Hard Fairness Versus Proportional Fairness in Wireless Communications: The Single-Cell Case

Giuseppe Caire, *Fellow, IEEE*, Ralf R. Müller, *Senior Member, IEEE*, and Raymond Knopp

***Abstract*—We consider a wireless communication system formed by a single cell with one base station and K user terminals. User channels are characterized by frequency-selective fading due to small-scale effects, modeled as a set of M parallel block-fading channels, and a frequency-flat distance-dependent path loss. We compare delay-limited systems with variable-rate systems under fairness constraints, in terms of the achieved system spectral efficiency C (bit/s/Hz) versus E_b/N_0. The considered delay-limited systems impose "hard-fairness": every user transmits at its desired rate on all blocks, independently of its fading conditions. The variable-rate system imposes "proportional fairness" via the popular Proportional Fair Scheduling (PFS) algorithm, currently implemented in 3G wireless for data (delay-tolerant) applications. We find simple iterative resource allocation algorithms that converge to the optimal delay-limited throughput for orthogonal (frequency-division multiple access (FDMA)/time-division multiple access (TDMA)) and optimal (superposition/interference cancellation) signaling. In the limit of large K and finite M we find closed-form expressions for C as a function of E_b/N_0. We show that in this limit, the optimal allocation policy consists of letting each user transmit on its best subchannel only. Also, we find a simple closed-form expression for the throughput of PFS in a cellular environment, that holds for any K and M. Finally, we obtain closed-form expressions for C versus E_b/N_0 in the low and high spectral efficiency regimes.**

The conclusions of our analysis in terms of system design guidelines are as follows: a) if hard fairness is a requirement, orthogonal access incurs a large throughput penalty with respect to the optimal (superposition coding) strategy, especially in the regime of high spectral efficiency; b) for high spectral efficiency, PFS does not provide any significant gain and may even perform worse than the optimal delay-limited system, despite the fact that the imposed fairness constraint is laxer; c) for low to moderate spectral efficiency, the stricter hard-fariness constraint incurs in a large throughput penalty with respect to PFS.

***Index Terms*—Code-division multiple access (CDMA), delay-limited capacity, proportional fair scheduling, uplink–downlink duality.**

Manuscript received February 10, 2006; revised October 5, 2006. This work was supported in part by the European Commission via the NEWCOM Network of Excellence under Contract IST NoE 507325 as well as the K-Plus programme of the Austrian Government.

G. Caire is with the Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089 USA (e-mail: caire@usc.edu).

R. R. Müller was with Forschungszentrum Telekommunikation Wien (FTW, Austria) and the Centre National de Recherche Scientifique (CNRS, France). He is now with Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway (e-mail: ralf@iet.ntnu.no).

R. Knopp is with the Institute Eurècom, 06904 Sophia-Antipolis, France (e-mail: raymond.knopp@eurecom.fr).

Communicated by P. Viswanath, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2007.892790

I. Introduction

We consider the uplink and the downlink of a wireless communication system with one base station and K user terminals. Each user is affected by a position-dependent path loss, fixed in time, and by a slowly time-varying frequency-selective fading channel modeled as M parallel block-fading channels [1].

We study the system throughput (total spectral efficiency) versus E_b/N_0 under hard fairness and proportional fairness constraints. By "hard fairness" we mean a system where each user transmits at its own desired rate, determined independently of the actual fading channel realization, and the system struggles to accommodate each user's rate request. This corresponds to the so-called *delay-limited* capacity of fading multiple-access channels [2]. When such strict rate constraint is relaxed, the notion of *throughput* (or ergodic) capacity region [3] becomes relevant: in our context, this is the long-term average rate region achievable when the users adapt their rate and power according to the actual channel conditions.[1] It is well known that the maximum *long-term average* throughput is achieved by letting only the user with the best channel transmit on each time–frequency coding interval (referred to as "slot" in the following) [3], [4]. However, in a cellular environment, where users are at different distance from the base station, this strategy would result in a very unfair resource allocation where basically only the users closest to the base station are allowed to transmit, while the users far from the base station would starve. To cope with this "near–far effect," various scheduling algorithms aiming at maximizing the long-term average throughput subject to some fairness constraint have been proposed. Among these, the Proportional Fair Scheduling (PFS) algorithm [5] enjoys many desirable properties and was adopted in some evolutionary 3G wireless communication standards [6], [7] for delay-tolerant data-oriented communications. By "proportional fairness" we mean a system that schedules the users according to the PFS policy.

Our analysis allows us to quantify the effect of imposing hard fairness versus proportional fairness in a cellular environment, for given M, K and channel statistics. For finite K, we find simple iterative resource allocation algorithms that provably converge to the optimal delay-limited throughput. Also, in the limit of very large K and finite M we find closed-form expressions for the delay-limited throughput. We show that, for both optimal and orthogonal signaling, the optimal strategy in

[1]When coding over an arbitrarily large number of fading blocks is allowed, the same ergodic capacity region can be achieved by fixed-rate variable-power transmission. However, due to our assumption of block-fading channel, in this work we assume that a coding interval spans a single fading realization. Hence, the variable rate and power scheme is in place.

the limit of large K consists of letting the users transmit on their own best subchannel only, irrespective of the other users. This result suggests a system where the users are able to "listen wideband," i.e., measure their channel gain on all the M subchannels, and "talk narrowband," i.e., they will transmit only on their best subchannel.

In the case of PFS, we find a simple closed-form expression for the throughput in the considered cellular environment that holds for any K and M.

Finally, we carry out a closed-form analysis of the throughput versus system E_b/N_0 in the high and low spectral efficiency regions, for all systems under consideration. Our analysis shows that, in the high spectral efficiency (high signal-to-noise ratio (high-SNR)) region, the penalty incurred by imposing hard fairness is generally small. Furthermore, in some cases of practical interest (with reasonably large but finite K), the optimal delay-limited system may outperform PFS for high spectral efficiency. On the contrary, the gain of PFS over *any* delay-limited system can be significant in the low spectral efficiency region (low SNR).

The proofs are mainly collected in the Appendices, in order to keep the flow of exposition.

II. Background

A. Capacity Region, Power Region, and $(E_b/N_0)_{\text{sys}}$

The K-user Gaussian multiple-access channel

$$Y = \sum_{k=1}^{K} X_k + N \tag{1}$$

has capacity region given by [8]

$$\sum_{k\in\mathcal{S}} R_k \le \log\left(1 + \frac{\sum_{k\in\mathcal{S}} E_k^{(r)}}{N_0}\right), \qquad \forall \mathcal{S} \subseteq \{1,\dots,K\} \tag{2}$$

with $E_k^{(r)}$ denoting the received energy per symbol of user k [9] and $N_0 = \mathbb{E}[|N|^2]$ denoting the noise power spectral density.

If orthogonal signaling, e.g., by time-division multiple access (TDMA) or frequency-division multiple access (FDMA), is used, the achievable rate region is given by

$$R_k \le \Theta_k \log\left(1 + \frac{E_k^{(r)}}{\Theta_k N_0}\right) \quad \forall k \tag{3}$$

subject to

$$\sum_{k=1}^{K} \Theta_k \le 1 \tag{4}$$

where Θ_k denotes the resource-sharing fraction (proportion of channel dimensions) given to user k. If these fractions are chosen appropriately, the optimal sum rate can be achieved. See Fig. 1 for an illustration in the case of $K = 2$ users.

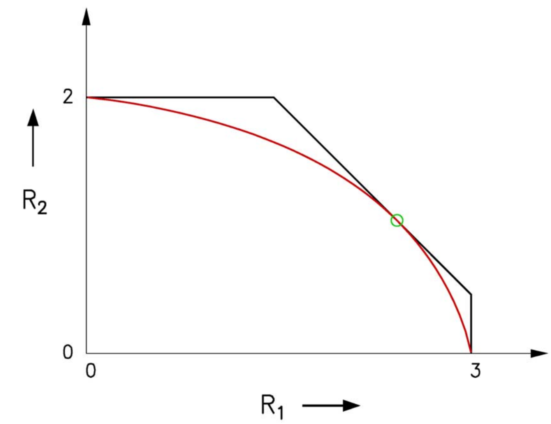

Fig. 1. Capacity region for two users with $E_1^{(r)}/N_0 = 7$ and $E_2^{(r)}/N_0 = 3$ [10]. The "lower" curve refers to orthogonal signaling.

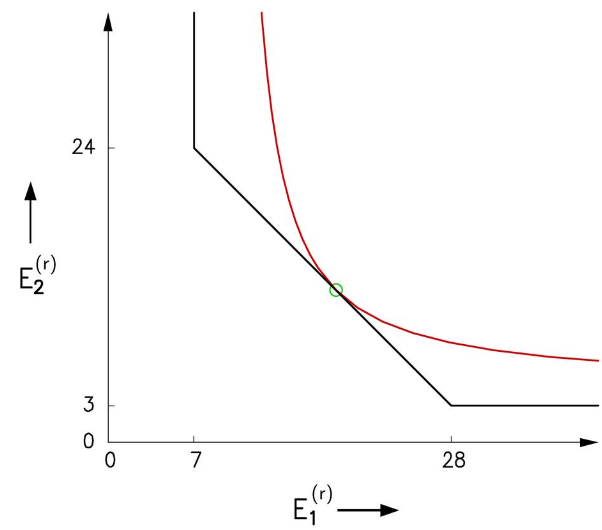

Fig. 2. Received power region for two users with $R_1 = 3, R_2 = 2$ bits, and $N_0 = 1$ [10]. The "upper" curve refers to orthogonal signaling.

The received power region supporting a given set of user rates is obtained solving the $2^K - 1$ equations in (2) and the K equations in (3) for the symbol energies. This yields

$$\sum_{k\in\mathcal{S}} E_k^{(r)} \ge N_0 \left[\exp\left(\sum_{k\in\mathcal{S}} R_k\right) - 1\right], \quad \forall\, \mathcal{S} \subseteq \{1,\dots,K\} \tag{5}$$

for optimal signaling and

$$E_k^{(r)} \ge \Theta_k N_0 [\exp(R_k/\Theta_k) - 1], \quad \forall k, \tag{6}$$

for orthogonal signaling, respectively, as shown in Fig. 2. Again, constraining to orthogonal signaling does not increase the required sum power if the fractions Θ_k are chosen appropriately.

Next, we introduce the propagation channel gains by formulating the problem in terms of *transmit* powers: namely, the received symbol $E_k^{(r)}$ energy is related to the transmit symbol energy E_k by

$$E_k^{(r)} = d_k E_k \tag{7}$$

where d_k denotes the channel (power) gain of user k. This leads to a rescaling of the axis in Fig. 2 as illustrated in Fig. 3. The rescaling of the axis implies that the minimum received power for given user rates is achieved by a unique vertex of the power

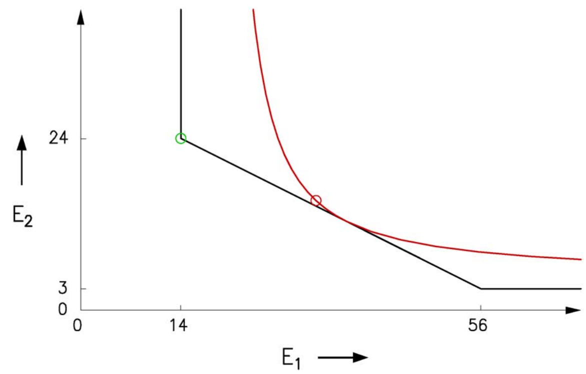

Fig. 3. Transmit power region for two users with $R_1 = 3, R_2 = 2$ bits, $N_0 = 1, d_1 = \frac{1}{2}$, and $d_2 = 1$ [10]. The "upper" curve refers to orthogonal signaling.

region (assuming all user gains distinct). At this vertex, the receiver can make use of successive decoding (stripping) without loss of performance. Clearly, users are decoded in decreasing order of the strengths of their channels [10], [2].

Constraining to orthogonal signaling implies an increase in total transmit power unless all channel gains are identical. In addition, the optimal choice of the fractions is influenced by the channel gains. In order to help users with bad channels, their fractions are increased at the expense of users with good channels.

The minimum total energy supporting a given rate K-tuple $\boldsymbol{R} = (R_1, \dots, R_K)$ with gains $\boldsymbol{d} = (d_1, \dots, d_K)$ is obtained by finding the symbol energies solution of

$$\min_{\boldsymbol{E} \in \mathbb{R}_+^K} \sum_{k=1}^{K} E_k \quad \text{subject to} \quad \boldsymbol{R} \in \mathcal{C}_{\rm MAC}(\boldsymbol{d}; \boldsymbol{E}), \tag{8}$$

where $\mathcal{C}_{\rm MAC}(\boldsymbol{d}; \boldsymbol{E})$ is the region defined in (2) after letting $E_k^{(r)} = d_k E_k$. Thanks to the fact that the received energy region is a contra-polymatroid [2], the solution of (8) is found explicitly as

$$E_{\pi_k} = \frac{N_0}{d_{\pi_k}} \left[\exp\left(\sum_{i \le k} R_{\pi_i} \right) - \exp\left(\sum_{i < k} R_{\pi_i} \right) \right] \tag{9}$$

where π is the permutation of $\{1, \dots, K\}$ that sorts the channel gains in increasing order, i.e.,

$$d_{\pi_1} \le \cdots \le d_{\pi_K}$$

where we define the associated "back-substitution" decoding order given by π_K (decoded first), $\pi_{K-1}, \dots, \pi_1$ (decoded last).

With orthogonal signaling, the minimum total energy supporting a given rate K-tuple $\boldsymbol{R}$ with gains $\boldsymbol{d}$ is obtained by solving

$$\min_{\boldsymbol{\Theta}, \boldsymbol{E} \in \mathbb{R}_+^K} \sum_{k=1}^{K} E_k \quad \text{subject to} \quad \boldsymbol{R} \in \mathcal{C}_{\rm orth}(\boldsymbol{d}; \boldsymbol{E}; \boldsymbol{\Theta}) \tag{10}$$

where $\mathcal{C}_{\rm orth}(\boldsymbol{d}; \boldsymbol{E}; \boldsymbol{\Theta})$ is the region defined by (3) and (4), after letting $E_k^{(r)} = d_k E_k$.

Assume that the channel gain vector is constant over the duration of a codeword and it is randomly distributed according to some joint probability law. The delay-limited capacity region [2] is the set of all rate K-tuples $\boldsymbol{R}$ that can be attained for all $\boldsymbol{d} \in \mathbb{R}_+^K$, subject to *average power constraints* $\mathbb{E}[E_k] \le \overline{E}_k$. In particular, in this work we are interested in the total system throughput (sum rate) versus the total average transmit energy. Following [11]–[13], we define the system E_b/N_0 under a coding strategy that supports user rates $\boldsymbol{R} = (R_1, \dots, R_K)$ with sum $\Gamma = \sum_{k=1}^K R_k$ subject to average transmit energy per symbol constraints $(\overline{E}_1, \dots, \overline{E}_K)$ as

$$\left(\frac{E_b}{N_0} \right)_{\rm sys} = \frac{\sum_{k=1}^K \bar{E}_k}{N_0 \Gamma} \tag{11}$$

where Γ is expressed in bits.[2] In the case of equal individual rates $R_k = \Gamma/K$, $(E_b/N_0)_{\rm sys}$ coincides with the individual user transmit E_b/N_0. In general, for finite K, an operating point $((E_b/N_0)_{\rm sys}, \Gamma)$ on the power/spectral efficiency plane is a function of both the signaling strategy and of the individual user rates $\boldsymbol{R}$, as well as of the channel gain joint distribution. In the following sections, we investigate a regime of large K for which, under mild assumptions on the user individual rates, the dependency on $\boldsymbol{R}$ disappears and the total instantaneous transmit energy converges to its average value.

