Reduced-Complexity Power-Efficient Wireless OFDMA using an Equally Probable CSI Quantizer

Antonio G. Marques*, Fadel F. Digham†, Georgios B. Giannakis (contact author)†, and F. Javier Ramos*
*Dept. of Signal Theory and Communications. Universidad Rey Juan Carlos. Fuenlabrada, 28943 Madrid (SPAIN)
Email: {antonio.garcia.marques, javier.ramos}@urjc.es
†Dept. of Electrical and Computer Engineering. University of Minnesota. Minneapolis, 55455 MN (USA)
Email: {georgios, fdigham}@ece.umn.edu

***Abstract*— Emerging applications involving low-cost wireless sensor networks motivate well optimization of multi-user orthogonal frequency-division multiple access (OFDMA) in the power-limited regime. In this context, the present paper relies on limited-rate feedback (LRF) sent from the access point to terminals to acquire quantized channel state information (CSI) in order to minimize the total average transmit-power under individual average rate and error probability constraints. Specifically, we introduce two suboptimal reduced-complexity schemes to: *(i)* allocate power, rate and subcarriers across users; and *(ii)* design accordingly the channel quantizer. The latter relies on the solution of *(i)* to design equally probable quantization regions per subcarrier and user. Numerical examples corroborate the analytical claims and reveal that the power savings achieved by our reduced-complexity LRF designs are close to those achieved by the optimal solution.**

I. INTRODUCTION

Orthogonal frequency-division multiplexing (OFDM) is the most common modulation for bandwidth limited wireline and wireless transmissions over frequency-selective multipath channels. OFDM transmissions over wireline or slowly fading wireless links have traditionally relied on deterministic or *perfect* (P-) channel state information at the transmitters (CSIT) to adaptively load power, bits and/or subcarriers so as to either maximize rate (capacity) for a prescribed transmit-power, or, minimize power subject to instantaneous rate constraints [9].

While the assumptions of P-CSI at the transmitters and receiver render analysis and design tractable, they may not be as realistic due to wireless channel variations and estimation errors, feedback delay, bandwidth limitation, and jamming induced errors [6]. These considerations motivate a *limited-rate feedback* (LRF) mode, where only *quantized* (Q-) CSIT is available through a (typically small) number of bits fed back from the receiver to the transmitters; see e.g., [10]. Q-CSIT entails a finite number of quantization regions describing different clusters of channel realizations [7], [10]. Upon estimating the channel, the receiver feeds back the index of the region individual uplink channels belong to (channel codeword), based on which each terminal adapts its transmission parameters accordingly. This LRF-based mode of operation fulfills two requirements: (i) the feedback is pragmatically affordable in most practical wireless links, and (ii) the Q-CIST is robust to channel uncertainties since transmitters adapt to a few regions rather than individual channel realizations.

Resource allocation in orthogonal frequency-division multiple access (OFDMA) minimizing the transmit-power *per symbol* based on P-CSIT was first studied in [9]. Relying on fixed (as opposed to adaptive) Q-CSIT, recent works deal with optimization of power or rate performance per OFDMA symbol [2], [5]. Different from these works, here we jointly adapt power, rate, and subcarrier resources based on Q-CSIT to minimize the *average* transmit-power. Our focus is on allocation algorithms with negligible on-line computational complexity. Moreover, we rely on the optimal allocation for designing a novel non-iterative channel quantizer that enforces equally probable quantization regions per user and subcarrier.

The rest of the paper is organized as follows. After introducing preliminaries on the setup we deal with (Section II), for a given quantizer design, we derive suboptimal subcarrier, power, and bit OFDMA allocation (Section III). Once the allocation is characterized, we capitalize on it for designing a quantizer with equally probable regions (Section IV). Numerical results and comparisons that corroborate our claims are presented (Section V), and concluding remarks finish this paper (Section VI)[1].

