Multiuser Diversity in Delay-Limited Cellular Wideband Systems

Giuseppe Caire
University of Southern California
Los Angeles, CA 90089
Email: caire@usc.edu

Ralf Müller
Norwegian Univ. of Science & Technology
7491 Trondheim, Norway
Email: ralf@iet.ntnu.no

Raymond Knopp
Eurecom Institute
Sophia-Antipolis, France
Email: raymond.knopp@eurecom.fr

***Abstract*— We consider the uplink and the downlink of a multiuser wireless system with one base station and K user terminals. Each user is affected by a propagation path loss, fixed in time, and by a slowly time-varying frequency-selective fading channel, modeled as M parallel block-fading channels. We investigate "multiuser diversity", i.e., the gain achieved by a multiuser system over a single-user system. In particular, we compare "delay-limited" systems with "delay-tolerant" systems based on Proportional Fair Scheduling. Our analysis allows us to quantify the loss incurred by a strict delay constraint in a cellular environment, for given M, K and channel statistics.**

I. BACKGROUND

A. Capacity region, power region and $(E_b/N_0)_{\rm sys}$

The K-user Gaussian multiple-access channel,

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

has capacity region given by [1]

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

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

If orthogonal signaling, e.g. by TDMA or FDMA, is used, the achievable rate region is given by

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

subject to $\sum\limits_{k=1}^{K} \Theta_k \leq 1$, 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.

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

$$\sum_{k\in\mathcal{S}} E_k^{(r)} \geq N_0\left[\exp\left(\sum_{k\in\mathcal{S}} R_k\right) - 1\right] \quad \forall\ \mathcal{S}\subseteq\{1,\ldots,K\} \tag{4}$$

for optimal signaling and

$$E_k^{(r)} \geq \Theta_k N_0\left[\exp\left(R_k/\Theta_k\right) - 1\right] \quad \forall k, \tag{5}$$

for orthogonal signaling. 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 $E_k^{(r)} = d_k E_k$, where d_k denotes the channel (power) gain of user k. This leads to a rescaling of the power region axes: the minimum received power for given user rates is achieved by a unique vertex of the power 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 strength of their channels [3], [4]. Constraining to orthogonal signaling implies an increase in total transmit power unless all channel gains are identical.

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

$$\begin{array}{c}\min_{\mathbf{E}\in\mathbb{R}_+^K}\ \sum_{k=1}^{K} E_k \\ \text{subject to}\ \ \mathbf{R}\in\mathcal{C}_{\rm MAC}(\mathbf{d};\mathbf{E}),\end{array} \tag{6}$$

where $\mathcal{C}_{\rm MAC}(\mathbf{d};\mathbf{E})$ 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 [4], the solution of (6) is found explicitly as

$$E_{\pi_k} = \frac{N_0}{d_{\pi_k}}\left[\exp\left(\sum_{i\leq k} R_{\pi_i}\right) - \exp\left(\sum_{i<k} R_{\pi_i}\right)\right] \tag{7}$$

where π is the permutation of $\{1,\ldots,K\}$ that sorts the channel gains in increasing order, i.e., $d_{\pi_1} \leq \cdots \leq d_{\pi_K}$ and the associated decoding order is given by $\pi_K, \pi_{K-1},\ldots,\pi_1$.

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

$$\begin{array}{c}\min_{\mathbf{\Theta},\mathbf{E}\in\mathbb{R}_+^K}\ \sum_{k=1}^{K} E_k \\ \text{subject to}\ \ \mathbf{R}\in\mathcal{C}_{\rm orth}(\mathbf{d};\mathbf{E};\mathbf{\Theta}),\end{array} \tag{8}$$

1-4244-0092-9/06/$20.00 ©2006 IEEE.

where $\mathcal{C}_{\rm orth}(\mathbf{d};\mathbf{E};\boldsymbol{\Theta})$ is the region defined by (3), 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 [4] is the set of all rate K-tuples $\mathbf{R}$ that can be attained for all $\mathbf{d} \in \mathbb{R}_+^K$, subject to *average power constraints* $\mathbb{E}[E_k] \leq \overline{E}_k$. In this work we are interested in the total system throughput (sum rate) versus the total average transmit energy. We define the system E_b/N_0 under a coding strategy that supports user rates $\mathbf{R} = (R_1, \ldots, R_K)$ with sum $\Gamma = \sum_{k=1}^K R_k$ subject to average transmit energy per symbol constraints $(\overline{E}_1, \ldots, \overline{E}_K)$ as

$$\left(\frac{E_b}{N_0}\right)_{\rm sys} = \frac{\sum_{k=1}^K \overline{E}_k}{N_0 \Gamma} \tag{9}$$

where Γ is expressed in bits.