B. Downlink

So far, we have treated the multiple-access case, modeling the uplink of a wireless system with a single base station and many users. Exploiting the recent result on the duality of the Gaussian multiple-access and broadcast channels [14], it is immediate to see that for any set of user rates $\boldsymbol{R}$, the required $(E_b/N_0)_{\rm sys}$ is the same for uplink and downlink.

For an orthogonal system this follows trivially from the fact that the uplink and downlink channel gains are identically distributed. In the case of optimal signaling, letting $\mathcal{C}_{\rm BC}(\boldsymbol{d}; E_{\rm tot})$ denote the Gaussian broadcast channel capacity region with gains $\boldsymbol{d}$ and transmit energy per symbol $E_{\rm tot}$, [14] showed that

$$\mathcal{C}_{\rm BC}(\boldsymbol{d}; E_{\rm tot}) = \bigcup_{\sum_{k=1}^K E_k = E_{\rm tot}} \mathcal{C}_{\rm MAC}(\boldsymbol{d}; \boldsymbol{E}). \tag{12}$$

Any point on the boundary of $\mathcal{C}_{\rm BC}(\boldsymbol{d}; E_{\rm tot})$ corresponds to the vertex of $\mathcal{C}_{\rm MAC}(\boldsymbol{d}; \boldsymbol{E})$ (for some choice of the individual transmit energies $(E_1, \dots, E_K)$ such that $\sum_k E_k = E_{\rm tot}$) associated to successive decoding in the order $\pi_K, \pi_{K-1}, \dots, \pi_1$. For what has been said before, this vertex is precisely the one that minimizes the total multiple-access channel transmit energy for given user rates $\boldsymbol{R}$ and channel gains $\boldsymbol{d}$. Therefore, for any realization of $\boldsymbol{d}$, the downlink achieves the rate K-tuple $\boldsymbol{R}$ with the same (minimal) total transmit energy of the uplink. It follows that uplink and downlink achieve the same set of optimal operating points $((E_b/N_0)_{\rm sys}, \Gamma)$. Therefore, from now on, we shall focus on the uplink, taking into account that all the results and conclusions are immediately applicable to the downlink.

Notice that also the coding/decoding schemes for uplink and downlink are very similar. In the uplink, each user k sends a Gaussian codeword of rate R_k and energy per symbol E_k.

[2]Throughout this paper, information rates are generally expressed in nats of analytical convenience. However, we use the obvious convention that, by definition, in all expressions that yield E_b/N_0 the information rates (in particular, the sum-rate Γ) are expressed in bits.

The receiver gets the superposition of all codewords plus noise and makes use of stripping decoding in the order of decreasing channel strength. In the downlink the transmitter sends the superposition of K independently selected Gaussian codewords such that codeword k has rate R_k and energy per symbol $\tilde{E}_k$. Each receiver makes use of stripping decoding in order to cancel interference of all users with smaller channel gain, while treating the users with larger channel gain as background noise. For the sake of completeness, we mention that the downlink signal energies are determined recursively from the uplink signal energies according to [14]

$$\begin{aligned}\tilde{E}_{\pi_K} &= \frac{N_0 E_{\pi_K}}{N_0 + \sum_{k=1}^{K-1} d_{\pi_k} E_{\pi_k}}\\ \tilde{E}_{\pi_{K-1}} &= \frac{\left(N_0 + d_{\pi_{K-1}} \tilde{E}_{\pi_K}\right) E_{\pi_{K-1}}}{N_0 + \sum_{k=1}^{K-2} d_{\pi_k} E_{\pi_k}}\\ &\vdots\\ \tilde{E}_{\pi_1} &= \frac{\left(N_0 + d_{\pi_1} \sum_{k=2}^{K} \tilde{E}_{\pi_k}\right) E_{\pi_1}}{N_0}. \end{aligned} \tag{13}$$

In [14] it is proved that $E_{\text{tot}} = \sum_{k=1}^{K} \tilde{E}_k = \sum_{k=1}^{K} E_k$.

III. Parallel Channels

We consider M parallel block-fading channels, where the channel gain of each user may differ from channel to channel. Namely, we let

$$Y^m = \sum_{k=1}^{K} \sqrt{d_k^m} X_k^m + N^m, \quad m = 1, \ldots, M. \tag{14}$$

This is an accurate model for frequency selectivity where m can be interpreted as the subband index. The system *spectral efficiency* is given by

$$\mathsf{C} = \frac{\Gamma}{M} \tag{15}$$

and it is expressed in bits per second per hertz (bit/s/Hz) or, equivalently, in bits per dimension.

The theoretical foundations of Gaussian parallel multiple-access channels were laid down in [15]. The capacity region of this channel can be achieved by letting each user split its information messages into M parallel streams, encode them independently, and send the resulting independent codewords over the parallel channels. The aggregate rate and aggregate energy per symbol of user k are given by

$$R_k = \sum_{m=1}^{M} R_k^m, \quad k = 1, \ldots, K \tag{16}$$

$$E_k = \sum_{m=1}^{M} E_k^m, \quad k = 1, \ldots, K \tag{17}$$

respectively, where R_k^m and E_k^m denote the rate and the energy per symbol allocated by user k on subchannel m. Letting $\boldsymbol{E}^m = (E_1^m, \ldots, E_K^m), \boldsymbol{R}^m = (R_1^m, \ldots, R_K^m)$, and $\boldsymbol{d}^m = (d_1^m, \ldots, d_K^m)$, the capacity region for given per-user energies $\boldsymbol{E} = (E_1, \ldots, E_K)$ and channel gains can be written as

$$\begin{aligned}&\mathcal{C}_{\text{MAC}}(\boldsymbol{d}^1, \ldots, \boldsymbol{d}^M; \boldsymbol{E})\\ &= \bigcup_{\substack{\sum_m E_k^m \le E_k\\ k=1,\ldots,K}} \left\{ \boldsymbol{R} = \sum_m \boldsymbol{R}^m : \boldsymbol{R}^m \in \mathcal{C}_{\text{MAC}}(\boldsymbol{d}^m; \boldsymbol{E}^m) \right\}. \end{aligned} \tag{18}$$

In other words, the partial rates R_k^m and energies E_k^m must obey the constraints (2) and (5) in each subchannel m.

For orthogonal multiple access, we let $\boldsymbol{\Theta}^m = (\Theta_1^m, \ldots, \Theta_K^m)$, where Θ_k^m denotes the resource-sharing fraction of user k over channel m. The achievable rate region under orthogonal signaling can be written as

$$\begin{aligned}\mathcal{C}_{\text{orth}}(\boldsymbol{d}^1, \ldots, \boldsymbol{d}^M; \boldsymbol{E}) = &\bigcup_{\substack{\sum_k \Theta_k^m \le 1\\ m=1,\ldots,M}} \bigcup_{\substack{\sum_m E_k^m \le E_k\\ k=1,\ldots,K}}\\ &\left\{ \boldsymbol{R} = \sum_m \boldsymbol{R}^m : \boldsymbol{R}^m \in \mathcal{C}_{\text{orth}}(\boldsymbol{d}^m; \boldsymbol{E}^m; \boldsymbol{\Theta}^m) \right\}. \end{aligned} \tag{19}$$

A. Delay-Limited Systems

As said earlier, in a delay-limited situation, the rates R_k are fixed *a priori*, and the system has to allocate transmit energies in order to let the rate K-tuple inside the achievable rate region. We wish to find the partial rates allocation (and the resource-sharing fractions in the case of orthogonal signaling) in order to minimize the required $(E_b/N_0)_{\text{sys}}$ to maintain a given rate K-tuple.

For optimal signaling, we make use of the fact that the successive decoding order depends only on the channel gains, but not on the rates. This has the important consequences that: 1) the decoding order is independent of the split of rates into partial rates; 2) the decoding order differs, in general, from subchannel to subchannel. Let π^m denote the permutation that sorts the gains $\boldsymbol{d}^m$ in increasing order. The required transmit energy per symbol of user π_k^m in channel m is given by

$$E_{\pi_k^m}^m = \frac{N_0}{d_{\pi_k^m}^m} \left[\exp\left(\sum_{i \le k} R_{\pi_i^m}^m \right) - \exp\left(\sum_{i < k} R_{\pi_i^m}^m \right) \right]. \tag{20}$$

Optimizing the partial rates R_k^m in order to minimize

$$\left(\frac{E_b}{N_0}\right)_{\text{sys}} = \frac{1}{\Gamma N_0} \sum_{k=1}^{K} \sum_{m=1}^{M} E_k^m \tag{21}$$

(recall that $\Gamma = \sum_k R_k$ is fixed by the user rates), subject to the constraints (16), is a convex optimization problem that can be solved with standard tools.

In particular, the constraint (16) is separable and the objective function (21) is convex. For such problems, it is generally sufficient to optimize with respect to $\{R_1^m : m = 1, \ldots, M\}$ while holding all other variables constant, then optimize with respect to $\{R_2^m : m = 1, \ldots, M\}$, etc., in an iterative fashion, in order to converge to a globally optimum point. This is referred to as the block-coordinate descent algorithm and convergence can be shown under relatively general conditions [16, Sec. 2.7].

Focusing on the optimization step with respect to $\{R_k^m : m = 1, \ldots, M\}$, we notice that when all the other rate variables are fixed, the objective function can be written in the form

$$\sum_{m=1}^{M} a_k^m \exp(R_k^m)$$

subject to $\sum_{m=1}^M R_k^m \ge R_k, R_k^m \ge 0$, where the coefficients a_k^m depend on the other system variables and can be easily evaluated. Using the result of Appendix B, we find the solution in the form

$$R_k^m = \left[\log \frac{\lambda_k}{a_k^m}\right]_+$$

were λ_k is the solution of $\sum_{m=1}^M [\log(\lambda_k / a_k^m)]_+ = R_k$.

For orthogonal signaling, the transmit energy per symbol of user k in subchannel m is given by

$$E_k^m = \Theta_k^m \frac{N_0}{d_k^m} \left(\exp\left(R_k^m / \Theta_k^m\right) - 1\right). \tag{22}$$

The minimization of (21) with respect to $\{\boldsymbol{\Theta}^m, \boldsymbol{R}^m : m = 1, \ldots, M\}$ is also a convex optimization problem. In fact, it can be checked that the function $g(x, y) = y \exp(x/y)$ is convex on $\mathbb{R}_+^2$. Since the constraints

$$\sum_{m=1}^{M} R_k^m \ge R_k, \quad k = 1, \ldots, K$$

and

$$\sum_{k=1}^{K} \Theta_k^m \le 1, \quad m = 1, \ldots, M$$

with $R_k^m \ge 0$ and $\Theta_k^m > 0$ are also separable, we can use again the coordinate descent algorithm and optimize in sequence, and iteratively, with respect to the variables R_k^m while keeping Θ_k^m fixed, and then with respect to the Θ_k^m's while keeping R_k^m fixed. From Appendix B, we find that the optimization with respect to $\{R_k^m : m = 1, \ldots, M\}$ yields

$$R_k^m = \Theta_k^m \left[\log(\lambda_k d_k^m)\right]_+$$

were λ_k is the solution of $\sum_{m=1}^M \Theta_k^m [\log(\lambda_k d_k^m)]_+ = R_k$.

The optimization with respect to $\{\Theta_k^m : k = 1, \ldots, K\}$ is a bit more complicated. The associated Lagrangian function is given by

$$\mathcal{L}(\boldsymbol{\Theta}^m) = \sum_{k=1}^{K} \frac{\Theta_k^m}{d_k^m} \exp(R_k^m / \Theta_k^m) + \mu^m \left(\sum_{k=1}^{K} \Theta_k^m - 1\right) \tag{23}$$

where μ^m is the Lagrange multiplier. First, notice that any point with $\Theta_k^m = 0$ cannot be a solution. Therefore, the Kuhn–Tucker conditions must hold with equality and yield

$$\left(1 - \frac{R_k^m}{\Theta_k^m}\right) \exp(R_k^m / \Theta_k^m) = 1 - \mu^m d_k^m.$$

Solving for Θ_k^m, we obtain

$$\Theta_k^m = \frac{R_k^m}{1 + W\left(\frac{d_k^m \mu^m - 1}{\mathrm{e}}\right)} \tag{24}$$

where $W : [-1/\mathrm{e}, +\infty) \to [-1, +\infty)$ is Lambert's W function [17], defined by the equation $W \exp(W) = x$. The value of the Lagrange multiplier $\mu^m \in \mathbb{R}_+$ can be found by solving (numerically) the equation

$$\sum_{k=1}^{K} \frac{R_k^m}{1 + W\left(\frac{d_k^m \mu^m - 1}{\mathrm{e}}\right)} = 1. \tag{25}$$

B. Delay-Tolerant Systems

In a delay-tolerant situation, the user rates can be adapted according to their instantaneous channel conditions. For simplicity, we consider the case of constant total power transmission (that is more relevant for the downlink, where the base station can operate always at its peak total power) and let $\mathsf{SNR} = E_{\text{tot}}/N_0$ denote the transmit SNR in each slot. A similar result is obtained if water-filling power allocation is used. Moreover, for high SNR, the difference between water filling and constant total power is negligible. Also, we assume that the channel gains are independent but not necessarily identically distributed across the users, and symmetrically distributed across the channels, that is, for any permutation π of $\{1, \ldots, M\}$, the joint cumulative distribution function (cdf) of the channel gains satisfies $F(d_k^1, \ldots, d_k^M) = F(d_k^{\pi_1}, \ldots, d_k^{\pi_M})$, for all k. This means that no subchannel is statistically worse or better than any other. However, the users might have different channel gain distributions, i.e., our analysis is not restricted to the case of symmetric users, as in [4], [5].

It is well known that the long-term average throughput under a total power constraint is maximized by letting only the user with the best channel transmit at any time and frequency (subchannel) [4], [18], [19]. A system based on such a "max-gain" allocation yields system E_b/N_0 parametrically given by the expression

$$\begin{aligned} \mathsf{C} &= \int_0^\infty \log_2(1 + x\mathsf{SNR}) \mathrm{d}F_{\max\{d\},K}(x) \\ \left(\frac{E_b}{N_0}\right)_{\text{sys}}^{\text{max-gain}} &= \frac{\mathsf{SNR}}{\mathsf{C}} \end{aligned} \tag{26}$$

where $F_{\max\{d\},K}(x)$ denotes the cdf of $\max\{d_1^m, \ldots, d_K^m\}$ (independent of m because of the above symmetry assumption).

Next, we consider a "near–far" situation typical of wireless cellular systems. Signal propagation is characterized by a frequency-flat factor that depends on the distance between the user terminal and the base station (path loss), and by a frequency-selective "small-scale" fading that depends on the local scattering environment around the user terminal [20]. The path loss varies so slowly in time with respect to the signal bandwidth that it can be considered constant forever. This corresponds to the realistic assumption that users do not change significantly their distance from the base station during a large number of consecutive slots. On the contrary, the small-scale fading changes in time depending on the channel Doppler bandwidth: its coherence time is such that the fading gain can be considered constant on each slot, but varying according to some stationary ergodic (possibly correlated) process from slot to slot. This model is referred to as block fading [1].

