Multiuser Mercury/waterfilling for Downlink OFDM with Arbitrary Signal Constellations

Angel Lozano
Bell Labs (Lucent Technologies)
Holmdel, NJ 07733, USA
Email: aloz@lucent.com

Antonia M. Tulino
Universitá di Napoli "Federico II"
Naples 80125, Italy
Email: atulino@ee.princeton.edu

Sergio Verdú
Princeton University
Princeton, NJ 08544, USA
Email: verdu@princeton.edu

***Abstract*— This paper formulates the power allocation policy that maximizes the region of mutual informations achievable in multiuser downlink OFDM channels. Arbitrary partitioning of the available tones among users and arbitrary modulation formats, possibly different for every user, are considered. The policy, derived for slowly fading channels tracked by the base station, adopts the form of a multiuser mercury/waterfilling procedure that generalizes the single-user mercury/waterfilling introduced in [1].**

I. Introduction

There is, of late, great interest in OFDM (orthogonal frequency-division multiplexing) for multiuser wireless downlinks. Although in general suboptimal in the face of instantaneous CSIT (channel state information at the transmitter), orthogonal multiplexing techniques are much more robust to CSIT inaccuracies than nonorthogonal schemes. Furthermore, OFDM has the added benefit of being naturally well suited to deal with frequency selectivity [2]. With the continued increase in signal bandwidths, OFDM is poised to be a central ingredient of most wireless systems to come.

The present paper formulates the optimum power allocation policy for multiuser OFDM downlinks with instantaneous CSIT. This policy, which maximizes the mutual information region, extends to the multiuser realm the single-user mercury/waterfilling policy presented in [1] enabling:

- Determination of the region of spectral efficiencies reliably achievable for any partition of the available tones and any modulation format.
- A benchmark against which suboptimal power allocation policies can be gauged.
- Assessment of the fundamental advantage of allocating power on the basis of instantaneous CSIT.

For the sake of brevity, the proofs of the various results are not included in the manuscript.

II. Models and Definitions

A. Multiuser OFDM

Consider a downlink channel partitioned into n orthogonal tones, sized such that each experiences (approximately) frequency-flat fading. A scalar signal is transmitted on every tone. A scheduler at the base station assigns each tone to one of k users, determines the signalling constellation to be used by each user on its assigned tones, and establishes user priorities from the nonnegative set $\{w_j\}_{j=1}^{k}$ such that

$$\sum_{j=1}^{k} w_j = 1. \tag{1}$$

Denoting by n_j the number of tones assigned to user j, the input-output relationship on the ith tone of the jth user is

$$Y_{i,j} = h_{i,j} X_{i,j} + W_{i,j} \qquad i = 1, \ldots, n_j \quad j = 1, \ldots, k \tag{2}$$

where $h_{i,j}$ is a complex gain while the noise $W_{i,j}$ is a zero-mean unit-variance complex Gaussian random variable independent of the noise on the other tones. Noting that the tones assigned to a given user may be nonadjacent, we define

$$\beta_j = \frac{n_j}{n} \tag{3}$$

as the fraction of the total bandwidth assigned to user j.

The complex signals $\{X_{i,j}\}$, zero-mean and mutually independent, must satisfy the power constraint

$$\frac{1}{n} \sum_{i=1}^{n_j} \sum_{j=1}^{k} E\left[|X_{i,j}|^2\right] \leq P \tag{4}$$

where P is not a function of time. It is convenient to introduce normalized unit-power signals

$$S_{i,j} = \frac{X_{i,j}}{E\left[|X_{i,j}|^2\right]}, \tag{5}$$

whose distribution is dictated by the modulation scheme used by the corresponding user, and a normalized power allocation

$$p_{i,j} = \frac{E\left[|X_{i,j}|^2\right]}{P}. \tag{6}$$

We can then define, for each tone, the channel state

$$\gamma_{i,j} = P\,|h_{i,j}|^2 \tag{7}$$

which represents the receive signal-to-noise ratio on the ith tone of the jth user when the power allocation is uniform, i.e., $p_{i,j} = 1\ \forall i, j$. (More generally, the receive signal-to-noise ratio is $p_{i,j}\gamma_{i,j}$.) With that, (2) under coherent reception is equivalent to