Although this paper deal explicitly with the uplink, it follows from the recent result on the duality of the Gaussian multiple-access and broadcast channels [5], that for any set of user rates $\mathbf{R}$ the required $(E_b/N_0)_{\rm sys}$ is the same for uplink and downlink. Therefore, all results presented here apply verbatim to the downlink.

II. WIDEBAND CHANNELS

We model a wideband system as a set of M parallel block-fading channels, where the channel gain of each user may differ from channel to channel. The resulting parallel multple-access channel is given by

$$Y^m = \sum_{k=1}^K \sqrt{d_k^m} X_k^m + N^m, \qquad m = 1, \ldots, M \tag{10}$$

This is an accurate model for frequency-selectivity where m can be interpreted as the sub-band index. The system *spectral efficiency* is given by $\mathsf{C} = \frac{\Gamma}{M}$, and it is expressed in bit/s/Hz.

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. The path loss varies so slowly in time with respect to the signal bandwidth that 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 [6].

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. In the following, it will be clear that an important role is played by the cdf $F_{s\max\{f\},n}(x)$, of the random variable $s\max\{f^1, \ldots, f^n\}$ where s is the path loss of a random user in the system and f^i is distributed as any of the fading components f_k^m. For the case where the users are uniformly distributed in a unit disc centered in the base station, but for a forbidden circular region of radius $\delta \ll 1$, and the frequency selective fading is i.i.d. chi-squared distributed (Rayleigh fading, independent in the frequency domain), in [7] we show that

$$F_{s\max\{f\},n}(x) = \frac{1}{1-\delta^2}\frac{1}{x^{2/\alpha}}\int_{x^{2/\alpha}\delta^2}^{x^{2/\alpha}} (1 - \exp(-y^{\alpha/2}))^n \, {\rm d}y \tag{11}$$

where α denotes the path-loss exponent. The integral in (11) can be given in closed form in some special cases, in particular, for for $\alpha = 2$ and $\alpha = 4$ (see [7]).

The theoretical foundations of Gaussian parallel multiple-access channels were laid down in [8]. 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

$$\begin{aligned} R_k &= \sum_{m=1}^M R_k^m, \qquad k = 1, \ldots, K & (12) \\ E_k &= \sum_{m=1}^M E_k^m, \qquad k = 1, \ldots, K & (13) \end{aligned}$$

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 $\mathbf{E}^m = (E_1^m, \ldots, E_K^m)$, $\mathbf{R}^m = (R_1^m, \ldots, R_K^m)$ and $\mathbf{d}^m = (d_1^m, \ldots, d_K^m)$, the capacity region for given per-user energies $\mathbf{E} = (E_1, \ldots, E_K)$ and channel gains can be written as

$$\mathcal{C}_{\rm MAC}(\mathbf{d}^1, \ldots, \mathbf{d}^M; \mathbf{E}) = \bigcup_{\substack{\sum_m E_k^m \leq E_k \\ k=1,\ldots,K}} \left\{ \mathbf{R} = \sum_m \mathbf{R}^m : \mathbf{R}^m \in \mathcal{C}_{\rm MAC}(\mathbf{d}^m; \mathbf{E}^m) \right\} \tag{14}$$

In other words, the partial rates R_k^m and energies E_k^m must obey the constraints (2) and (4) in each of the subchannels 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

$$\mathcal{C}_{\rm orth}(\mathbf{d}^1, \ldots, \mathbf{d}^M; \mathbf{E}) = \bigcup_{\substack{\sum_k \Theta_k^m \leq 1 \\ m=1,\ldots,M}} \bigcup_{\substack{\sum_m E_k^m \leq E_k \\ k=1,\ldots,K}} \left\{ \mathbf{R} = \sum_m \mathbf{R}^m : \mathbf{R}^m \in \mathcal{C}_{\rm orth}(\mathbf{d}^m; \mathbf{E}^m; \boldsymbol{\Theta}^m) \right\} \tag{15}$$

A. Delay-limited systems

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. The goal of an efficient resource allocation strategy is to find the partial rates allocation (and the resource-sharing fractions in the case of orthogonal signaling) in order to minimize the required $(E_b/N_0)_{\rm sys}$ to maintain a given rate K-tuple.

Consider the case of optimal signaling. Let π^m denote the permutation that sorts the gains $\mathbf{d}^m$ in increasing order. The resulting optimization problem minimizes

$$\left(\frac{E_b}{N_0}\right)_{\rm sys} = \frac{1}{\Gamma}\sum_{k=1}^{K}\sum_{m=1}^{M}\frac{1}{d^m_{\pi^m_k}}\left[\exp\left(\sum_{i\leq k} R^m_{\pi^m_i}\right) - \exp\left(\sum_{i<k} R^m_{\pi^m_i}\right)\right] \tag{16}$$

(recall that $\Gamma = \sum_k R_k$ is fixed by the user rates), subjects to the constraints (12). This is a convex optimization problem that can be solved with standard tools. In particular, since the constraint (12) is separable and the objective function (16) is convex, in [7] we find a simple iterative block-coordinate descent algorithm that provably converges to the optimum. Details are omitted for lack of space.

