \left\{ \mathbb{S}_k \geq 0 \quad \forall k = 1, \cdots, K \right. \quad (146)
$$

$$
\left. \mathbb{S}_k \leq \frac{\mathbb{S}_0^{(k)}}{1 + 2 \sum_{\boldsymbol{\pi}} \lambda_{\boldsymbol{\pi}} \sum_{j < \pi^{-1}(k)} \frac{\theta_{\pi_j}}{\theta_k} \frac{\sum_{n=1}^{N} \mathrm{E}[\alpha_{\pi_j,n} \mathbf{1}\{n_{\pi_j}^\star = n\}] \mathrm{E}[\alpha_{k,n} \mathbf{1}\{n_k^\star = n\}]}{\sum_{n=1}^{N} \mathrm{E}[\alpha_{k,n}^2 \mathbf{1}\{n_k^\star = n\}]}} \right\}
$$

and that the one-shot policy $\boldsymbol{\beta}^\star$ is second-order optimal.

Remark. In analogy with (145), it can be shown that, for all $\mathcal{A} \subseteq \{1, \ldots, K\}$,

$$
g^{(\mathcal{A})}(\{\gamma_j\}_{j \in \mathcal{A}}) = \frac{1}{N} S_N^\star(\{N \gamma_j\}_{j \in \mathcal{A}}; \mathcal{A})
$$

given by the recursion

$$
\begin{aligned}
S_n^\star(\{P_j\}_{j \in \mathcal{A}}; \mathcal{A}) &= \mathrm{E}\left[\log\left(1 + \sum_{j \in \mathcal{A}} \alpha_j P_j\right) \prod_{j \in \mathcal{A}} \mathbf{1}\{\alpha_j \geq s_{n-1}^{(j)}\}\right] \\
&+ \prod_{j \in \mathcal{A}} \Pr[\alpha_j < s_{n-1}^{(j)}] S_{n-1}^\star(\{P_j\}_{j \in \mathcal{A}}; \mathcal{A})
\end{aligned}
$$

with initial condition $S_0^\star(\{P_j\}_{j \in \mathcal{A}}; \mathcal{A}) = 0$.

H Proof of Theorem 13

By repeating the same steps that led to (147) in Appendix G, it follows easily that the term $\mathcal{K}_{k,j}$ is given by

$$\mathcal{K}_{k,j} = 2\frac{\partial_{j,k}g(\mathbf{0})}{\partial_{k,k}g(\mathbf{0})} \tag{147}$$

where the function $g(\boldsymbol{\gamma})$ is defined as

$$g(\boldsymbol{\gamma}) = \mathrm{E}\left[\frac{1}{N}\sum_{n=1}^{N}\log\left(1 + \sum_{k=1}^{K}\alpha_{k,n}N\gamma_k\,1\{n = n_k^\star\}\right)\right] \tag{148}$$

and where $n_k^\star$ is the index of the maximum, i.e.,

$$n_k^\star = \arg\max_n\{\alpha_{k,1},\cdots,\alpha_{k,n},\cdots,\alpha_{k,N}\}$$

The partial derivatives are given by

$$\partial_k g(\boldsymbol{\gamma}) = \mathrm{E}\left[\sum_{n=1}^{N}\frac{\alpha_{k,n}1\{n = n_k^\star\}}{1 + \sum_{k=1}^{K}\alpha_{k,n}N\gamma_k\,1\{n = n_k^\star\}}\right] \tag{149}$$

by

$$
\begin{aligned}
\partial_{k,k}g(\mathbf{0}) &= -N\,\mathrm{E}\left[\sum_{n=1}^{N}\left(\frac{\alpha_{k,n}1\{n = n_k^\star\}}{1 + \sum_k \alpha_{k,n}N\gamma_k\,1\{n = n_k^\star\}}\right)^2\right]\Bigg|_{\boldsymbol{\gamma}=\mathbf{0}} \\
&= -N\,\mathrm{E}\left[(\max_n\{\alpha_{k,n}\})^2\right] \tag{150}
\end{aligned}
$$

and by

$$
\begin{aligned}
\partial_{j,k}g(\mathbf{0}) &= -N\,\mathrm{E}\left[\sum_{n=1}^{N}\frac{\alpha_{k,n}1\{n = n_k^\star\}\alpha_{j,n}1\{n = n_j^\star\}}{(1 + \sum_k \alpha_{k,n}N\gamma_k\,1\{n = n_k^\star\})^2}\right]\Bigg|_{\boldsymbol{\gamma}=\mathbf{0}} \\
&= -N\,\mathrm{E}\left[\max_n\{\alpha_{k,n}\}\max_n\{\alpha_{j,n}\}1\{n_k^\star = n_j^\star\}\right] \\
&= -\mathrm{E}\left[\max_n\{\alpha_{k,n}\}\max_n\{\alpha_{j,n}\}\right] \tag{151}
\end{aligned}
$$