$$Y_{i,j} = \sqrt{\gamma_{i,j}\, p_{i,j}}\, S_{i,j} + W_{i,j} \tag{8}$$

subject to

$$\frac{1}{n}\sum_{i=1}^{n_j}\sum_{j=1}^{k} p_{i,j} \leq 1. \tag{9}$$

B. Fading Channels and CSIT

For every user j, the channel states $\{\gamma_{i,j}\}_{i=1}^{n_j}$ on the tones assigned to that user have the same marginal distribution, determined by the location of that particular user, and are known by its receiver. In particular,

$$E\left[\gamma_{i,j}\right] = \bar{\gamma}_j \qquad\qquad i = 1, \ldots, n_j \tag{10}$$

where $\bar{\gamma}_j$ is a measure of the local-average signal-to-noise ratio at the location of user j. The $\{\gamma_{i,j}\}_{i=1}^{n_j}$ are independent if the frequency separation of the corresponding tones exceeds some finite coherence bandwidth. Moreover, $\{\gamma_{i,j}\}_{i=1}^{n_j}$ are independent from their counterparts for all other users.

We shall consider slowly fading channels where $\{\gamma_{i,j}\}$ do not change appreciably during each codeword. The transmitter can then track the fading states. (Precisely, the formulation only requires that the amplitudes $\{|\gamma_{i,j}|\}$ be tracked).

In the case of Rayleigh fading, which shall be invoked in many of our examples, every $\gamma_{i,j}$ has an exponential density

$$f_{\gamma_{i,j}}(\xi) = \frac{e^{-\xi/\bar{\gamma}_j}}{\bar{\gamma}_j} \qquad\qquad i = 1, \ldots, n_j. \tag{11}$$

III. Mutual Information and MMSE

Our measure of performance is the mutual information, which specifies the maximum spectral efficiency achievable with arbitrary reliability for a given modulation format. Given a scalar Gaussian-noise channel of the form $Y = \sqrt{\rho}\, S + W$, we denote its mutual information by

$$\mathcal{I}(\rho) = I(S; \sqrt{\rho}\, S + W) \tag{12}$$

which is maximized when S is Gaussian, for which $\mathcal{I}(\rho) = \log(1+\rho)$ where the base of the logarithm determines the information units. While ideal, Gaussian signals cannot be realized in practice because of their continuous and unbounded support. Rather, signals usually conform to discrete constellations with limited peak-to-average ratios. For a discrete constellation (m-QAM, m-PSK, etc) consisting of m points, $\{s_\ell\}_{\ell=1}^{m}$, taken with probabilities $\{q_\ell\}_{\ell=1}^{m}$ such that $\sum_{\ell=1}^{m} q_\ell = 1$,

$$\mathcal{I}(\rho) = -\log(\pi e) - \int f_m(y,\rho)\log f_m(y,\rho)\, dy \tag{13}$$

where the integration extends to the complex plane while

$$f_m(y,\rho) = \frac{1}{\pi}\sum_{\ell=1}^{m} q_\ell\, e^{-|y-\sqrt{\rho}\, s_\ell|^2}. \tag{14}$$

A defining feature of any discrete constellation is the minimum distance between any two of its points, which we indicate by

$$d = \min_{\substack{k,\ell \\ k\neq\ell}} |s_k - s_\ell|. \tag{15}$$

The fact that, for non-Gaussian signals, the mutual information cannot in general be expressed in closed form greatly complicates optimization procedures that entail its differentiation. Propitiously, a recently unveiled relationship [3] affirms that, regardless of the distribution of the signal S,

$$\frac{d}{d\rho}\mathcal{I}(\rho) = \mathsf{MMSE}(\rho) \tag{16}$$

where $\mathcal{I}(\cdot)$ is in nats/s/Hz and the function $\mathsf{MMSE}(\cdot)$ returns the minimum mean-square error in estimating S by observing Y. This minimum mean-square error is achieved by the conditional-mean estimator