We model these two effects by letting $d_k^m = s_k f_k^m$, where s_k denotes the path loss of user k (symbol s stands for "slow") and f_k^m is the frequency-selective block fading of user k in channel m (symbol f stands for "fast"). Clearly, s_k and f_k^m are mutually statistically independent, as they are due to completely different propagation effects. Appendix A provides expressions for the channel gain statistics that we use in our numerical results, that are relevant to typical cellular systems. In particular, we notice that the symmetry assumption is satisfied when the users are at fixed distance from the base station and each frequency-selective channel obeys the classical uncorrelated scattering wide-sense-stationary fading model [20], since the small-scale fading is identically distributed at each frequency.

In such a near–far situation, the above "max-gain" scheduling would result in a very unfair sharing of the channel resource, letting basically only the users that are very close to the base station to transmit. Then, the PFS algorithm has been proposed to alleviate this problem [5]. PFS allocates user k on channel m at any given slot t if $\hat{k}_m = k$, where

$$\hat{k}_m(t) = \arg \max_{k'=1,\dots,K} \frac{\log(1 + d_{k'}^m(t)\mathsf{SNR})}{T_{k'}(t)} \tag{27}$$

where $d_k^m(t)$ denotes the gain of the mth channel of user k at slot time t and $T_k(t)$ denotes the long-term average throughput of user k at time t. The long-term average user throughputs are recursively computed by

$$\begin{aligned} T_k(t) = (1-\gamma)&T_k(t-1) \\ &+\gamma \sum_{m=1}^{M} \log(1 + d_k^m(t)\mathsf{SNR}) 1\{\hat{k}_m(t) = k\}. \end{aligned} \tag{28}$$

The parameter γ regulates the size of the time window over which fairness is imposed. Very small γ implies very large delay jitter and large total average throughput. On the contrary, large γ forces the system to schedule users in an almost round-robin fashion, as in the case of conventional TDMA/FDMA. Here, in order to stress the difference between the delay-limited system and the opportunistic (variable-delay) system, we shall consider the PFS in the limit of vanishing γ.

In our case, we have that $d_k^m(t) = s_k f_k^m(t)$, where s_k is fixed (distance dependent) and $f_k^m(t)$ is a stationary ergodic process with respect to t, independent of s_k, identically distributed for all k, and symmetrically distributed with respect to m. Then, the limit $\bar{T}_k = \lim_{t\to\infty} \frac{1}{t}\sum_{\tau=1}^{t} T_k(\tau)$ exists. By letting $\gamma \to 0$ and initializing $T_k(0) = \bar{T}_k$, the PFS scheduling rule becomes

$$\hat{k}_m(t) = \arg \max_{k'=1,\dots,K} \frac{\log(1 + d_{k'}^m(t)\mathsf{SNR})}{\bar{T}_{k'}}. \tag{29}$$

It can be shown (see [21], [22]) that the PFS in the limit of vanishing γ maximizes $\sum_{k=1}^{K} \log \bar{T}_k$ over all possible scheduling algorithms (proportional fairness property). We shall take this as the defining property of PFS in this regime. We have the following statement.

Theorem 1: For any given K and fixed path loss components $\boldsymbol{s} = (s_1, \dots, s_K)$, under the channel gain statistics defined above, the long-term average throughput achieved by PFS is given by

$$\sum_{k=1}^{K} \bar{T}_k = \frac{M}{K} \sum_{k=1}^{K} \int_0^\infty \log(1 + s_k x \mathsf{SNR}) \mathrm{d}F_{\max\{f\},K}(x) \tag{30}$$

where $F_{\max\{f\},K}(x)$ is the cdf of $\max\{f_1^m, \dots, f_K^m\}$ (independent of the channel index m).

Proof: See Appendix C. □

Assume now that the path loss s_k is identically distributed for all users (e.g., in Appendix A we consider the case of users independently and uniformly distributed in a disk-shaped cell). As a corollary of Theorem 1, it follows that the average spectral efficiency C as a function of $(E_b/N_0)_{\text{sys}}$, where expectation is taken also with respect to the (random) path loss, is given implicitly by

$$\begin{aligned} \mathsf{C} &= \int_0^\infty \log_2(1 + x\mathsf{SNR}) \mathrm{d}F_{s\max\{f\},K}(x) \\ \left(\frac{E_b}{N_0}\right)_{\text{sys}}^{\text{PFS}} &= \frac{\mathsf{SNR}}{\mathsf{C}} \end{aligned} \tag{31}$$

where $F_{s\max\{f\},K}(x)$ denotes the cdf of $s\max\{f_1^m, \dots, f_K^m\}$ and s is distributed as the path loss of a random user in the system.

This parallels the expression (26) for the (unfair) max-gain scheduling, with the replacement of $F_{\max\{d\},K}$ (cdf of the maximum over the combined path loss and fading gain) with $F_{s\max\{f\},K}(x)$ (cdf of the maximum over the fading, with random path loss).

Figs. 4 and 5 compare the spectral efficiency achieved by PFS and delay-limited systems for finite number of users $K = 10, 20, 30, 50, 100$ (in Fig. 5 we show only the case $K = 10$ and $K = \infty$ for the optimized-orthogonal system for the sake of clarity). In all cases, spectral efficiency *improves* with K. This effect is known as *multiuser diversity*. However, the effect of multiuser diversity is quite different in the delay-limited and delay-tolerant setting. While for the delay-tolerant systems increasing K yields a gain in terms of $(E_b/N_0)_{\text{sys}}$ for all spectral efficiencies (roughly, a horizontal shift of the C versus $(E_b/N_0)_{\text{sys}}$ curve), for the delay-limited systems increasing K yields a change only for large spectral efficiency. This effect will be analyzed in depth in Section V, where the high and low spectral efficiency region will be explicitly addressed.

IV. Delay-Limited Systems for a Large Number of Users

In this section, we study the delay-limited systems in the limit of large K. As we shall see, this asymptotic analysis yields both elegant closed-form expressions for $(E_b/N_0)_{\text{sys}}$ as a function of spectral efficiency C, and some interesting considerations on system design that will be pointed out later on. We make the following assumptions:

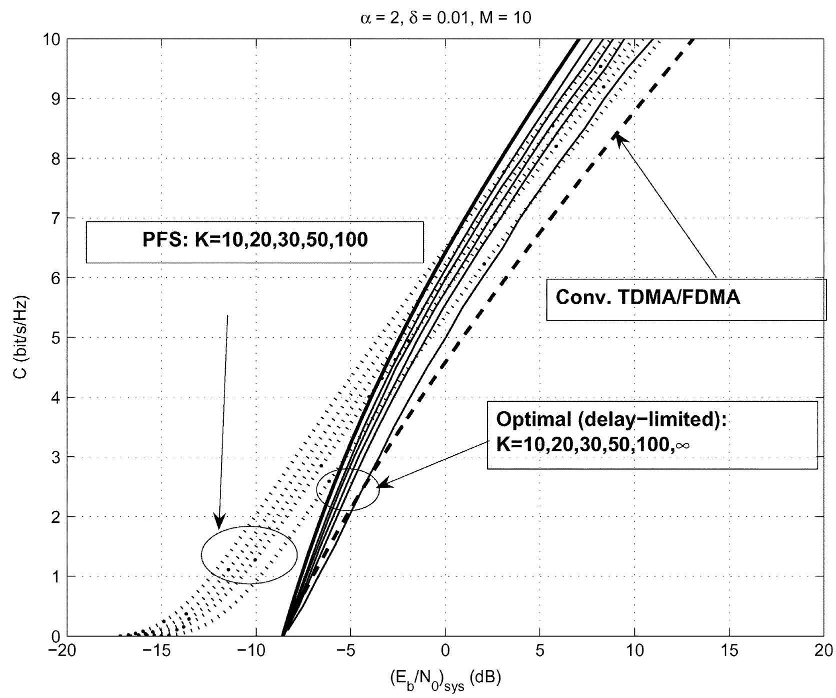

Fig. 4. Spectral efficiency versus system E_b/N_0 for PFS and optimal delay-limited signaling. The curve for conventional TDMA/FDMA and $K = \infty$ users is shown for comparison. The channel parameters are $M = 10$, path loss exponent $\alpha = 2$, and radius of the forbidden region around the base station (see Appendix A) $\delta = 0.01$.

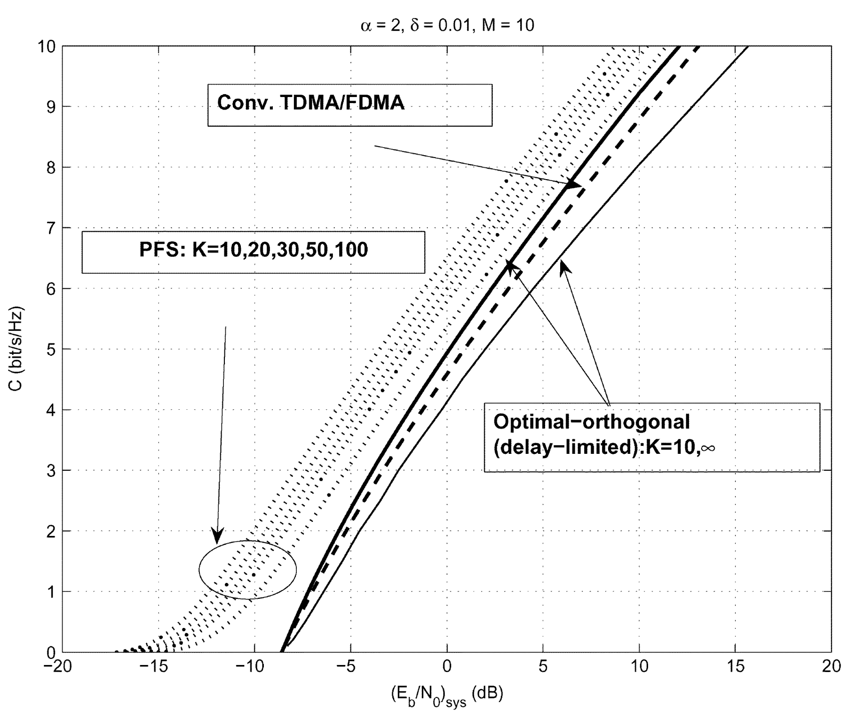

Fig. 5. Spectral efficiency versus system E_b/N_0 for PFS and optimized-orthogonal delay-limited signaling. The curve for conventional TDMA/FDMA and $K = \infty$ users is shown for comparison. The channel parameters are $M = 10$, path loss exponent $\alpha = 2$, and radius of the forbidden region around the base station (see Appendix A) $\delta = 0.01$.

[A1] M is fixed while K becomes arbitrarily large.

[A2] As $K \to \infty$, the empirical joint channel gain distribution, defined by

$$F^{(K)}(x^1,\dots,x^M)=\frac{1}{K}\sum_{k=1}^{K}\prod_{m=1}^{M}1\{d_k^m \le x^m\} \quad (32)$$

converges almost surely to a given deterministic cdf $F(x^1,\dots,x^M)$. Moreover, $F(\cdot)$ is assumed to be symmetric, in the sense already defined before. In particular, the marginal cdfs of $F(\cdot)$ are identical.

[A3] For a given system throughput Γ, the user individual rates are given by $R_k = \frac{\Gamma}{K}\nu_k$, where ν_k is the rate allocation factor for user k. As $K \to \infty$, the empirical rate allocation distribution, defined by

$$G^{(K)}(x) = \frac{1}{K}\sum_{k=1}^{K}1\{\nu_k \le x\} \quad (33)$$

converges almost surely to a given deterministic cdf $G(x)$ with mean 1 and support in $[a,b]$ as $K \to \infty$, where $0 \le a \le b < \infty$ are constants independent of K.

[A4] The rate allocation factors are fixed *a priori*, independently of the realization of the channel gains. Therefore, the empirical joint distribution of $\{(d_k^1,\dots,d_k^M,\nu_k) : k=1,\dots,K\}$ converges to the product cdf $F(x^1,\dots,x^M)G(z)$. We remark here that this assumption reflects the delay-limited nature of the problem: the user rates are fixed *a priori* and independently of the channel gain realization.

The performance of delay-limited systems in the limit of large number of users is given by the following results.

Theorem 2: Under the assumptions A1, A2, A3, and A4, as $K \to \infty$ the minimum $(E_b/N_0)_{\rm sys}$ for given system spectral efficiency C is given by

$$\left(\frac{E_b}{N_0}\right)_{\rm sys} = \log(2)\int_0^\infty 2^{\mathsf{C}F_{\max}(x)}\frac{\mathrm{d}F_{\max}(x)}{x} \quad (34)$$

where $F_{\max}(x)$ is the limit cdf of the empirical distribution of $\max\{d_k^1,\dots,d_k^M\}$, as $K \to \infty$. This is achieved by letting each user transmit on its best subchannel only, and by using superposition coding and successive decoding on each subchannel.

Proof: See Appendix D. □

Theorem 3: Under the assumptions A1, A2, A3 and A4, as $K \to \infty$ the minimum $(E_b/N_0)_{\rm sys}$ for given spectral efficiency C achieved by orthogonal signaling is given by

$$\left(\frac{E_b}{N_0}\right)_{\rm sys} = \log(2)\int_0^\infty \frac{\exp\left(1+W\left(\frac{\mu x-1}{e}\right)\right)-1}{1+W\left(\frac{\mu x-1}{e}\right)}\frac{\mathrm{d}F_{\max}(x)}{x} \quad (35)$$

where $F_{\max}(x)$ is defined as in Theorem 2, $W(x)$ is Lambert's W function, and where μ is the solution of

$$\int_0^\infty \frac{\mathrm{d}F_{\max}(x)}{1+W\left(\frac{\mu x-1}{e}\right)} = \frac{1}{\mathsf{C}\log(2)}. \quad (36)$$

This is achieved by letting each user transmit on its own best subchannel only, and by using orthogonal signaling with optimized fractions on each subchannel.

Proof: See Appendix E. □

We shall compare the optimal and the optimized-orthogonal delay-limited systems of Theorems 2 and 3 with a conventional TDMA/FDMA system, where each user chooses its own best channel to transmit, but resource allocation (the fractions $\boldsymbol{\Theta}^m$) are proportional to the users' requested rates, disregarding the actual channel gains. Interestingly, most "radio resource management" schemes in today's wireless systems follow approximately this rule and therefore they are suboptimal. The performance of conventional TDMA/FDMA is given by

Theorem 4: Under the assumptions A1, A2, A3, and A4, as $K \to \infty$ the $(E_b/N_0)_{\rm sys}$ for given system spectral efficiency C, achieved by letting each user transmit on its best subchannel only and allocating a fraction of channel uses proportional to its individual rate, is given by

$$\left(\frac{E_b}{N_0}\right)_{\rm sys} = \frac{2^{\mathsf{C}}-1}{\mathsf{C}}\int_0^\infty \frac{\mathrm{d}F_{\max}(x)}{x} \quad (37)$$

where $F_{\max}(x)$ is defined as in Theorem 2.

Proof: See Appendix F. □

Not surprisingly, a conventional TDMA/FDMA system that does not make use of optimized resource allocation fractions $\boldsymbol{\Theta}^m$ as given by Theorem 4 behaves like a single-user system with spectral efficiency C and channel gain $d \sim F_{\max}(x)$. This is because each user, in order to maintain its own rate on every slot, has to invert its channel as if it was alone in the system. In fact, (37) coincides with the spectral efficiency versus E_b/N_0 for a single-user system under the *channel inversion* power control strategy.