where the last equality follows because the events $\{\max_n\{\alpha_{k,n}\} \leq x\}$ and $\{n_k^\star = i\}$ are independent (notice the same user index k). Then, since $\{n_k^\star = i\}$ is uniformly distributed on $\{1,\cdots,N\}$, and since the events $\{n_k^\star = i\}$ and $\{n_j^\star = i\}$ are independent (notice the different user indexes), it follows that $\Pr(1\{n_k^\star = n_j^\star\}) = \frac{1}{N}$. Finally, by substituting (150) and (151) in (147) we obtain (45).

References

[1] A.Goldsmith and P.P.Varaiya. Capacity of fading channels with channel state information. *IEEE Trans. on Inform. Theory*, 43(6):1986–1992, November 1997.

[2] R. S. Cheng and S. Verdú. Gaussian multiaccess channels with ISI: capacity region and multiuser water-filling. *IEEE Trans. on Inform. Theory*, 39(3):773 –785, May 1993.

[3] I. Csiszar and J. Korner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1982.

[4] D.P.Bertsekas. *Dynamic programming and optimal control, Vol.1 and Vol.2*. Athena Scientific, 1995.

[5] D.Tse and S.Hanly. Multiaccess fading channels-Part I: Polymatroid structure, optimal resource allocation and throughput capacities. *IEEE Trans. on Inform. Theory*, 44(7):2796–2815, November 1998.

[6] E.Biglieri, J.Proakis, and S.Shamai. Fading channels: information-theoretic and communication aspects. *IEEE Trans. on Inform. Theory*, 44(6):2619–2692, October 1998.

[7] R. Gallager. *Communications and Cryptography – Two sides of one tapestry*, chapter An inequality on the capacity region of multiaccess fading channels, pages 129–139. Kluwer, Boston, MA, 1994.

[8] G.Caire, G.Taricco, and E.Biglieri. Optimum power control over fading channel. *IEEE Trans. on Inform. Theory*, 45(5):1468–1489, July 1999.

[9] A. J. Goldsmith and M. Effros. The capacity region of broadcast channels with intersymbol interference and colored gaussian noise. *IEEE Trans. on Inform. Theory*, 47(1):219 –240, Jannuary 2001.

[10] J.G.Proakis. *Digital Communications (3nd ed.)*. McGraw Hill, New York, 1995.

[11] Lifang Li and A.Goldsmith. Outage capacities and optimal power allocation for fading multiple-access channels. In *IEEE Wireless Communications and Networking Conference, 1999 (WCNC1999)*, volume 1, pages 485–489, Seattle, 1999.

[12] Lifang Li and Goldsmith A.J. Capacity and optimal resource allocation for fading broadcast channels. part i. ergodic capacity. part ii. outage capacity. *IEEE Trans. on Inform. Theory*, 47(3):1083–1127, March 2001.

[13] L.Ozarow, S.Shamai, and A.D.Wyner. Information theoretic considerations for cellular mobile radio. *IEEE Trans. on Vehic. Tech.*, 43(2):359–378, May 1994.

[14] R.Knopp and P.A.Humblet. Information capacity and power control in single-cell multiuser communications. In *IEEE International Conference on Communications, 1995 (ICC '95), 'Gateway to Globalization'*, volume 1, pages 331–335, Seattle, July 1995.

[15] R.Negi and J.Cioffi. Delay-constrained capacity with causal feedback. *IEEE Trans. on Inform. Theory*, 48(9):2478 –2494, September 2002.

[16] S.Hanly and D.Tse. Multiaccess fading channels-Part II: Delay limited capacities. *IEEE Trans. on Inform. Theory*, 44(7):2816–2831, November 1998.

[17] S.Verdú. On channel capacity per unit cost. *IEEE Trans. on Inform. Theory*, 36(5):1019–1030, September 1990.

[18] S.Verdú. Spectral efficiency in the wideband regime. *IEEE Trans. on Inform. Theory*, 48(6):1319 –1343, June 2002.

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

[20] S. Verdú, G. Caire, and D. Tuninetti. Is TDMA optimal in the low power regime? In *Proceedings 2002 IEEE International Symposium on Information Theory*, Lausanne (CH), 2002.