$$\hat{S}(y,\rho) = E\left[S|\sqrt{\rho}\, S + W = y; \rho\right] \tag{17}$$

which is, in general, a nonlinear function of the observation y. (It becomes linear if S is Gaussian.) Therefore,

$$\mathsf{MMSE}(\rho) = E\left[\left|S - \hat{S}(\sqrt{\rho}\, S + W, \rho)\right|^2\right] \tag{18}$$

with expectation over both S and W. Since S is unit power, $\mathsf{MMSE}(\cdot) \in [0,1]$. The inverse of $\mathsf{MMSE}(\cdot)$ with respect to the composition of functions is denoted by $\mathsf{MMSE}^{-1}(\cdot) \in [0,\infty)$.

For a Gaussian signal, (17) becomes

$$\hat{S}(y,\rho) = \frac{\sqrt{\rho}}{1+\rho}\, y \tag{19}$$

leading to $\mathsf{MMSE}(\rho) = 1/(1+\rho)$ and, in turn, to

$$\mathsf{MMSE}^{-1}(\zeta) = \frac{1}{\zeta} - 1. \tag{20}$$

For discrete constellations, (17) yields

$$\hat{S}(y,\rho) = \frac{\sum_{\ell=1}^{m} q_\ell\, s_\ell\, e^{-|y-\sqrt{\rho}\, s_\ell|^2}}{\sum_{\ell=1}^{m} q_\ell\, e^{-|y-\sqrt{\rho}\, s_\ell|^2}} \tag{21}$$

from which the $\mathsf{MMSE}(\cdot)$ follows via (18), which can be easily implemented as a low-pass filter driven by the estimation error $|S-\hat{S}|^2$. Alternatively, the $\mathsf{MMSE}(\cdot)$ can be tabulated and stored in memory for each of the constellations in use.

IV. Multiuser Mercury/waterfilling

Since $\{|\gamma_{i,j}|\}$ are known by the transmitter, the power allocation is a function thereof. Every realization of $\{\gamma_{i,j}\}$ thus gives rise to a k-dimensional region containing all of the feasible k-tuples $\{\mathcal{R}_1, \ldots, \mathcal{R}_k\}$ where

$$\mathcal{R}_j(p_{1,j}, \ldots, p_{n_j,j}) = \frac{1}{n_j}\sum_{i=1}^{n_j}\mathcal{I}_j(p_{i,j}\gamma_{i,j}) \tag{22}$$

is the mutual information attained by user j on its assigned tones with the function $\mathcal{I}_j(\cdot)$ reflecting the signalling scheme used by that user. The boundary of this region can be fully characterized by means of the weighted function $\sum_{j=1}^{k} w_j\, \mathcal{R}_j(\cdot)$ for all priorities $\{w_j\}_{j=1}^{k}$ satisfying (1). We thus seek the power allocation $\{p_{i,j}^{\star}\}$ that solves

$$\{p_{i,j}^{\star}\} = \arg \max_{\{p_{i,j}\}: \frac{1}{n}\sum_{i,j} p_{i,j} \leq 1} \frac{1}{n}\sum_{j=1}^{k}\frac{w_j}{\beta_j}\sum_{i=1}^{n_j}\mathcal{I}_j(p_{i,j}\gamma_{i,j}). \tag{23}$$

yielding the optimum boundary $\{\mathcal{R}_1^{\star}, \ldots, \mathcal{R}_k^{\star}\}$.