For the sake of further comparisons with the PFS system, we notice here that for the channel model accounting for path loss and frequency selective fading $d_k^m = s_k f_k^m$, introduced in Section III-B and, in greater details, in Appendix A, the limit cdf $F_{\max}(x)$ coincides with $F_{s\max\{f\},M}(x)$, i.e., the cdf of the product of a random path loss s with $\max\{f_k^1,\dots,f_k^M\}$. Intuitively, since $K \gg M$, the PFS is able to exploit a much larger multiuser diversity (order K) than the delay-limited systems (order M). This will be evident from the results of next section.

V. Low and High Spectral Efficiency Behaviors

We wish to quantify the multiuser diversity gain of the optimal delay-limited systems of Theorems 2 and 3 over the conventional TDMA/FDMA system of Theorem 4. Then, we shall also quantify the loss incurred by delay-limited systems with respect to opportunistic variable-delay variable-rate systems based on PFS.

Our comparison is based on the spectral efficiency C as a function of $(E_b/N_0)_{\rm sys}$. In particular, we focus on the low and high spectral efficiency regimes, as defined in [9], [11]. In general, the low spectral efficiency behavior ($\mathsf{C} \downarrow 0$) is characterized by the minimum system E_b/N_0, denoted by $(E_b/N_0)_{\min}$ and the *wideband slope* $\mathcal{S}_0$, such that [9]

$$\left(\frac{E_b}{N_0}\right)_{\rm sys}\bigg|_{\rm dB} = \left(\frac{E_b}{N_0}\right)_{\min}\bigg|_{\rm dB} + \frac{\mathsf{C}}{\mathcal{S}_0}10\log_{10}(2)+o(\mathsf{C}). \quad (38)$$

The high spectral efficiency behavior ($\mathsf{C}\to\infty$) is characterized by the high-SNR slope $\mathcal{S}_\infty$ and by the horizontal decibel penalty $\mathcal{L}_\infty$, such that [11]

$$\left(\frac{E_b}{N_0}\right)_{\rm sys}\bigg|_{\rm dB} = \frac{\mathsf{C}}{\mathcal{S}_\infty}10\log_{10}(2) -10\log_{10}(\mathsf{C}) + \mathcal{L}_\infty 10\log_{10}(2) + o(1). \quad (39)$$

We start with the low spectral efficiency regime. All the delay-limited systems achieve the same $(E_b/N_0)_{\min}$ that can be obtained by letting $\mathsf{C}\to 0$ in (34), (35), and in (37), and it is given by

$$\left(\frac{E_b}{N_0}\right)_{\min} = \log(2)\int_0^\infty \frac{{\rm d}F_{\max}(x)}{x}. \quad (40)$$

The advantage of optimal over conventional delay-limited signaling is evidenced by the wideband slope, provided by the following theorem.

Theorem 5: Under the assumptions A1, A2, A3, and A4, as $K\to\infty$ the wideband slope $\mathcal{S}_0$ (in bit/dimension/3 dB) of the spectral efficiency versus $(E_b/N_0)_{\rm sys}$ curve for the delay-limited systems is given by

$$\mathcal{S}_0^{\rm optimal} = \frac{\int \frac{{\rm d}F_{\max}(x)}{x}}{\int \frac{F_{\max}(x)}{x}{\rm d}F_{\max}(x)} \quad (41)$$

$$\mathcal{S}_0^{\rm opt.orthogonal} = \frac{2\int \frac{{\rm d}F_{\max}(x)}{x}}{\left(\int \frac{1}{\sqrt{x}}{\rm d}F_{\max}(x)\right)^2} \quad (42)$$

$$\mathcal{S}_0^{\rm conv.tdma/fdma} = 2. \quad (43)$$

Proof: See Appendix H. □

The low spectral efficiency behavior of the PFS system is easily obtained from (31) by letting $\mathsf{SNR}\to 0$ and using the results of [9]. We have

$$\left(\frac{E_b}{N_0}\right)_{\min}^{\rm PFS} = \frac{\log(2)}{\int_0^\infty x{\rm d}F_{s\max\{f\},K}(x)} = \frac{\log(2)}{\mathbb{E}[s]\mathbb{E}[\max\{f_1^m,\dots,f_K^m\}]}. \quad (44)$$

Notice that, under mild conditions on the fading distribution, $(E_b/N_0)_{\min}^{\rm PFS}$ goes to zero as $K\to\infty$. For the channel statistics of Appendix A, f_k^m are independent and identically distributed (i.i.d.) central chi-squared distributed and, for large K, we have that [23]

$$P\left(\log K - \log\log K \le \max\{f_1^m,\dots,f_K^m\} \le \log K + \log\log K\right) \ge 1 - O\left(\frac{1}{\log K}\right) \quad (45)$$

Since $\mathbb{E}[s]$ is finite, $(\frac{E_b}{N_0})_{\min}^{\rm PFS}$ goes to zero as $O(\frac{1}{\log K})$.

As far as the wideband slope is concerned, direct application of the results in [11] yields

$$\mathcal{S}_0^{\rm PFS} = \frac{2\left(\mathbb{E}[s]\mathbb{E}\left[\max\{f_1^m,\dots,f_K^m\}\right]\right)^2}{\mathbb{E}[s^2]\mathbb{E}\left[\left(\max\{f_1^m,\dots,f_K^m\}\right)^2\right]}. \quad (46)$$

When (45) holds, it is easy to see that

$$\lim_{K\to\infty}\mathcal{S}_0^{\rm PFS} = \frac{2\mathbb{E}[s]^2}{\mathbb{E}[s^2]}.$$

By comparing (40) and (44) under the channel statistics of Appendix A, we notice that for low spectral efficiency the gain of the opportunistic scheme over the delay-limited scheme is twofold: on the one hand it achieves larger multiuser diversity as $K\gg M$, on the other hand it achieves a "Jensen's inequality" gain due to the convexitiy of $1/x$. We conclude that for low spectral efficiency the cost of imposing a strict constraint on rate and delay is very high. In fact, the optimal delay-limited system does not benefit in terms of $(E_b/N_0)_{\min}$ over a conventional orthogonal system (or a single-user system). In this regime, multiuser diversity appears only as a second-order effect, as a gain in the wideband slope.

Next, we focus on the high spectral efficiency regime. In this case, the spectral efficiency slope of optimal delay-limited signaling in the limit of large K is easily obtained from the definition (39) and Theorem 2 as

$$\begin{aligned}\mathcal{S}_\infty^{\rm optimal} &= \lim_{(E_b/N_0)_{\rm sys}\to\infty}\frac{\mathsf{C}}{\left(\frac{E_b}{N_0}\right)_{\rm sys}\Big|_{\rm dB}}10\log_{10}(2)\\ &= \lim_{\mathsf{C}\to\infty}\frac{\mathsf{C}\log(2)}{\log\left(\int_0^\infty \exp(\mathsf{C}\log(2)F_{\max}(x))\frac{{\rm d}F_{\max}(x)}{x}\right)}\\ &= 1\end{aligned} \quad (47)$$

where the last step follows as an application of Varadhan's integral lemma [24]. High-SNR slope equal to 1 is not surprising, and it is a common feature of *any* scheme that makes full use of all system degrees of freedom. After rather trivial calculations it is easy to show that all other systems considered in this paper achieve the same $\mathcal{S}_\infty = 1$. However, they may differ significantly in their horizontal decibel penalty, as it will be illustrated in the remainder of this section. The conventional TDMA/FDMA system yields (calculation is immediate)

$$\mathcal{L}_\infty^{\rm conv.tdma/fdma} = \log_2\left(\int_0^\infty \frac{{\rm d}F_{\max}(x)}{x}\right). \quad (48)$$

For the optimal delay-limited signaling, we have the following surprising behavior, already noticed in [10, Sec. 5.2.2] for the case of frequency-flat path-loss only.

Theorem 6: Under the assumptions A1, A2, A3, and A4, as $K\to\infty$ the horizontal decibel penalty of the optimal delay-limited system is given by

$$\mathcal{L}_\infty^{\rm optimal} = -\log_2\left(F_{\max}^{-1}\left(1-\frac{1}{\mathsf{C}}\right)\right) + O(1) \quad (49)$$

Proof: See Appendix I. □

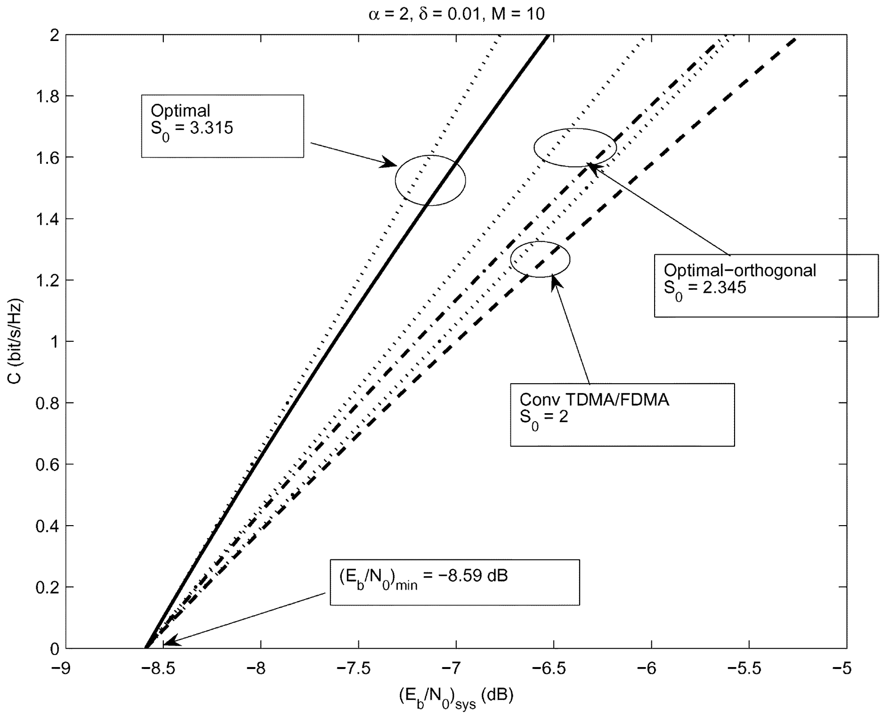

Fig. 6. Spectral efficiency versus system E_b/N_0 for the optimal, optimized-orthogonal, and conventional TDMA/FDMA delay-limited systems for $K = \infty$. The channel parameters are $M = 10$, path loss exponent $\alpha = 2$, and radius of the forbidden region around the base station (see Appendix A) $\delta = 0.01$. The dotted lines correspond to the low spectral efficiency approximation (38).

In particular, for $M = 1$ and $d_k^m = s_k$ (path-loss only), when s_k is distributed as in (52), we have that $\mathcal{L}_\infty^{\text{optimal}}$ diverges to $-\infty$ as

$$-\frac{\alpha}{2}\log_2 \mathsf{C}$$

where α is the path-loss exponent. This is the same result found in [10, Sec. 5.2.2] using a more refined (but more complicated) direct calculation. For the case $d_k^m = f_k^m$ (frequency-selective fading only, path-loss equal to 1), then $F_{\max}(x)$ is given by (54) and we obtain that $\mathcal{L}_\infty^{\text{optimal}}$ diverges to $-\infty$ as

$$-\log_2 \log \mathsf{C}.$$

In both cases, and more in general, in all cases where $F_{\max}(x)$ is strictly increasing for all sufficiently large x, the horizontal decibel "penalty" diverges to $-\infty$, indicating that optimal delay-limited signaling yields unbounded decibel *gain* over the corresponding conventional TDMA/FDMA system.

The following result provides the horizontal decibel penalty of optimized-orthogonal delay-limited signaling.

Theorem 7: Under the assumptions A1, A2, A3, and A4, as $K \to \infty$ the horizontal decibel penalty of the optimized-orthogonal delay-limited system is given by

$$\mathcal{L}_\infty^{\text{opt.orthogonal}} = -\int_0^\infty \log_2(x)\mathrm{d}F_{\max}(x) \qquad (50)$$

Proof: See Appendix J. □

Finally, for the opportunistic PFS system we obtain, after simple direct calculation

$$\mathcal{L}_\infty^{\text{PFS}} = \lim_{\mathsf{SNR}\to\infty}\Bigg(\log_2 \mathsf{SNR} - \int_0^\infty \log_2(1 + x\mathsf{SNR})\mathrm{d}F_{s,\max\{f\},K}(x)\Bigg)$$
$$= -\int_0^\infty \log_2(x)\mathrm{d}F_{s,\max\{f\},K}(x). \qquad (51)$$

In the limit of large K, for channel statistics such that (45) holds, we have the behavior $\mathcal{L}_\infty^{\text{PFS}} = O(\log\log K)$, typical of multiuser diversity systems [25], [5], [23].

Comparing (48), (50), and (51) we notice that, since $-\log x$ is convex,

$$\mathcal{L}_\infty^{\text{conv.tdma/fdma}} \le \mathcal{L}_\infty^{\text{opt.orthogonal}}$$

by Jensen's inequality. Furthermore, for the channel statistics of Appendix A, where $F_{\max}(x) = F_{s\max\{f\},M}(x)$, in the usual case where the number of users is much larger than the number of subchannels $(K \gg M)$ we have that

$$\mathcal{L}_\infty^{\text{opt.orthogonal}} \le \mathcal{L}_\infty^{\text{PFS}}.$$

This quantifies the advantage of optimized-orthogonal versus conventional TDMA/FDMA signaling, and the advantage of PFS (which is also an orthogonal system) over the orthogonal delay-limited systems in the high spectral efficiency region.

Note that the gain of PFS comes only from $K > M$ and the diversity associated with it. For M growing large, the gain

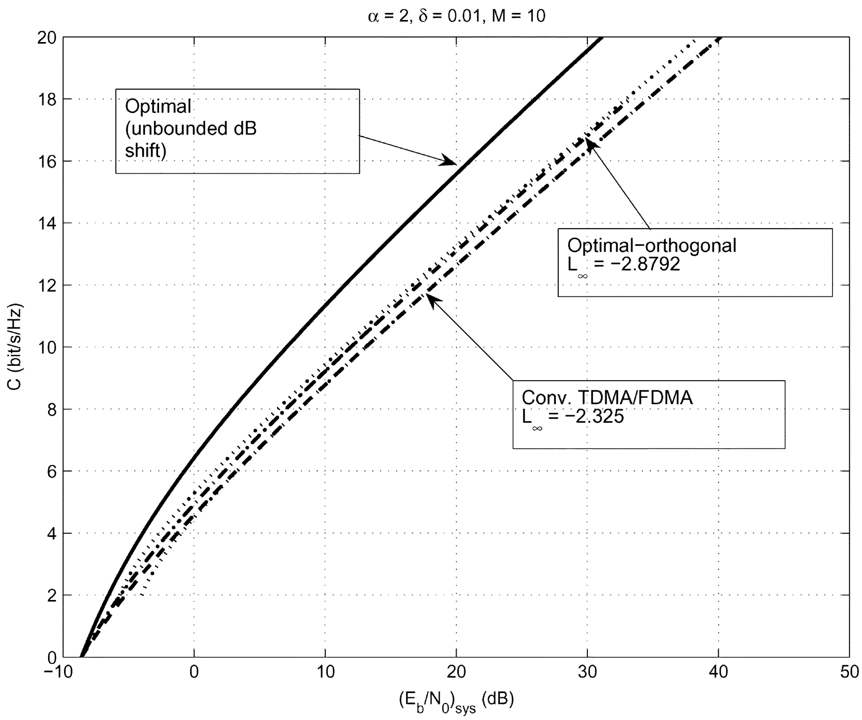

Fig. 7. Spectral efficiency versus system E_b/N_0 for the optimal, optimized-orthogonal, and conventional TDMA/FDMA delay-limited systems for $K = \infty$. The channel parameters are $M = 10$, path loss exponent $\alpha = 2$, and radius of the forbidden region around the base station (see Appendix A) $\delta = 0.01$. The dotted lines correspond to the high spectral efficiency approximation (39).