For orthogonal signaling, the resulting optimization problem consists of minimizing

$$\left(\frac{E_b}{N_0}\right)_{\rm sys} = \frac{1}{\Gamma}\sum_{k=1}^{K}\sum_{m=1}^{M}\frac{\Theta^m_k}{d^m_k}\left(\exp(R^m_k/\Theta^m_k) - 1\right) \tag{17}$$

with respect to $\{\mathbf{\Theta}^m, \mathbf{R}^m : m = 1, \ldots, M\}$. This is also a convex optimization problem, since the function $g(x, y) = x\exp(x/y)$ is convex. Furthermore, the constraints $\sum_{m=1}^{M} R^m_k \geq R_k$, $\sum_{k=1}^{K}\Theta^m_k \leq 1$ with $R^m_k \geq 0$ and $\Theta^m_k > 0$ are separable. Then, in [7] we find a simple iterative block-coordinate descent algorithm that converges to the optimum.

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 $\mathsf{SNR} = E_{\rm tot}/N_0$ denote the transmit SNR in each slot. 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) [9]–[11].

Because of the different path loss (due to different distance to the base station), the channel gains are independent but not necessarily identically distributed across the users. 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 [12]. PFS allocates user k on channel m at any given slot t if $\widehat{k}_m = k$, where

$$\widehat{k}_m(t) = \arg\max_{k'=1,\ldots,K}\frac{\log(1 + d^m_{k'}(t)\mathsf{SNR})}{T_{k'}(t)} \tag{18}$$

where $d^m_k(t)$ denotes the gain of the m-th channel of user k at slot time t and $T_k(t)$ denotes the long-term average throughput of user k at time t. We omit further details on PFS for lack of space and because they are available in the literature. In [7] we prove the following result:

Theorem 1: For any given K and under the channel gain statistics defined above, the spectral efficiency C vs. system E_b/N_0 achieved by PFS is parametrically given by

$$\begin{aligned}\mathsf{C} &= \int_0^\infty \log_2(1 + x\mathsf{SNR})\mathrm{d}F_{s\max\{f\},K}(x)\\ \left(\frac{E_b}{N_0}\right)^{\rm PFS}_{\rm sys} &= \frac{\mathsf{SNR}}{\mathsf{C}}\end{aligned} \tag{19}$$

□

III. DELAY-LIMITED SYSTEMS FOR LARGE K

In this section we study the delay-limited systems in the limit of large K. This asymptotic analysis yields both elegant closed-form expressions for $(E_b/N_0)_{\rm sys}$ as a function of C, and some interesting considerations on system design. We make the following assumptions:

(A1) M is fixed while K becomes arbitrarily large.

(A2) As $K \to \infty$, the empirical joint channel gain distribution, converges almost surely to a given deterministic cumulative distribution function (cdf) $F(x^1, \ldots, x^M)$. Moreover, $F(\cdot)$ is assumed to be symmetric, in the sense already defined before.

(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 converges almost surely to a given deterministic cdf $G(x)$ with mean 1 and support in $[a, b]$ as $K \to \infty$, where $0 \leq a \leq b < \infty$ 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 $\{(d^1_k, \ldots, d^M_k, \nu_k) : k = 1, \ldots, K\}$ converges to the product cdf $F(x^1, \ldots, x^M)G(z)$. 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, proved in [7].

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

$$\left(\frac{E_b}{N_0}\right)_{\rm sys} = \log(2)\int_0^\infty \frac{2^{\mathsf{C}F_{s\max\{f\},M}(x)}\mathrm{d}F_{s\max\{f\},M}(x)}{x} \tag{20}$$

This is achieved by letting each user transmit on its best subchannel only, and by using superposition coding and successive decoding on each subchannel. □

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

$$\left(\frac{E_b}{N_0}\right)_{\rm sys} = \log(2)\int_0^\infty \frac{\exp\left(1+W\left(\frac{\mu x-1}{e}\right)\right)-1}{1+W\left(\frac{\mu x-1}{e}\right)}\frac{{\rm d}F_{s\max\{f\},M}(x)}{x} \tag{21}$$

where $W(x)$ is Lambert's W function [13] and where μ is the solution of

$$\int_0^\infty \frac{{\rm d}F_{s\max\{f\},M}(x)}{1+W\left(\frac{\mu x-1}{e}\right)} = \frac{1}{\mathsf{C}\log(2)} \tag{22}$$

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. □

In a conventional TDMA/FDMA system, 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 $K \rightarrow \infty$ the $(E_b/N_0)_{\rm sys}$ vs. C for a conventional TDMA/FDMA system is given by

$$\left(\frac{E_b}{N_0}\right)_{\rm sys} = \frac{2^{\mathsf{C}}-1}{\mathsf{C}}\int_0^\infty \frac{{\rm d}F_{s\max\{f\},M}(x)}{x} \tag{23}$$

□

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 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 (23) coincides with the spectral efficiency vs. E_b/N_0 for a single-user system under the *channel inversion* power control strategy.