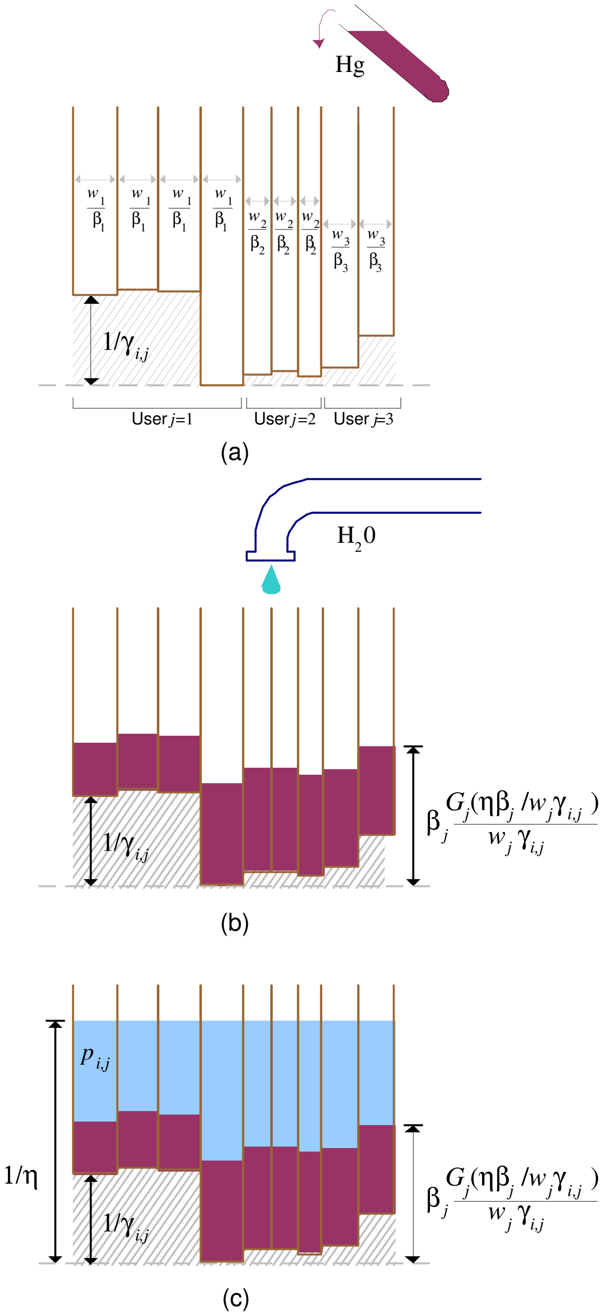

Fig. 1. Multiuser mercury/waterfilling with $k = 3$ users and with adjacent tones assigned to each user for clarity. (a) On vessels of base $(w_j/\beta_j) \times 1$ solid up to $1/\gamma_{i,j}$, pour mercury as shown. (b) Waterfill until the water levels reach $1/\eta$. (c) The volume of water in the (i,j)th vessel gives $p_{i,j}$.

Theorem 1: The unique power allocation that solves (23) is

$$p^\star_{i,j} = 0 \qquad \gamma_{i,j} \le \frac{\beta_j}{w_j}\eta \tag{24}$$

$$p^\star_{i,j} = \frac{1}{\gamma_{i,j}} \mathsf{MMSE}_j^{-1}\left(\frac{\beta_j \eta}{w_j \gamma_{i,j}}\right) \qquad \gamma_{i,j} > \frac{\beta_j}{w_j}\eta \tag{25}$$

with η such that (9) is met with strict equality.

The strategy spelled by Theorem 1 is as follows. No power is allocated to tones whose channel state is below a threshold, $(\beta_j/w_j)\eta$, that is directly proportional to the bandwidth fraction of the corresponding user and inversely proportional to its priority. Active tones, in turn, are allocated the exact amount of power needed to render $\gamma_{i,j}\mathsf{MMSE}_j(p^\star_{i,j}\gamma_{i,j})$ equal to the threshold of their respective users.

Computationally, Theorem 1 boils down to solving a single nonlinear equation on η, from which $\{p^\star_{i,j}\}$ are then simply mapped via (24) and (25).