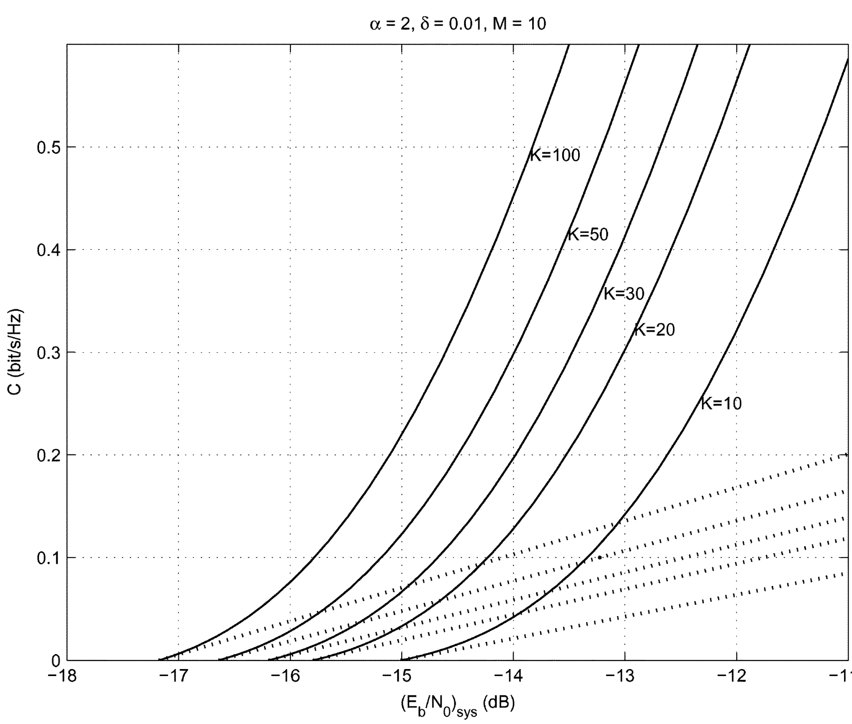

Fig. 8. Spectral efficiency versus system E_b/N_0 for the PFS system for $K = 10, 20, 30, 50, 100$ (C is increasing with K). The channel parameters are $M = 10$, path loss exponent $\alpha = 2$, and radius of the forbidden region around the base station (see Appendix A) $\delta = 0.01$. The dotted lines correspond to the low spectral efficiency approximation (38).

vanishes. Thus, the wider the band, the less advantage for PFS in terms of spectral efficiency, despite the looser delay constraint.

Figs. 6–9 show the low and high spectral efficiency behavior of all systems considered. In particular, we observe that, consistently with Theorem 6, the gain of the optimal delay-limited signaling over orthogonal signaling becomes unbounded as $\mathsf{C} \to \infty$, although for this case of channel statistics the gain grows quite slowly.

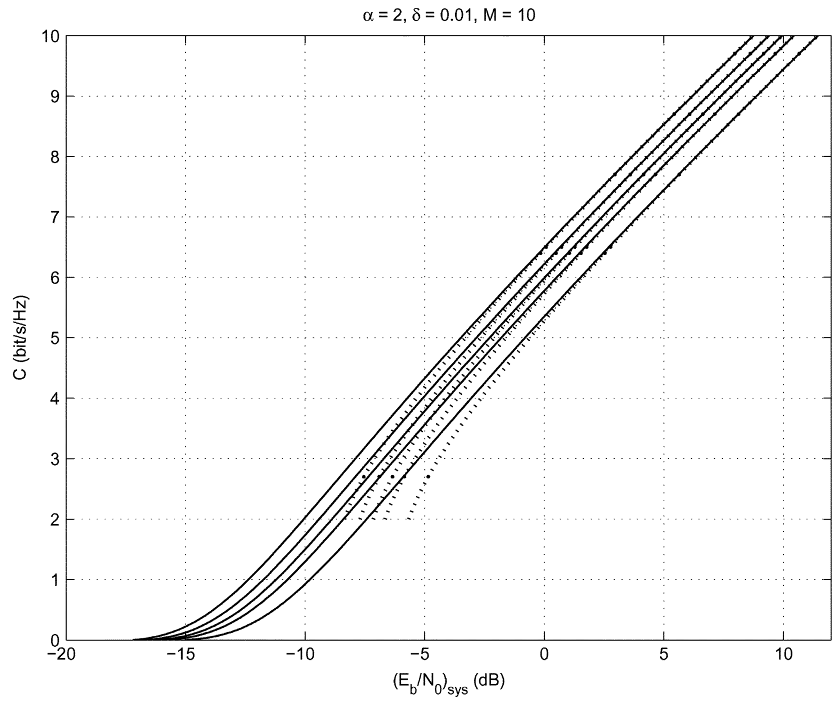

Fig. 9. Spectral efficiency versus system E_b/N_0 for the PFS system for $K = 10, 20, 30, 50, 100$ (C is increasing with K). The channel parameters are $M = 10$, path loss exponent $\alpha = 2$, and radius of the forbidden region around the base station (see Appendix A) $\delta = 0.01$. The dotted lines correspond to the high spectral efficiency approximation (39).

VI. Conclusion

The main message of this paper is that, in order to achieve high system spectral efficiency while coping with fairness constraints, it is not necessary to relax the individual users rate requirements from "instantaneous" (i.e., given target rates in any coding block) to "long-term averages." On the contrary, exploiting the differences in the user path losses, and in the fading multipath frequency selectivity, hard fairness can be achieved without substantial losses (even more: for sufficiently high spectral efficiency the delay-limited system can outperform PFS). The inherent diversity given by path losses (near–far diversity), is a new form of *multiuser diversity* that PFS is not able to exploit. This might have an impact on the system design in certain applications where fixed rates and strict delays are important, such as video streaming over the wireless cell. Furthermore, an interesting question arises about the near–far diversity effect in a multiple-cell environment. In fact, since the users at the cell border are decoded last, they do not have to boost their power as in a conventional power-controlled wireless system. These users are responsible for most of the outer-cell interference and therefore we expect that the optimal delay-limited policy (even for the single cell presented here) yields significant gains also in a multiple-cell environment. This, as well as possible variations involving coordinated processing at multiple-cell sites, is left for future work.

However, in order to take full advantage of the near–far diversity, successive decoding at the base station (uplink) and superposition coding with successive interference cancellation at the user terminals (downlink) is needed. The optimized-orthogonal delay-limited scheme does not provide significant gains with respect to conventional TDMA/FDMA.

On the practical side, we notice that the asymptotically optimal strategy for both optimal and orthogonal signaling schemes consists of letting each user transmit on its own best subchannel only, irrespectively of the other users. This result suggests a system where the users are able to measure their SNR over a large number of (narrowband) subchannels and pick the most favorable subchannel for transmission. For a large number of users, this can be done in a completely decentralized fashion, i.e., independently of the other users. This feature reminds certain current proposals referred to as "Cognitive Radio" (see, for example, [26]). However, we hasten to say that in order to achieve optimality the transmission power (and resource-sharing factors in the case of orthogonal signaling) must be coordinated by the base station for each group of users sharing the same subchannel.

Appendix A
Channel Statistic

In cellular communications, signal propagation is typically characterized by a frequency-flat factor that depends on the distance between the user terminal and the base station (path loss), and by a frequency-selective "small-scale" fading that depends on the local scattering environment around the user terminal [20]. The path loss varies so slowly in time with respect to the signal bandwidth that it can be considered constant forever. This corresponds to the realistic assumption that users do not change significantly their distance from the base station during a large number of consecutive slots. On the contrary, the small-scale fading changes in time depending on the channel Doppler bandwidth. In practice, its coherence time is such that it can be considered constant on each slot, but changing according to some stationary ergodic (possibly correlated) process from slot to slot. This model is referred to as block fading [1].

We take into account these two effects by letting $d_k^m = s_k f_k^m$, where s_k denotes the path loss of user k (symbol s stands for "slow") and f_k^m is the frequency-selective block fading of user k in channel m (symbol f stands for "fast"). Clearly, s_k and f_k^m are mutually statistically independent, as they are due to completely different propagation effects.

Path loss is typically modeled by a monomial signal decay with distance [27], [28]. This is why received signal strength imbalance due to path loss is usually referred to as the "near–far effect." The path loss takes on the form $s_k = D_k^{-\alpha}$, where D_k denotes the distance from the base station and α is the path loss exponent, ranging typically in the interval [2], [4]. Considering users uniformly distributed within a circular cell with unit radius, where the base station is placed at the center of the cell, this results in the following cdf:

$$F_s(x) = \begin{cases} 0, & x < 1 \\ 1 - x^{-2/\alpha}, & x \geq 1. \end{cases} \tag{52}$$

Although the path-loss distribution (52) is widely used in the analysis of *conventional* cellular systems [27], [28], it presents the annoying fact that the path-loss diverges when the distance between terminal and base station becomes small. While this has little effect in the analysis of conventional systems, it might yield completely meaningless conclusions in the case of systems that take full advantage of the channel knowledge at the transmitter. For example, in the case of PFS this model yields $(E_b/N_0)_{\min}^{(\text{PFS}} = 0$ for any number of users K, since the denominator of (44) is unbounded.

For this reason, we shall assume that the users are uniformly distributed in the unit-circle cell, but for a forbidden circular region of radius δ centered around the base station, where $0 < \delta < 1$ is a fixed system constant. Under this model, the path-loss cdf is given by

$$F_s(x) = \begin{cases} 0, & x < 1 \\ 1 - \frac{x^{-2/\alpha} - \delta^2}{1-\delta^2}, & 1 \leq x < \delta^{-\alpha} \\ 1, & x \geq \delta^{-\alpha} \end{cases} \tag{53}$$

where we let the path loss at the cell border be equal to 1. Clearly, all our results in terms of $(E_b/N_0)_{\text{sys}}$ must be scaled by a factor $D_0^{-\alpha}$ equal to the path loss at the cell border, where D_0 denotes the actual radius of the cell. Of course, the conclusions of our analysis do not change after this scaling, which applies to all systems in the same way. However, the numerical results in terms of spectral efficiency versus system E_b/N_0 need to be rescaled in order to obtain practically meaningful values. The path loss statistics (53) are even closer to reality, as for distances in the order of wavelengths or below, the propagation loss does not scale monomially with distance due to electromagnetic near-field effects. Furthermore, with users independently and uniformly distributed in the cell, it is clear that our assumption A2 holds. In fact, the empirical cdf of the path losses $\{s_k, \ldots, s_K\}$ converges to $F_s(x)$ in (53) as $K \to \infty$.

Frequency-selective block fading is modeled by a channel transfer function $H_k(\omega)$ that, for every fixed frequency ω, is a zero-mean Gaussian circularly symmetric random variable (Rayleigh fading). Under very mild conditions on the channel impulse response statistics, $H_k(\omega)$ is identically distributed for every ω. By slicing the channel bandwidth into M subbands, each of which is smaller than the channel coherence bandwidth [20], the physical channel is well modeled by M parallel channels each affected by a channel gain $f_k^m = |H_k(\omega_m)|^2$ for $m = 1, \ldots, M$, where ω_m denotes the center frequency of each subband. Assuming that the subbands are sufficiently far apart, we may consider $f_k^1, \ldots, f_k^M$ as i.i.d. Hence, the distribution of $\max\{f_k^m, \ldots, f_k^M\}$ is given by

$$F_{\max\{f\},M}(x) = (1 - \mathrm{e}^{-x})^M. \tag{54}$$

Recall that $F_{\max}(x)$ in Theorem 2 is defined as the cdf of the random variable

$$\max\{d_k^1, \ldots, d_k^M\} = s_k \max\{f_k^1, \ldots, f_k^M\}.$$

Hence, it coincides with $F_{s,\max\{f\},M}(x)$, as introduced in Section III-B in the analysis of the PFS. Then, in the remainder of this appendix we shall focus on obtaining convenient expressions for $F_{s,\max\{f\},M}(x)$.

If path loss and Rayleigh fading occur simultaneously and independently, we have

$$\begin{aligned} F_{s\max\{f\},M}(x) &= \int F_s(x/y)\,\mathrm{d}F_{\max\{f\},M}(y) &(55)\\ &= F_{\max\{f\},M}(x\delta^\alpha) \\ &\quad + \int_{x\delta^\alpha}^{x} \left(1 - \frac{(y/x)^{2/\alpha} - \delta^2}{1-\delta^2}\right) \mathrm{d}F_{\max\{f\},M}(y) &(56)\\ &= \frac{1}{1-\delta^2}\frac{1}{x^{2/\alpha}} \int_{x^{2/\alpha}\delta^2}^{x^{2/\alpha}} F_{\max\{f\},M}(y^{\alpha/2})\,\mathrm{d}y &(57)\end{aligned}$$

where the last line is obtained after a change of variable and integration by parts.

The integral in (55) can be given in closed form in some special cases. In particular, for $\alpha = 2$ after some algebra we find

$$\begin{aligned} F_{s\max\{f\},M}^{\alpha=2}(x) &= \frac{1}{(1-\delta^2)x}\int_{x\delta^2}^{x}(1-\exp(-y))^M\,\mathrm{d}y \\ &= 1 - \frac{1}{(1-\delta^2)x}\sum_{i=1}^{M}\frac{1}{i} \\ &\quad \times [(1-\exp(-x))^i - (1-\exp(-x\delta^2))^i]. \quad (58)\end{aligned}$$

For $\alpha = 4$, in a similar manner we find

$$\begin{aligned} F_{s\max\{f\},M}^{\alpha=4}(x) = 1 - \frac{1}{(1-\delta^2)\sqrt{x}}\sum_{i=1}^{M}\binom{M}{i}(-1)^i \\ \times \sqrt{\frac{\pi}{i}}[Q(\sqrt{2ix}) - Q(\sqrt{2ix}\delta^2)] \quad (59)\end{aligned}$$

where, as usual, $Q(z) = \frac{1}{\sqrt{2\pi}}\int_z^\infty \exp(-u^2/2)\,\mathrm{d}u$ is the Gaussian tail function.

APPENDIX B
A WATER-FILLING-LIKE OPTIMIZATION PROBLEM

In resource allocation for both optimal signaling and optimized-orthogonal signaling, addressed in Section III-A, a key step in the coordinate descent iterative algorithm takes on the following form:

$$\min_{\boldsymbol{x}} \sum_{i=1}^{n} a_i \exp(x_i/\theta_i) \tag{60}$$

subject to $\sum_{i=1}^{n} x_i = b$ and $x_i \geq 0$ for all $i = 1, \ldots, n$, where a_i, θ_i and b are fixed positive constants.