The above results show that, in the regime of large K, the optimal delay-limited strategy for both optimal and orthogonal signaling consists of letting each user transmit on its own best subchannel only, irrespectively of the other users. Our results suggest 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. This feature is referred to as "Cognitive Radio", a technology that is gaining an increasing interest also in the standardization environment [14].

IV. COMPARISONS

Fig.1 compares the spectral efficiency for the case $\alpha = 2$, $\delta = 0.01$, $M = 10$ parallel channels, for the PFS with $K = 10, 20, 30, 50$ and 100 users (spectral efficiency increases with K as $\log\log K$ for high C), and the optimal, optimized orthogonal and conventional TDMA/FDMA delay-limited systems (we used the asymptotic expressions for large K). We conclude that while the advantage of PFS is very significant at low spectral efficiency, imposing a strict delay constraint does not incur much loss in practice (say, up to $K \sim 100$ users per cell) for high spectral efficiency. A more refined analysis in terms of minimum E_b/N_0, wideband slope and high SNR dB penalty (along with closed form results) can be found in [7] and allows us to quantify exactly the effect of multiuser diversity for both delay-limited and delay-tolerant systems.

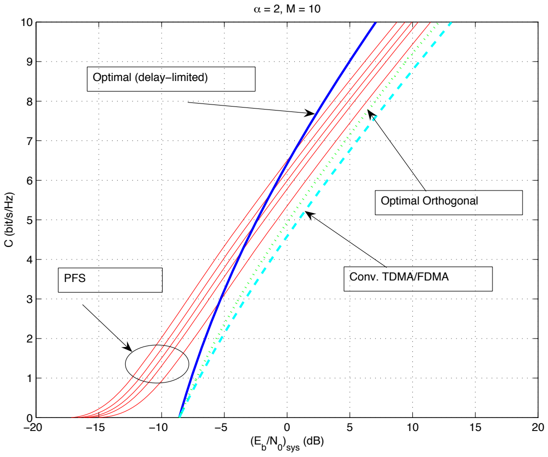

Fig. 1. System comparison in terms of C vs. $(E_b/N_0)_{\rm sys}$.

REFERENCES

[1] Thomas M. Cover and Joy A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, 1991.

[2] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. on Inform. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.

[3] Ralf R. Müller, *Power and Bandwidth Efficiency of Multiuser Systems with Random Spreading*, Shaker–Verlag, Aachen, Germany, 1999.

[4] Stephen V. Hanly and David N.C. Tse, "Multi–access Fading Channels: Part II: Delay–Limited Capacities," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2816–2831, Nov. 1998.

[5] Nihar Jindal, Sriram Vishwanath, and Andrea Goldsmith, "On the Duality of Gaussian Multiple-Access and Broadcast Channels," *IEEE Transactions on Information Theory*, vol. 50, no. 5, pp. 768–783, May 2004.

[6] Ezio Biglieri, John Proakis, and Shlomo Shamai (Shitz), "Fading Channels: Information–Theoretic and Communications Aspects," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[7] G. Caire, R. Müller, and R. Knopp, "Multiuser Diversity in Wireless Systems: Delay-Limited vs. Scheduling," submitted to IEEE Trans. on Inform. Theory, January 2006.

[8] Roger S. Cheng and Sergio Verdú, "Gaussian Multiaccess Channels with ISI: Capacity Region and Multiuser Water–Filling," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 773–785, May 1993.

[9] R. Knopp and P.A.Humblet, "Information Capacity and Power Control in Single-Cell Multiuser Communications," in *Proc. of IEEE International Conference on Communications (ICC)*, Seattle, WA, June 1995, pp. 331–335.

[10] D. Tse, "Optimal Power Allocation over Parallel Gaussian Broadcast Channels," in *IEEE Int. Symp. Inform. Theory (ISIT'97)*, Ulm Germany, June 29-July 4 1997, p. 27.

[11] 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. Inform. Theory*, vol. 47, no. 3, pp. 1083–1127, March 2001.

[12] P.Viswanath, D.N.C.Tse, and R.Laroia, "Opportunistic Beamforming Using Dumb Antennas," *IEEE Trans. on Inform. Theory*, vol. 48, no. 6, June 2002.

[13] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert W Function," *Advances in Computational Mathematics*, vol. 5, pp. 329–359, 1996.

[14] "URL: `http://www.fcc.gov/oet/cognitiveradio/`," Federal Communications Commission home page.