In order to provide a graphical interpretation, let us define the auxiliary function

$$G_j(\zeta) = \begin{cases} 1/\zeta - \mathsf{MMSE}_j^{-1}(\zeta) & \zeta \in [0,1] \\ 1 & \zeta > 1 \end{cases} \tag{26}$$

which, for a Gaussian signal, reduces to $G_j(\zeta) = 1$. Theorem 1 is tantamount to the following multiuser mercury/waterfilling procedure (cf. Fig. 1):

(a) For each of the n tones, set up a vessel of base $(w_j/\beta_j) \times 1$, solid up to a height $1/\gamma_{i,j}$.
(b) Choose η. Pour mercury onto each of the vessels until its height (including the solid) reaches $\frac{\beta_j}{w_j\gamma_{i,j}} G_j(\frac{\beta_j\eta}{w_j\gamma_{i,j}})$.
(c) Waterfill, keeping a common upper level of water, until the water level reaches $1/\eta$.
(d) The volume of water in the (i,j)th vessel gives $p_{i,j}$.

As in single-user mercury/waterfilling, the mercury regulates the water admitted by each vessel thereby accounting for the respective signalling constellations. The new feature in multiuser mercury/waterfilling is the variability in vessel widths, which also regulates the water admission but in relation with the user priorities and assigned bandwidths. Since no mercury is poured onto vessels whose signal is Gaussian, multiuser mercury/waterfilling reverts to a multiuser waterfilling whenever all of the signals are Gaussian:

$$p^\star_{i,j} = 0 \qquad \gamma_{i,j} \le \frac{\beta_j}{w_j}\eta \tag{27}$$

$$p^\star_{i,j} = \frac{w_j/\beta_j}{\eta} - \frac{1}{\gamma_{i,j}} \qquad \gamma_{i,j} > \frac{\beta_j}{w_j}\eta \tag{28}$$

In this case, we can in fact clear the parameter η and obtain an alternative fixed-point form for the solution.

Corollary 1: With Gaussian signals,

$$p^\star_{i,j} = 0 \qquad \gamma_{i,j} \le \frac{\eta}{w_j}$$

$$p^\star_{i,j} = \frac{\frac{w_j}{\beta_j}\left(1 - \mathsf{MMSE}_j(p^\star_{i,j}\gamma_{i,j})\right)}{\frac{1}{n}\sum_{\kappa=1}^{k}\frac{w_\kappa}{\beta_\kappa}\sum_{\ell=1}^{n_\kappa}\left(1 - \mathsf{MMSE}_\kappa(p^\star_{\ell,\kappa}\gamma_{\ell,\kappa})\right)} \qquad \gamma_{i,j} > \frac{\eta}{w_j}$$

Although in principle less appealing than (27)–(28), the fixed-point characterization in Corollary 1 has the advantage of generalizing to nonorthogonal parallel channels [4].

V. Low-Power and High-Power Regimes

In the low-power regime, multiuser mercury/waterfiling behaves as follows for quadrature-symmetric signals, a class that encompasses ideal Gaussian signals as well as discrete constellations such as m-QAM and m-PSK [5].

Proposition 1: For $P \to 0$, multiuser mercury/waterfilling allocates power only to the tone(s) with the largest factor $w_j \gamma_{i,j}/\beta_j$. If several tones share the largest such factor, then power is uniformly distributed thereon.

In contrast with the low-power regime, where optimality in terms of mutual information boils down to quadrature symmetry only, in the high-power regime the nonidealities of discrete constellations become overtly manifest.

Proposition 2: If the signals are Gaussian, then for $P \to \infty$

$$p^\star_{i,j} = \frac{w_j}{\beta_j} + \mathcal{O}(\frac{1}{P}) \qquad i = 1, \ldots, n_j \tag{29}$$

whereas, if the signals conform to discrete constellations with minimum distance d_j for user j, then

$$p^\star_{i,j} = \frac{\alpha}{|h_{i,j}|^2 d_j^2} + \mathcal{O}(\frac{\log P}{P}) \tag{30}$$

with

$$\alpha = \frac{1}{\frac{1}{n}\sum_{\kappa=1}^{k}\sum_{\ell=1}^{n_\kappa}\frac{1}{|h_{\ell,\kappa}|^2 d_\kappa^2}}. \tag{31}$$

Notice the discrepancy between the leading terms in the high-power expansion with ideal Gaussian signals and with discrete constellations. With the former, the power allocated to a tone is dominated by the priority and bandwidth fraction of the respective user. With the latter, in contrast, the user priority and bandwidth fraction become immaterial for large P. Only the channel states and the constellation minimum distances are of essence, with more power allocated to tones whose users are employing richer constellations but with less power on tones with stronger channel states.