This problem is convex. The Kuhn–Tucker conditions are given as

$$\frac{a_i}{\theta_i} \exp(x_i/\theta_i) - \lambda \geq 0 \tag{61}$$

where λ is the Lagrange multiplier for the equality constraint and where the ith inequality must hold with equality if the optimal solution has $x_i > 0$.

Solving for x_i in (61) we find that

$$x_i = \theta_i \left[\log \lambda - \log \frac{a_i}{\theta_i}\right]_+ \tag{62}$$

satisfies the Kuhn–Tucker conditions, and it is therefore the (unique) optimal point.

Finally, the Lagrange multiplier is obtained by solving

$$\sum_{i=1}^{n} \theta_i \left[\log \lambda - \log \frac{a_i}{\theta_i}\right]_+ = b$$

This equation can be efficiently solved as follows. Let π denote the sorting permutation such that

$$\frac{a_{\pi_1}}{\theta_{\pi_1}} \leq \frac{a_{\pi_2}}{\theta_{\pi_2}} \leq \cdots \leq \frac{a_{\pi_n}}{\theta_{\pi_n}}.$$

Then, the solution must take on the form

$$\log \lambda = \frac{b + \sum_{i=1}^{k} \theta_{\pi_i} \log \frac{a_{\pi_i}}{\theta_{\pi_i}}}{\sum_{i=1}^{k} \theta_{\pi_i}}$$

where k is the smallest integer $1, 2, \ldots, n$ such that

$$\frac{b + \sum_{i=1}^{k} \theta_{\pi_i} \log \frac{a_{\pi_i}}{\theta_{\pi_i}}}{\sum_{i=1}^{k} \theta_{\pi_i}} \leq \log \frac{a_{\pi_{k+1}}}{\theta_{\pi_{k+1}}}$$

(where we define $\log \frac{a_{\pi_{n+1}}}{\theta_{\pi_{n+1}}} = +\infty$).

APPENDIX C
PROOF OF THEOREM 1

Consider K users, with fixed path losses $s_1, \ldots, s_k$. Because of the symmetry of the small-scale fading distribution, the average throughput is the same for all subchannels. Hence, it is sufficient to focus on a single subchannel and multiply the final result by M. We shall drop the subchannel superscript m for the sake of notation simplicity. We denote by $f_1, \ldots, f_K$ the small-scale fading gains. They are i.i.d. and evolve in time according to a stationary ergodic process. Therefore, the K users are completely symmetric with respect to their small scale fading.

Consider a system that allocates the channel to the user $k = \arg\max\{f_1, \ldots, f_K\}$. This scheduling algorithm clearly achieves individual average throughputs given by

$$\tilde{T}_k = \frac{1}{K} \int_0^\infty \log(1 + s_k x \mathsf{SNR}) \mathrm{d}F_{\max\{f\},K}(x). \tag{63}$$

We let $\bar{T}_k$ and β_k denote the average throughput of user k and the fraction of slots allocated to user k under PFS, respectively. PFS is defined by the allocation rule

$$\hat{k}(t) = \arg \max_{k'=1,\ldots,K} \frac{\log(1 + s_{k'} f_{k'}(t) \mathsf{SNR})}{\bar{T}_{k'}}. \tag{64}$$

The average throughput of user k under PFS can be written as

$$\begin{aligned} \bar{T}_k &= \liminf_{t\to\infty} \frac{1}{t} \sum_{\tau=1}^{t} \log(1 + s_k f_k(\tau) \mathsf{SNR}) 1\{\hat{k}(\tau) = k\} \\ &= \liminf_{t\to\infty} \frac{|\mathcal{T}_k(t)|}{t} \frac{1}{|\mathcal{T}_k(t)|} \sum_{\tau \in \mathcal{T}_k(t)} \log(1 + s_k f_k(\tau) \mathsf{SNR}) \end{aligned} \tag{65}$$

where we have defined the set $\mathcal{T}_k(t)$ as the set of all slot times τ for which $\hat{k}(\tau) = k$. The rule (64) is a stationary policy characterized by some decision regions $\mathcal{D}_k$ that depend on $\boldsymbol{s}$ and on $\bar{T}_1, \ldots, \bar{T}_K$, which form a partition of $\mathbb{R}_+^K$. Now, we define a new random variable $f_{\max}^K(\tau)$ that has the same cdf $F_{\max\{f\},K}(x)$ of $\max\{f_1(\tau), \ldots, f_K(\tau)\}$, but it is statistically independent of $f_1(\tau), \ldots, f_K(\tau)$. We consider the following experiment: at each time $\tau \in \mathcal{T}_k(t)$, we generate independently $f_{\max}^{(K)}(\tau) \sim F_{\max\{f\},K}(x)$. By definition, we have that for $t \to \infty$ the $\limsup$ in probability of the random variable

$$\frac{1}{|\mathcal{T}_k(t)|} \sum_{\tau \in \mathcal{T}_k(t)} \log(1 + s_k f_k(\tau) \mathsf{SNR})$$

is smaller or equal to the limit in probability of

$$\frac{1}{|\mathcal{T}_k(t)|} \sum_{\tau \in \mathcal{T}_k(t)} \log(1 + s_k f_{\max}^K(\tau) \mathsf{SNR})$$

where the latter, by ergodicity, exists and it is equal to

$$\int_0^\infty \log(1 + s_k x \mathsf{SNR}) \mathrm{d}F_{\max\{f\},K}(x).$$

Since, again by ergodicity and by definition of β_k, the limit

$$\lim_{t\to\infty} \sum_{\tau=1}^{t} \frac{|\mathcal{T}_k(t)|}{t} \to \beta_k$$

holds almost surely, we have the upper bound

$$\bar{T}_k \leq \beta_k \int_0^\infty \log(1 + s_k x \mathsf{SNR}) \mathrm{d}F_{\max\{f\},K}(x). \tag{66}$$

This upper bound[3] together with (63) and the defining property of PFS, implying that $\sum_k \log \bar{T}_k \geq \sum_k \log \tilde{T}_k$, yield the inequality

$$\sum_k \frac{1}{K} \log \frac{1/K}{\beta_k} \leq 0 \qquad (67)$$

The left-hand side (LHS) of (67) is the divergence (cross entropy) between the uniform probability mass function $1/K$ and the probability mass function β_k. Since divergence is nonnegative [8] and it is zero if and only if the two probability mass functions are equal, it follows that $\beta_k = 1/K$, that is, the PFS scheme equally shares the channel between all users irrespective or their path loss. It follows also that $\tilde{T}_k = \bar{T}_k$ for all k. This concludes the proof.

Finally, we remark that PFS adaptively and automatically performs the following operation: it symmetrizes the channel gain distribution of all users by disregarding the fixed path loss (near–far effect) and by looking only at the ergodic fading component that, under our assumptions, is identically distributed across the users. Each user is served at its own peak channel gain.

APPENDIX D
PROOF OF THEOREM 2

We express Γ in nats, as it simplifies notation. Moreover, we shall consider bounded fading [2], i.e., such that $d_k^m \in [0, D]$, for some constant $D < \infty$ independent of K, with probability 1. Let $R_k^m = \frac{1}{K}\Gamma\nu_k^m$, for all $k = 1, \dots, K$ and $m = 1, \dots, M$, be a partial rate allocation. Using (20) in (21), we can write the minimization of $(E_b/N_0)_{\rm sys}$ as

minimize

$$\frac{1}{\Gamma}\sum_{k=1}^{K}\sum_{m=1}^{M}\frac{1}{d^m_{\pi^m_k}}\exp\left(\frac{\Gamma}{K}\sum_{i<k}\nu^m_{\pi^m_i}\right)\left(\exp\left(\frac{\Gamma}{K}\nu^m_{\pi^m_k}\right)-1\right) \qquad (68)$$

subject to

$$\sum_{m=1}^{M}\nu_k^m = \nu_k, \quad k = 1, \dots, K$$

and to the nonnegativity constraints $\nu_k^m \geq 0$ for all k, m. For large K

$$\left(\exp\left(\frac{\Gamma}{K}\nu^m_{\pi^m_k}\right)-1\right) = \frac{\Gamma}{K}\nu^m_{\pi^m_k} + o(1/K)$$

Therefore, the objective function for large K can be written as

$$\frac{1}{K}\sum_{k=1}^{K}\sum_{m=1}^{M}\frac{\nu^m_{\pi^m_k}}{d^m_{\pi^m_k}}\exp\left(\frac{\Gamma}{K}\sum_{i<k}\nu^m_{\pi^m_i}\right) \qquad (69)$$

with associated Lagrangian function

$$\mathcal{L} = \frac{1}{K}\sum_{k=1}^{K}\sum_{m=1}^{M}\frac{\nu^m_{\pi^m_k}}{d^m_{\pi^m_k}}\exp\left(\frac{\Gamma}{K}\sum_{i<k}\nu^m_{\pi^m_i}\right) - \sum_{k=1}^{K}\lambda_k^{(K)}\left(\sum_{m=1}^{M}\nu_k^m - \nu_k\right). \qquad (70)$$

[3]Notice that this upper bound holds for any stationary scheduling policy with fixed decision regions.

Notice that for each K we have an optimization problem, and we denote by $\lambda_k^{(K)}$ the kth Lagrange multiplier of the problem with size K users.

By differentiating with respect to ν_k^m, and by letting $\pi_i^m = k$, i.e., user k is ranked in the ith position on channel m, we obtain the Kuhn–Tucker conditions in the form

$$\frac{1}{d_k^m}\exp\left(\frac{\Gamma}{K}\sum_{\ell<i}\nu^m_{\pi^m_\ell}\right) + \frac{\Gamma}{K}\sum_{\ell>i}\frac{\nu^m_{\pi^m_\ell}}{d^m_{\pi^m_\ell}}\exp\left(\frac{\Gamma}{K}\sum_{\ell'<\ell}\nu^m_{\pi^m_{\ell'}}\right) \geq \lambda_k^{(K)} \qquad (71)$$

where we have multiplied both sides by K and have replaced $K\lambda_k^{(K)}$ by $\lambda_k^{(K)}$.

We guess the following solution: let $m_k = \arg\max_\ell\{d_k^\ell\}$ denote the index of the subchannel for which user k has maximum gain. Then, we let

$$\begin{aligned}\nu_k^m &= \nu_k, \quad \text{for } m = m_k\\ \nu_k^m &= 0, \quad \text{for } m \neq m_k.\end{aligned} \qquad (72)$$

Next, we substitute this solution in the Kuhn–Tucker conditions (71) and show that they are satisfied in the limit of $K \to \infty$. Suppose that the value $d_k^m = x$ is ranked in the ith position by the permutation π^m. Then, define $\mathcal{K}(m, x)$ as the set of users k such that $m_k = m$ (i.e., having maximum gain on channel m) such that their maximum channel gain is not larger than x. For the allocation (72) we have

$$\frac{1}{K}\sum_{\ell<i}\nu^m_{\pi^m_\ell} = \frac{1}{K}\sum_{k'\in\mathcal{K}(m,x)}\nu_{k'} \to \frac{1}{M}F_{\max}(x) \qquad (73)$$

where the last line follows from Lemma 1 in Appendix G. Furthermore, for $\ell > i$ we notice that the channel gain of a user ranked in position ℓ of permutation π^m must be larger than x. Defining $\mathcal{K}(m, (x, D])$ to be the set of users such that $m_k = m$ and their gain is in the interval $(x, D]$, we have

$$\begin{aligned}\frac{1}{K}\sum_{\ell>i}\frac{\nu^m_{\pi^m_\ell}}{d^m_{\pi^m_\ell}}\exp\left(\frac{\Gamma}{K}\sum_{\ell'<\ell}\nu^m_{\pi^m_{\ell'}}\right) &= \frac{1}{K}\sum_{k'\in\mathcal{K}(m,(x,D])}\frac{\nu_{k'}}{d^m_{k'}}\exp\left(\frac{\Gamma}{K}\sum_{k''\in\mathcal{K}(m,d^m_{k'})}\nu_{k''}\right)\\ &\to \frac{1}{M}\int_x^D \exp\left(\frac{\Gamma}{M}F_{\max}(z)\right)\frac{\mathrm{d}F_{\max}(z)}{z}\end{aligned} \qquad (74)$$

where again the last line follows from Lemma 1.

Using (73) and (74), we can write (71) in the limit for large K as

$$\frac{1}{x}\exp\left(\frac{\Gamma}{M}F_{\max}(x)\right) + \frac{\Gamma}{M}\int_x^D \exp\left(\frac{\Gamma}{M}F_{\max}(z)\right)\frac{\mathrm{d}F_{\max}(z)}{z} \geq \lambda_k \qquad (75)$$

for $d_k^m = x$, where λ_k denotes the limit of the Lagrange multiplier $\lambda_k^{(K)}$ for $K \to \infty$. The right-hand side (RHS) of (75) is seen as a function $g(x)$ of x. The guessed solution satisfies asymptotically the Kuhn–Tucker conditions if, for all k,, $g(x) = \lambda_k$ for $x = d_k^{m_k}$, i.e., for x equal to the maximum channel gain of user k, and $g(x) \geq \lambda_k$ for $x = d_k^m, m \neq m_k$. This condition holds if $g(x)$ is a nonincreasing function of x. The first derivative of $g(x)$ is given by

$$\frac{dg(x)}{dx} = -\frac{1}{x^2}\exp\left(\frac{\Gamma}{M}F_{\max}(x)\right) \tag{76}$$

which is negative for all $x \in [0, D]$. We conclude that the solution (72), that allocates each user k on its own best channel only, is asymptotically optimal for large K.

Finally, substituting this solution in (69), we define $\mathcal{K}(m)$ as the set of users k such that $m_k = m$, and $\mathcal{K}(m, x)$ as before and, using repeatedly Lemma 1 in Appendix G, we obtain

$$\begin{aligned} &\frac{1}{K}\sum_{k=1}^{K}\sum_{m=1}^{M}\frac{\nu_{\pi_k^m}^m}{d_{\pi_k^m}^m}\exp\left(\frac{\Gamma}{K}\sum_{i<k}\nu_{\pi_i^m}^m\right) \\ &= \sum_{m=1}^{M}\frac{1}{K}\sum_{k\in\mathcal{K}(m)}\frac{\nu_k}{d_k^m}\exp\left(\frac{\Gamma}{K}\sum_{k'\in\mathcal{K}(m,d_k^m)}\nu_{k'}\right) \\ &\to \int_0^D \frac{1}{x}\exp\left(\frac{\Gamma}{M}F_{\max}(x)\right)\mathrm{d}F_{\max}(x) \end{aligned} \tag{77}$$

where the last line follows from the fact that the subchannels are statistically symmetric, therefore, the factor $1/M$ of Lemma 1 cancels with the sum with respect to m of M equal terms. This proves Theorem 2 for bounded channel gains. Finally, the proof can be extended to channel gains having a distribution with unbounded support using standard arguments and under mild regularity conditions on the channel gain distribution. In particular, for channel gains having a continuous probability density function (pdf) with bounded moments the extension is straightforward.