VI. Ergodic Characterization

The optimum power allocation $\{p^\star_{i,j}\}$ and the corresponding user mutual informations $\{\mathcal{R}^\star_j\}_{j=1}^k$ can be regarded as random variables whose distributions are induced by the channel states, $\{\gamma_{i,j}\}$. For delay-tolerant applications, nonetheless, the time averages of the mutual informations acquire operational significance. Under ergodic fading, these time averages equal $\bar{\mathcal{R}}^\star_j = E[\mathcal{R}^\star_j]$. The corresponding average power allocation,

$$\bar{p}^\star_j = E[p^\star_{i,j}] \qquad i = 1, \ldots, n_j \tag{32}$$

meaningfully conveys how the power is allocated on average. Note that this average allocation is common to all the tones of a given user because of their identical marginal fading distribution and equal signalling constellation.

Example 1: Consider an access point streaming data to $k = 2$ users over respective frequency-flat Rayleigh-faded channels with equal bandwidth assigned to each user (i.e., $\beta_1 = \beta_2 = 1/2$) and with $\bar{\gamma}_1|_{\text{dB}} = \bar{\gamma}_2|_{\text{dB}} = 5$.[1] The average multiuser mercury/waterfilling power allocation as function of the priority w_1, is depicted in Fig. 2 parameterized by the constellation used by both users. The corresponding ergodic mutual information region boundaries are shown in Fig. 3.

[1] $x|_{\text{dB}} = 10 \log_{10} x$.

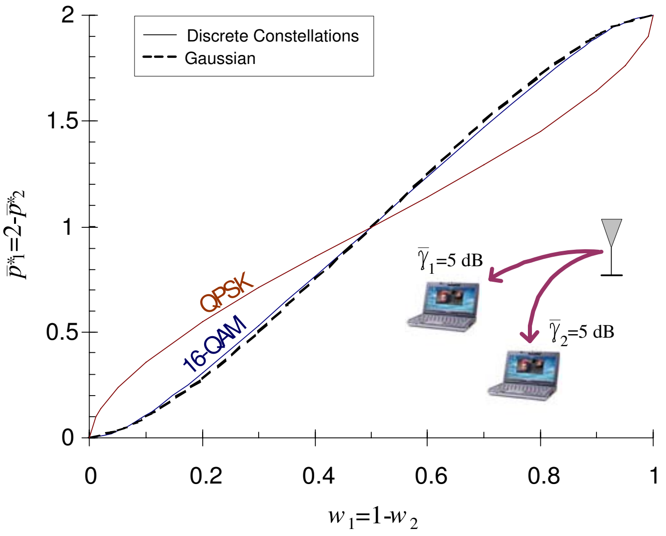

Fig. 2. Average power allocation as function of w_1, for $k = 2$ users having $\bar{\gamma}_1|_{\text{dB}} = \bar{\gamma}_2|_{\text{dB}} = 5$. Both channels are frequency-flat Rayleigh-faded with bandwidth partitioning $\beta_1 = \beta_2 = 1/2$.

With the average conditions in Example 1, the channel states are frequently in a range where 16-QAM is rich enough to resemble an ideal Gaussian signal whereas QPSK is not.

Example 2: Consider the same scenario of Example 1, except with $\bar{\gamma}_1|_{\text{dB}} = 10$ and $\bar{\gamma}_2|_{\text{dB}} = 0$. The average multiuser mercury/waterfilling power allocation and the ergodic mutual information region boundaries are shown in Figs. 4 and 5, respectively.

In Example 2, the various signalling constellations behave similarly whenever user $j = 2$ is prioritized because it is often in low-power conditions. When user $j = 1$ is favored, however, there is a large disparity between the power allocation and mutual informations for the several signalling formats.

Another instance in which, notwithstanding the delay tolerances, the average mutual informations provide a meaningful characterization is that of strong frequency selectivity per user, whereby (22) by itself provides an effective averaging mechanism. To gain insight, we consider the limiting regime where $n_j \to \infty$, $j = 1, \ldots, k$.[2] From the asymptotic independence of the channel states for each user, the $\{\mathcal{R}^\star_j\}$ converge in the mean-square sense to nonrandom limits that depend only on the signalling constellations and fading distributions. Precisely,

$$\mathcal{R}^\star_j \to \int_{\frac{\beta_j}{w_j}\eta}^{\infty} \mathcal{I}_j\left(\mathsf{MMSE}^{-1}\left(\frac{\beta_j}{w_j}\frac{\eta}{\xi}\right)\right) f_{\gamma_{i,j}}(\xi)\, d\xi \tag{33}$$

with η the solution of

$$\sum_{j=1}^{k} \beta_j \int_{\frac{\beta_j}{w_j}\eta}^{\infty} \frac{1}{\xi}\, \mathsf{MMSE}_j^{-1}\left(\frac{\beta_j}{w_j}\frac{\eta}{\xi}\right) f_{\gamma_{i,j}}(\xi)\, d\xi = 1. \tag{34}$$

[2] Note that, by virtue of (9), the total transmit power is also growing without bound as the system bandwidth increases.

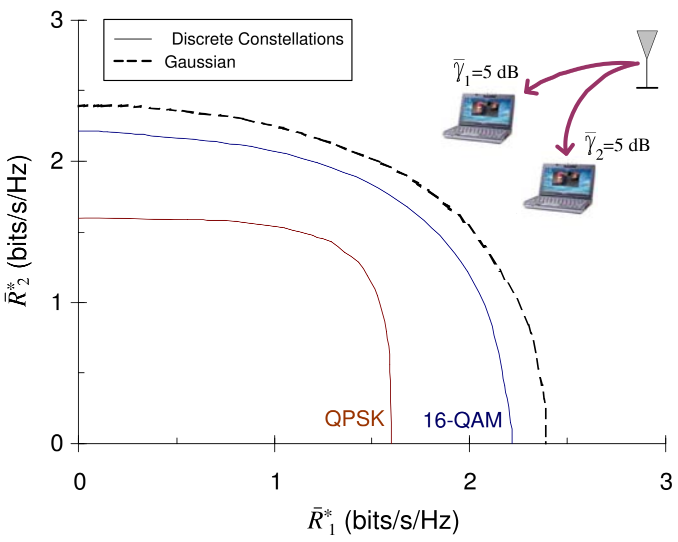

Fig. 3. Average mutual information regions for $k = 2$ users having $\bar{\gamma}_1|_{\text{dB}} = \bar{\gamma}_2|_{\text{dB}} = 5$. Both channels are frequency-flat Rayleigh-faded with bandwidth partitioning $\beta_1 = \beta_2 = 1/2$.

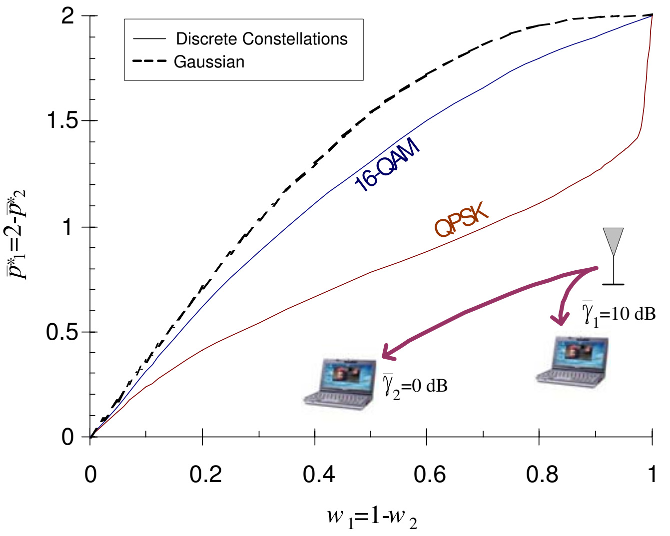

Fig. 4. Average power allocation as function of w_1, for $k = 2$ users having $\bar{\gamma}_1|_{\text{dB}} = 0$ and $\bar{\gamma}_2|_{\text{dB}} = 10$. Both channels are frequency-flat Rayleigh-faded with bandwidth partitioning $\beta_1 = \beta_2 = 1/2$.