APPENDIX E
PROOF OF THEOREM 3

Again, for simplicity we express Γ in nats and consider bounded fading in $[0, D]$, for some constant $D < \infty$ independent of K, with probability 1. It is convenient to introduce the new variables $r_k^m = R_k^m/\Theta_k^m$. The minimization of system E_b/N_0 is expressed as

$$\min_{\boldsymbol{\Theta}^m, \boldsymbol{r}^m, m=1,\dots,M} \frac{1}{\Gamma}\sum_{k=1}^{K}\sum_{m=1}^{M}\frac{\Theta_k^m}{d_k^m}\left(\exp(r_k^m) - 1\right) \tag{78}$$

subject to

$$\sum_{m=1}^{M}\Theta_k^m r_k^m = R_k, \quad k = 1,\dots,K$$

and to

$$\sum_{k=1}^{K}\Theta_k^m \leq 1, \quad m = 1,\dots,M$$

and under the nonnegativity constraints $\Theta_k^m \geq 0, r_k^m \geq 0$. It is clear that the solution of (78) is obtained when the constraints on the fractions Θ_k^m are satisfied with equality (otherwise, one would not use some channel dimensions without any benefit). Therefore, the only inequality constraints are the nonnegativity constraints. The Lagrangian of the problem (78) is given by

$$\begin{aligned} \mathcal{L} = &\frac{1}{\Gamma}\sum_{k=1}^{K}\sum_{m=1}^{M}\frac{\Theta_k^m}{d_k^m}\left(\exp(r_k^m) - 1\right) \\ &-\sum_{k=1}^{K}\lambda_k^{(K)}\left[\sum_{m=1}^{M}\Theta_k^m r_k^m - R_k\right] + \sum_{m=1}^{M}\mu^m\left[\sum_{k=1}^{K}\Theta_k^m - 1\right] \end{aligned} \tag{79}$$

Again, we denote by $\lambda_k^{(K)}$ the kth Lagrange multiplier of the problem with size K users. By differentiating with respect to Θ_k^m and with respect to r_k^m we obtain the Kuhn–Tucker conditions

$$\frac{\partial\mathcal{L}}{\partial\Theta_k^m} = \frac{1}{\Gamma d_k^m}\left(\exp(r_k^m) - 1\right) - \lambda_k^{(K)} r_k^m + \mu^m \geq 0 \tag{80}$$

and

$$\frac{\partial\mathcal{L}}{\partial r_k^m} = \frac{\Theta_k^m}{\Gamma d_k^m}\exp(r_k^m) - \lambda_k^{(K)}\Theta_k^m \geq 0 \tag{81}$$

We guess the following solution: for every k, let $m_k = \arg\max_\ell\{d_k^\ell\}$. Then, we let

$$\begin{aligned} \Theta_k^{m_k} &> 0 \\ r_k^{m_k} &= R_k/\Theta_k^{m_k} \end{aligned} \tag{82}$$

and $\Theta_k^m = r_k^m = 0$ for all $m \neq m_k$. Next, we substitute this solution in the Kuhn–Tucker conditions (71) and show that they are satisfied in the limit of $K \to \infty$.

For each k and $m = m_k$, the Kuhn–Tucker conditions (80) and (81) must hold with equality. From (81) we get

$$\frac{1}{\Gamma d_k^{m_k}}\exp(r_k^{m_k}) = \lambda_k^{(K)}. \tag{83}$$

Substituting (83) into (80), after some algebra, we obtain

$$r_k^{m_k} = 1 + W\left(\frac{\mu^{m_k}\Gamma d_k^{m_k} - 1}{e}\right) \tag{84}$$

where $W(x)$ is Lambert's W function.

If solving for the Lagrange multipliers μ^m we find $\mu^m = \mu > 0$ for all $m = 1,\dots,M$, then the guessed solution is seen to satisfy the Kuhn–Tucker conditions. In fact, for $m = m_k$ both (80) and (81) hold with equality, while for $m \neq m_k$, (81) holds with equality (since $\Theta_k^m = 0$) and (80) holds with inequality, since from (84) we have

$$\frac{\mu\Gamma d_k^{m_k} - 1}{\mathrm{e}} = \left(R_k/\Theta_k^{m_k} - 1\right)\exp\left(R_k/\Theta_k^{m_k} - 1\right)$$

which implies $\mu > 0$. Hence, we shall solve for the Lagrange multipliers and show that as $K \to \infty$, the solution has the form $\mu^m = \mu$, independent of m.

By using $\Theta_k^{m_k} r_k^{m_k} = R_k$ and (84) and by summing over k we obtain, for all $m = 1, \ldots, M$

$$\sum_{k \in \mathcal{K}(m)} \frac{R_k}{1 + W\left(\frac{\mu^m \Gamma d_k^m - 1}{e}\right)} = 1 \tag{85}$$

where $\mathcal{K}(m)$ is the subset of users such that $m_k = m$, i.e., the users that share subchannel m. Letting $R_k = \frac{\Gamma}{K}\nu_k$, we have

$$\frac{1}{K} \sum_{k \in \mathcal{K}(m)} \frac{\Gamma \nu_k}{1 + W\left(\frac{\mu^m \Gamma d_k^m - 1}{e}\right)} = 1. \tag{86}$$

Direct application of Lemma 1 of Appendix G to (86) yields that the Lagrange multiplier equations for all m in the limit of $K \to \infty$ converge to

$$\int_0^D \frac{\mathrm{d}F_{\max}(x)}{1 + W\left(\frac{\mu^m \Gamma x - 1}{e}\right)} = \frac{M}{\Gamma}. \tag{87}$$

Since this equation is the same for all m, it follows that the solution of the Lagrange multipliers is $\mu^m = \mu$ for all m, in the limit of $K \to \infty$. For what has been said before, this implies that the guessed solution satisfies the Kuhn–Tucker conditions in the limit of large K and hence it is asymptotically optimal.

Using the asymptotic solution in the objective function of (78), we obtain

$$\frac{1}{\Gamma} \sum_{m=1}^{M} \sum_{k \in \mathcal{K}(m)} \frac{\frac{\Gamma \nu_k}{K}\left(\exp\left(1 + W\left(\frac{\mu \Gamma d_k^m - 1}{\mathrm{e}}\right)\right) - 1\right)}{d_k^m \left(1 + W\left(\frac{\mu \Gamma d_k^m - 1}{\mathrm{e}}\right)\right)}$$
$$\to \int_0^D \frac{\exp\left(1 + W\left(\frac{\mu \Gamma x - 1}{\mathrm{e}}\right)\right) - 1}{1 + W\left(\frac{\mu \Gamma x - 1}{\mathrm{e}}\right)} \frac{\mathrm{d}F_{\max}(x)}{x}. \tag{88}$$

The final expression of Theorem 3 is obtained by replacing $\mu\Gamma$ by μ (without loss of generality) and extending the result to channel gain distributions with unbounded support, under mild regularity conditions.

Appendix E
Proof of Theorem 4

Theorem 4 is an immediate application of Lemma 1 in Appendix G. Let $\mathcal{K}(m)$ denote the subset of users such that $m_k = m$, i.e., the users that share subchannel m. By definition of the conventional TDMA/FDMA scheme, the fraction Θ_k^m of channel uses for user $k \in \mathcal{K}(m)$ is given by

$$\Theta_k^m = \frac{\nu_k}{\sum_{k' \in \mathcal{K}(m)} \nu_{k'}}. \tag{89}$$

Therefore, its energy allocation is

$$E_k^m = \frac{\nu_k N_0}{d_k^m \sum_{k' \in \mathcal{K}(m)} \nu_{k'}} \left(\exp\left(\frac{\Gamma}{K} \sum_{k' \in \mathcal{K}(m)} \nu_{k'}\right) - 1\right) \tag{90}$$

for $m = m_k$, and zero for $m \neq m_k$.

Using this in (21) with Γ expressed in bits and applying Lemma 1 we get

$$\left(\frac{E_b}{N_0}\right)_{\text{sys}} = \frac{1}{\Gamma} \sum_{m=1}^{M} \frac{1}{K} \sum_{k \in \mathcal{K}(m)} \frac{\nu_k / d_k^m}{\frac{1}{K} \sum_{k' \in \mathcal{K}(m)} \nu_{k'}} \times \left(2^{\frac{\Gamma}{K} \sum_{k' \in \mathcal{K}(m)} \nu_{k'}} - 1\right)$$
$$\to \frac{2^{\Gamma/M} - 1}{\Gamma/M} \int_0^\infty \frac{\mathrm{d}F_{\max}(x)}{x}. \tag{91}$$

This concludes the proof.

Appendix G
Lemma 1

Let $\mathcal{A} \subseteq [0, D]$ be an interval, and let $\mathcal{A}(m)$ denote the set of all k such that $\max\{d_k^1, \ldots, d_k^M\} \in \mathcal{A}$ and $m_k = m$. Also, let $g(\cdot)$ denote a continuous measurable function in $(0, D]$. Under Assumptions A2, A3, and A4

$$\frac{1}{K} \sum_{k \in \mathcal{A}(m)} g(d_k^m)\nu_k \to \frac{1}{M} \int_{\mathcal{A}} g(x) \mathrm{d}F_{\max}(x) \tag{92}$$

with probability 1, as $K \to \infty$ and M is fixed.

Proof: It follows immediately from the convergence of the empirical cdf of $\{(d_k^1, \ldots, d_k^M, \nu_k)\}_{k=1}^K$ that, as $K \to \infty$, we have

$$\frac{1}{K} \sum_{k \in \mathcal{A}(m)} g(d_k^m)\nu_k = \frac{|\mathcal{A}(m)|}{K} \frac{1}{|\mathcal{A}(m)|} \sum_{k \in \mathcal{A}(m)} g(d_k^m)\nu_k$$
$$\to \Pr(k \in \mathcal{A}(m)) \, \mathbb{E}\left[g\left(d_k^m\right)\nu_k \,|\, k \in \mathcal{A}(m)\right]. \tag{93}$$

We have

$$\begin{aligned} &\Pr(k \in \mathcal{A}(m)) \\ &\quad = \Pr\left(\max\{d_k^1, \ldots, d_k^M\} \in \mathcal{A} \,|\, m_k = m\right) \Pr(m_k = m) \\ &\quad = \frac{1}{M} \int_{\mathcal{A}} \mathrm{d}F_{\max}(x). \end{aligned} \tag{94}$$

This is easily seen by observing that, by symmetry of F, the maximum of the vector $d_k^1, \ldots, d_k^M$ can occur at any position m equally likely. Thus, $\Pr(m_k = m) = 1/M$. Moreover, conditioning on $m_k = m$ has no effect on the distribution of $\max\{d_k^1, \ldots, d_k^M\}$. Thus

$$\Pr(\max\{d_k^1, \ldots, d_k^M\} \in \mathcal{A} | m_k = m) = \int_{\mathcal{A}} \mathrm{d}F_{\max}(x).$$

By Assumption A4, since the joint cdf of $(d_k^1, \ldots, d_k^M)$ and ν_k is the product cdf FG and by Assumption A3 G has mean 1, then

$$\begin{aligned} &\mathbb{E}[g(d_k^m)\nu_k | k \in \mathcal{A}(m)] \\ &\quad = \mathbb{E}\left[g\left(d_k^m\right) \,|\, k \in \mathcal{A}(m)\right] \mathbb{E}[\nu_k | k \in \mathcal{A}(m)] \\ &\quad = \mathbb{E}\left[g\left(d_k^m\right) \,|\, k \in \mathcal{A}(m)\right] \mathbb{E}[\nu_k] \\ &\quad = \mathbb{E}\left[g\left(d_k^m\right) \,|\, k \in \mathcal{A}(m)\right]. \end{aligned} \tag{95}$$

Finally, from the preceding discussion we have that

$$\begin{aligned}&\mathbb{E}\left[g\left(d_k^m\right) \mid k \in \mathcal{A}(m)\right]\\&\quad=\mathbb{E}\left[g\left(\max\{d_k^1,\ldots,d_k^M\}\right) \mid \max\{d_k^1,\ldots,d_k^M\} \in \mathcal{A}\right]\\&\quad=\frac{1}{\int_{\mathcal{A}} \mathrm{d}F_{\max}(x)}\int_{\mathcal{A}} g(x)\mathrm{d}F_{\max}(x) \qquad (96)\end{aligned}$$

By using (94)–(96) in (93) the result follows. □

Appendix H
Proof of Theorem 5

Let us start by recalling some results and notations in [9]. For a complex baseband channel with white additive noise with power spectral density N_0 and subject to an input power constraint, let $\mathcal{C}(\mathsf{SNR})$ denote the capacity–cost function where, without loss of generality, the input cost is defined in terms of the channel SNR, as defined in [9], and let $\mathsf{C}(E_b/N_0)$ denote the spectral efficiency (in bit/dimension, or bit/s/Hz) as a function of E_b/N_0. In general, there exists a value $(E_b/N_0)_{\min} \geq 0$ such that $\mathsf{C}(E_b/N_0)=0$ for all $(E_b/N_0) \leq (E_b/N_0)_{\min}$ and $\mathsf{C}(E_b/N_0)>0$ for all $(E_b/N_0) > (E_b/N_0)_{\min}$. The derivative of C with respect to E_b/N_0 expressed in decibels, evaluated at $(E_b/N_0)_{\min}$ and normalized by 3 dB, is called the *wideband slope* and it is usually indicated by $\mathcal{S}_0$. The engineering significance of $\mathcal{S}_0$ is discussed in depth in [9]. In particular, it is shown that systems with the same $(E_b/N_0)_{\min}$ might indeed have very different behaviors in the wide band (i.e., very large but not infinite bandwidth) regime, and this behavior is captured by $\mathcal{S}_0$.

Theorem 9 in [9] provides a nice way to compute the wideband slope in terms of the channel capacity–cost function (expressed in nats)

$$\mathcal{S}_0=\frac{2(\mathcal{C}'(0))^2}{-\mathcal{C}''(0)} \qquad (97)$$

where $\mathcal{C}'(0)$ and $\mathcal{C}''(0)$ denote the first and second derivatives of $\mathcal{C}(\mathsf{SNR})$ for $\mathsf{SNR} \downarrow 0$.

In our case, Theorems 2–4 yield directly $(E_b/N_0)_{\rm sys}$ in terms of C. By using the fact that, by definition, $\mathsf{SNR}=(E_b/N_0)_{\rm sys}\mathsf{C}$, it follows that our capacity–cost function $\mathcal{C}$ is given by the implicit parameterization in terms of C given by

$$\begin{aligned}\mathsf{SNR}&=\mathsf{C}f(\mathsf{C})\\\mathcal{C}(\mathsf{SNR})&=\mathsf{C}\log(2) \qquad (98)\end{aligned}$$

where, for notational simplicity, we indicate briefly by $f(\mathsf{C})$ the functions that yield $(E_b/N_0)_{\rm sys}$ in terms of C, given by (34), (35), and (37).

From basic calculus, it follows that

$$\begin{aligned}\mathcal{C}'(\mathsf{SNR})&=\log(2)\frac{\mathrm{d}\mathsf{C}}{\mathrm{d}\mathsf{SNR}}\\&=\frac{\log(2)}{\frac{\mathrm{d}\mathsf{SNR}}{\mathrm{d}\mathsf{C}}}\\&=\frac{\log(2)}{f(\mathsf{C})+\mathsf{C}f'(\mathsf{C}).} \qquad (99)\end{aligned}$$

Letting $\mathsf{SNR} \downarrow 0$ is equivalent to letting $\mathsf{C} \downarrow 0$; therefore, we obtain

$$\mathcal{C}'(0)=\frac{\log(2)}{f(0)}.$$

Similarly, for the second derivative we have

$$\mathcal{C}''(\mathsf{SNR})=\frac{-(\log(2))^2[2f'(\mathsf{C})+\mathsf{C}f''(\mathsf{C})]}{[f(\mathsf{C})+\mathsf{C}f'(\mathsf{C})]^3}. \qquad (100)$$

Letting again $\mathsf{C} \downarrow 0$, we find

$$\mathcal{C}''(0)=\frac{-2\log(2)f'(0)}{f(0)^3}.$$

By using these expressions in (97), we obtain

$$\mathcal{S}_0=\log(2)\frac{f(0)}{f'(0)}. \qquad (101)$$

The results for the optimal delay-limited system (Theorem 2) and for the conventional TDMA/FDMA system (Theorem 4) follow immediately. Details are omitted as they are straightforward.