The tone powers $\{p^{\star}_{i,j}\}$ remain random, but their empirical distributions for the various users converge in probability to nonrandom limits that can be found from η via (24)–(25).

Specifically for Gaussian signals and Rayleigh fading, we can invoke (11) and (20) to obtain more explicit versions of (34) and (33). Under these conditions

$$\mathcal{R}_j \rightarrow E_1\left(\frac{\beta_j}{w_j \bar{\gamma}_j}\eta\right) \qquad j = 1, \ldots, k \qquad (35)$$

in nats/s/Hz, with $E_1(\zeta) = \int_1^{\infty} t^{-1} e^{-\zeta t} dt$ an exponential integral and with η the solution to

$$\sum_{j=1}^{k}\left(\frac{w_j\, e^{-\frac{\beta_j}{w_j\bar{\gamma}_j}\eta}}{\eta} - \frac{\beta_j}{\bar{\gamma}_j} E_1\left(\frac{\beta_j}{w_j \bar{\gamma}_j}\eta\right)\right) = 1. \qquad (36)$$

VII. Summary

We have formulated the multiuser mercury/waterfilling power allocation policy for OFDM downlinks with arbitrary tone partitioning and modulation formats.[3] The more general problem of jointly assigning tones and allocating power could also be explored, with the multiuser mercury/waterfilling procedure as a building block.

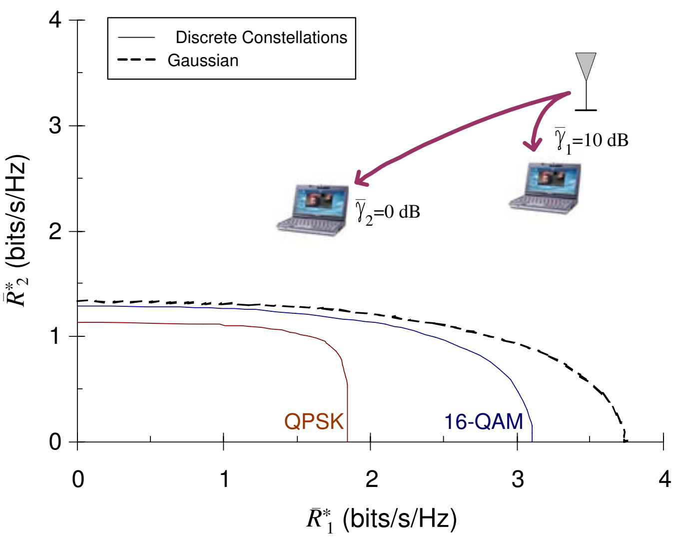

Fig. 5. Average mutual information regions for $k = 2$ users having $\bar{\gamma}_1|_{\text{dB}} = 0$ and $\bar{\gamma}_2|_{\text{dB}} = 10$. Both channels are frequency-flat Rayleigh-faded with bandwidth partitioning $\beta_1 = \beta_2 = 1/2$.

References

[1] A. Lozano, A. M. Tulino, and S. Verdú, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Trans. on Inform. Theory*, vol. 52, no. 7, July 2006.
[2] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Magazine*, vol. 28, no. 5, pp. 5–14, May 1990.
[3] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. on Inform. Theory*, vol. 51, no. 4, pp. 1261–1283, Apr. 2005.
[4] A. M. Tulino, A. Lozano, and S. Verdú, "Capacity-achieving input covariance for single-user multi-antenna channels," *IEEE Trans. on Wireless Communications*, Jan. 2006.
[5] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. on Inform. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.

[3]Although the modulation format has been considered uniform over the tones assigned to each user, this restriction can be easily removed.