The case of the optimal orthogonal system (Theorem 3) is a bit more complicated since the function f is given implicitly, parameterized by the Lagrange multiplier μ (see (35) and (36)). Since as $\mathsf{C} \downarrow 0$ then $\mu \downarrow 0$, the computation of $f(0)$ is straightforward and yields

$$f(0)=\log(2)\int\frac{\mathrm{dF_{max}(x)}}{x}. \qquad (102)$$

In fact, is should be noticed that $f(0)$ coincides by definition with $(E_b/N_0)_{\rm sys,min}$. In order to obtain $f'(0)$, we can write

$$f'(\mathsf{C})=\frac{\log(2)}{e}\int\frac{W(\exp(1+W)+1)W'\mu'}{(1+W)^2}dF_{\max}(x) \qquad (103)$$

where we use the shorthand notation W to indicate $W(\frac{\mu x-1}{e})$ and where W' denotes the first derivative of Lambert's W function and μ' denotes the first derivative of the Lagrange multiplier with respect to C. For small C, which implies small μ, we have the approximation which becomes tight as $\mathsf{C} \to 0$

$$W(z) \approx -1+\sqrt{2(ez+1)}.$$

Using $z=(\mu x-1)/e$ in the above, we obtain

$$W \approx -1+\sqrt{2\mu x}$$

and

$$W' \approx \frac{e}{\sqrt{2\mu x}}.$$

Using the approximation for W in (36), we obtain

$$\mu=\left(\log(2)\int\frac{\mathrm{d}F_{\max}(x)}{\sqrt{2x}}\right)^2\mathsf{C}^2 \qquad (104)$$

and eventually

$$\mu'=2\left(\log(2)\int\frac{\mathrm{d}F_{\max}(x)}{\sqrt{2x}}\right)^2\mathsf{C}. \qquad (105)$$

By collecting all these expressions and using them in (103), it is not difficult to see that

$$\lim_{\mathsf{C}\downarrow 0} f'(\mathsf{C}) = \left(\log(2)\int \frac{\mathrm{d}F_{\max}(x)}{\sqrt{2x}}\right)^2.$$

This, together with (102), used in (101) yields the desired result.

Appendix I
Proof of Theorem 6

From the definition (39) and using the fact that $\mathcal{S}_\infty = 1$, we have

$$\mathcal{L}_\infty = \frac{\mathsf{C}}{\log(2)}\left[\frac{1}{\mathsf{C}}\log(\mathsf{C}f(\mathsf{C})) - \log(2) + o(1/\mathsf{C})\right], \quad \mathsf{C}\to\infty \tag{106}$$

where $f(\mathsf{C})$ denotes $(E_b/N_0)_{\mathrm{sys}}$ as a function of C for the delay-limited optimal signaling in the limit for $K\to\infty$. Using Theorem 2, we obtain

$$\frac{1}{\mathsf{C}}\log(\mathsf{C}f(\mathsf{C})) = \frac{\log\log(2)}{\mathsf{C}} + \frac{1}{\mathsf{C}}\log\left(\mathsf{C}\int_0^\infty \exp(\mathsf{C}\log(2)F_{\max}(x))\frac{\mathrm{d}F_{\max}(x)}{x}\right) \tag{107}$$

With the change of variable $y = \mathsf{C}F_{\max}(x)$, (107) can be written as

$$\frac{1}{\mathsf{C}}\log(\mathsf{C}f(\mathsf{C})) = O(1/\mathsf{C}) + \frac{1}{\mathsf{C}}\log\left(\int_0^{\mathsf{C}} \exp(\log(2)y)\frac{\mathrm{d}y}{F_{\max}^{-1}(y/\mathsf{C})}\right) \tag{108}$$

The integral inside the logarithm can be approximated by the Riemann sum

$$\frac{\mathsf{C}}{N}\sum_{n=0}^{N-1}\frac{\exp(\log(2)n\mathsf{C}/N)}{F_{\max}^{-1}(n/N)}$$

for a sufficiently large integer N. Since we are interested in the limit for large C, we can take $N = \lceil\mathsf{C}\rceil$ and obtain a sufficiently accurate approximation, in the sense that the integral differs from the Riemann sum by no more than a constant term $O(1)$.

The term that exponentially dominates the sum inside the logarithm is for $n = \lceil\mathsf{C}\rceil - 1$. Thus, we obtain

$$\frac{1}{\mathsf{C}}\log(\mathsf{C}f(\mathsf{C})) = O(1/\mathsf{C}) + \log(2)\left(1 - \frac{1}{\lceil\mathsf{C}\rceil}\right) - \frac{1}{\mathsf{C}}\log F_{\max}^{-1}\left(1 - \frac{1}{\lceil\mathsf{C}\rceil}\right). \tag{109}$$

Using (109) in (106) we obtain (in the limit of large C)

$$\mathcal{L}_\infty = \frac{\mathsf{C}}{\log(2)}\left[O(1/\mathsf{C}) - \frac{1}{\mathsf{C}}\log F_{\max}^{-1}\left(1-\frac{1}{\mathsf{C}}\right)\right] = -\log_2\left(F_{\max}^{-1}\left(1-\frac{1}{\mathsf{C}}\right)\right) + O(1). \tag{110}$$

This concludes the proof.

Appendix J
Proof of Theorem 7

The difficulty in proving Theorem 7 lies in the fact that in this case C versus $(E_b/N_0)_{\mathrm{sys}}$ are given in parametric form as functions of the Lagrange multiplier μ, in Theorem 3.

The Lagrange multiplier is defined by the constraint equation

$$\frac{1}{\mathsf{C}\log 2} = \int \frac{\mathrm{d}F_{\max}(x)}{1 + W\left(\frac{\mu x - 1}{\mathrm{e}}\right)}. \tag{111}$$

For large C, the LHS of (111) goes to zero. Since Lambert's W function is increasing and not smaller than -1, the RHS can only become small as the argument of Lambert's W function approaches infinity. Thus, we conclude that $\mu\to\infty$ as $\mathsf{C}\to\infty$.

For large argument, Lambert's W function can be approximated as [29]

$$W(z) = \log(z) - \log\log(z) + O\left(\frac{\log\log(z)}{\log(z)}\right). \tag{112}$$

This yields

$$1 + W\left(\frac{\mu x - 1}{\mathrm{e}}\right) = \log(\mu x - 1) - \log(\log(\mu x - 1) - 1) + O\left(\frac{\log\log(\mu)}{\log(\mu)}\right). \tag{113}$$

Noticing that

$$\log(\mu x - 1) = \log(\mu) + \log(x) + O(\mu^{-1}) \tag{114}$$

$$\log(\log(\mu x - 1) - 1) = \log\log(\mu) + O\left(\frac{1}{\log(\mu)}\right) \tag{115}$$

we find

$$1 + W\left(\frac{\mu x - 1}{\mathrm{e}}\right) = \log(\mu) - \log\log(\mu) + \log(x) + O\left(\frac{\log\log(\mu)}{\log(\mu)}\right). \tag{116}$$

We shall use repeatedly the following series expansion for the reciprocal function. Let $g(z) = a(z) - b(z)$ such that $b(z)/a(z)\to 0$ as $z\to\infty$. Hence, for sufficiently large z

$$\frac{1}{g(z)} = \frac{1}{a(z)}\sum_{i=0}^\infty\left(\frac{b(z)}{a(z)}\right)^i. \tag{117}$$

Applying this to (116) we obtain

$$\frac{1}{1 + W\left(\frac{\mu x - 1}{\mathrm{e}}\right)} = \frac{1}{\log(\mu)} + \frac{\log\log(\mu) - \log(x)}{\log^2(\mu)} + O\left(\frac{\log^2\log(\mu)}{\log^3(\mu)}\right) \tag{118}$$

and, integrating with respect to the fading distribution, we get

$$\int \frac{\mathrm{d}F_{\max}(x)}{1 + W\left(\frac{\mu x - 1}{\mathrm{e}}\right)} = \frac{1}{\log(\mu)} + \frac{\log\log(\mu) - \int\log(x)\mathrm{d}F(x)}{\log^2(\mu)} + O\left(\frac{\log^2\log(\mu)}{\log^3(\mu)}\right). \tag{119}$$

Using again the fact (117), C as a function of μ is given as

$$\mathsf{C}\log(2) = \log(\mu) - \log\log(\mu) + \int \log(x)\mathrm{d}F_{\max}(x) + O\left(\frac{\log\log(\mu)}{\log(\mu)}\right). \quad (120)$$

Note that even constant terms in the integral cannot be neglected against $\log(\mu)$ even for large μ, as C will later appear in the exponent and additive terms become factors to 2^{C}. Only terms that actually go to zero disappear, as they will be converted to unit factors.

Now, we turn to the calculation of the system E_b/N_0

$$\left(\frac{E_b}{N_0}\right)_{\text{sys}} = \log(2)\int \frac{\exp\left(1+W\left(\frac{\mu x-1}{\mathrm{e}}\right)\right)-1}{1+W\left(\frac{\mu x-1}{\mathrm{e}}\right)}\frac{\mathrm{d}F_{\max}(x)}{x}. \quad (121)$$

With the series expansion (116), for fading distributions with $\int x^{-1}\mathrm{d}F(x) < \infty$, we find

$$\frac{\exp\left(\left(1+W\left(\frac{\mu x-1}{\mathrm{e}}\right)\right)-1\right.}{x} = \frac{\mu}{\log(\mu)}O\left(e^{\frac{\log\log(\mu)}{\log(\mu)}}\right) - \frac{1}{x} \quad (122)$$

$$= \frac{\mu}{\log(\mu)} + O(1). \quad (123)$$

This gives

$$\left(\frac{E_b}{N_0}\right)_{\text{sys}} = \log(2)\int \frac{\mathrm{d}F_{\max}(x)}{1+W\left(\frac{\mu x-1}{e}\right)}\left(\frac{\mu}{\log(\mu)} + O(1)\right) \quad (124)$$

$$= \frac{1}{\mathsf{C}}\frac{\mu}{\log(\mu)} + O\left(\mathsf{C}^{-1}\right). \quad (125)$$

Using (120) we have that

$$2^{\mathsf{C}} = \frac{\mu}{\log(\mu)}e^{\int \log(x)\mathrm{d}F(x)}O\left(e^{\frac{\log\log(\mu)}{\log(\mu)}}\right) \quad (126)$$

and, finally, we obtain

$$\left(\frac{E_b}{N_0}\right)_{\text{sys}} = \frac{2^{\mathsf{C}}}{\mathsf{C}}2^{-\int \log_2(x)\mathrm{d}F_{\max}(x)}O\left(e^{\frac{-\log\log(\mu)}{\log(\mu)}}\right) + O\left(\mathsf{C}^{-1}\right)$$

$$\left.\left(\frac{E_b}{N_0}\right)_{\text{sys}}\right|_{\text{dB}} = \mathsf{C}10\log_{10}(2) - 10\log_{10}(\mathsf{C}) - \int \log_2(x)\mathrm{d}F_{\max}(x)10\log_{10}(2) + o(1). \quad (127)$$

Comparing the above expression with (39), and taking into account that we have already established that $\mathcal{S}_\infty = 1$, the theorem is proved.

REFERENCES

[1] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information—Theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[2] S. V. Hanly and D. N. C. Tse, "Multi-access fading channels: Part II: Delay-limited capacities," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2816–2831, Nov. 1998.

[3] D. N. C. Tse and S. V. Hanly, "Multi-access fading channels: Part I: Polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2796–2815, Nov. 1998.

[4] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. IEEE In. Conf. Communications (ICC)*, Seattle, WA, Jun. 1995, pp. 331–335.

[5] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1277–1294, Jun. 2002.

[6] *CDMA2000 High Rate Packet DataAir Interface Specification*, TIA/EIA/3GPP2 Std. IS-856/3GPP2 C.S.0024, Dec. 2001, v3.0.

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

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[9] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, Jun. 2002.

[10] R. R. Müller, *Power and Bandwidth Efficiency of Multiuser Systems with Random Spreading*. Aachen, Germany: Shaker-Verlag, 1999.

[11] S. Shamai (Shitz) and S. Verdú, "The impact of frequency-flat fading on the spectral efficiency of CDMA," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1302–1327, May 2001.

[12] G. Caire, R. R. Müller, and T. Tanaka, "Iterative multiuser joint decoding: Optimal power allocation and low-complexity implementation," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1950–1973, Sep. 2004.

[13] A. Roumy, S. Guemghar, G. Caire, and S. Verdú, "Maximizing the spectral efficiency of coded CDMA under successive decoding," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 152–164, Jan. 2004.

[14] N. Jindal, S. Vishwanath, and A. Goldsmith, "On the duality of Gaussian multiple-access and broadcast channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 768–783, May 2004.

[15] R. S. Cheng and S. Verdú, "Gaussian multiaccess channels with ISI: Capacity region and multiuser water-filling," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 773–785, May 1993.

[16] D. Bertsekas, *Non-Linear Programming: 2nd Ed.*. Belmont, MA: Athena Scientific, 1999.

[17] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert W function," *Adv. Comput. Math.*, vol. 5, pp. 329–359, 1996.

[18] D. N. C. Tse, "Optimal power allocation over parallel Gaussian broadcast channels," in *Proc. IEEE Int. Symp. Information Theory*, Ulm, Germany, Jun./Jul. 1997, p. 27.

[19] L. Li and A. J. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels: Part I: Ergodic capacity and part II: Outage capacity," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 1083–1127, Mar. 2001.

[20] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2001.

[21] S. Borst and P. Whiting, Dynamic rate control algorithms for HDR throughput optimization, Bell Labs., Lucent Technologies, 2000, Tech. Memo. 10009626000828-21TM.

[22] V. K. Lau, "Proportional fair space-time scheduling for wireless communications," *IEEE Trans. Commun.*, vol. 53, no. 8, pp. 1353–1360, Aug. 2005.

[23] M. Sharif and B. Hassibi, "On the capacity of a MIMO broadcast channel with partial side information," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 506–522, Feb. 2005.

[24] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed. New York: Springer-Verlag, 1998.

[25] R. Knopp, "Coding and Multiple-Access over Fading Channels," Ph.D. dissertation, EPFL, Lausanne, Switzerland, May 1997.

[26] Federal Communications Commission home page [Online]. Available: http://www.fcc.gov/oet/cognitiveradio/

[27] W. C. Y. Lee, *Mobile Communications Design Fundamentals*. New York: Wiley, 1993.

[28] T. S. Rappaport, *Wireless Communications*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[29] R. M. Corless, D. J. Jeffrey, and D. E. Knuth, "A sequence of series for the Lambert W function," in *Proc. 1997 Int. Symp. Symbolic and Algebraic Computation*, Maui, HI, 1997, pp. 197–204.