II. PRELIMINARIES AND PROBLEM STATEMENT

We consider a wireless OFDMA system (see Fig. 1) with M users, indexed by $m \in [1, M]$, sharing K subcarriers (subchannels), indexed by $k \in [1, K]$. The *instantaneous* (per symbol) power and rate user m loads on subcarrier k are denoted by $p_{k,m}$ and $r_{k,m}$, respectively. With these as entries we form $K \times M$ instantaneous power and rate matrices $\mathbf{P}$ and $\mathbf{R}$, that is $[\mathbf{P}]_{k,m} := p_{k,m}$ and $[\mathbf{R}]_{k,m} := r_{k,m}$. For a given

[1]*Notation:* Lower and upper case boldface letters are used to denote (column) vectors and matrices, respectively; $(\cdot)^T$ denotes transpose; $[\cdot]_{k,l}$ the (k,l)th entry of a matrix, and $[\cdot]_k$ the kth entry of a vector; $\mathbf{X} \geq \mathbf{0}$ means all entries of $\mathbf{X}$ are nonnegative; $\mathbf{F}_N$ stands for the normalized FFT matrix with entries $[\mathbf{F}_N]_{n,k} = e^{-j\frac{2\pi}{N}kn}$, $n, k = 0, \ldots, N-1$; $f_{\mathbf{X}}(\mathbf{X})$ denotes the joint probability density function (PDF) of a matrix $\mathbf{X}$; likewise, $f_x(x)$ denotes the PDF of a scalar x; $\mathbb{E}_{\mathbf{X}}[\cdot]$ stands for the expectation operator over $\mathbf{X}$; $\lfloor\cdot\rfloor$ ($\lceil\cdot\rceil$) denotes the floor (ceiling) operation; $\mathbf{I}_{\{\cdot\}}$ is short for the indicator function; i.e., $\mathbf{I}_{\{x\}} = 1$ if x is true and zero otherwise; and LHS(x) denotes the left hand side of equation (x).

Work in this paper was supported by the US ARL under the CTA Program, Cooperative Agreement No. DAAD19-01-2-0011; by the USDoD ARO grant No. W911NF-05-1-0283; and by the Government of C.A. Madrid under grant No. P-TIC-000223-0505.

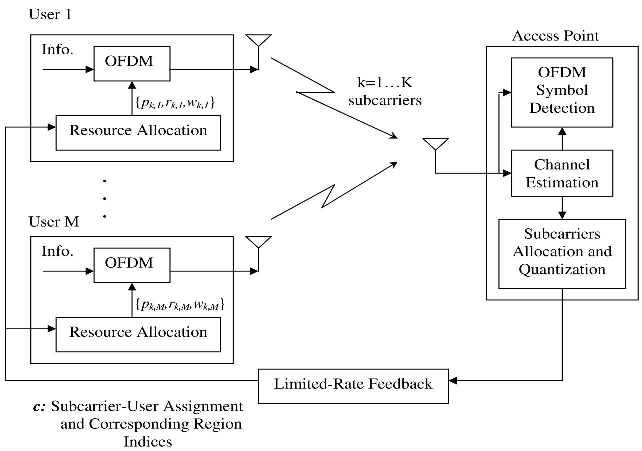

Fig. 1. System block diagram.

feedback update, we consider a time sharing user access per subcarrier; i.e., time division multiple access (TDMA)[2]. This sharing process is described by the $K \times M$ weight matrix $\mathbf{W}$ whose (k, m)th entry $w_{k,m}$ represents the percentage of time the kth subcarrier is utilized by the mth user. Clearly, $\sum_{m=1}^{M} w_{k,m} \leq 1, \forall k$, and the average power and rate over the transmission period between successive feedback updates is $p_{k,m}w_{k,m}$ and $r_{k,m}w_{k,m}$ for the kth subcarrier of user m.

Each user's discrete-time baseband equivalent impulse response of the corresponding frequency-selective fading channel is $\mathbf{h}_m := [h_{m,0}, \ldots, h_{m,N_m}]^T$, where: $N_m := \lfloor D_{m,\max}/T_s \rfloor$ denotes the channel order, $D_{m,\max}$ the maximum delay spread, T_s the sampling period, and $N_{\max} := \max_{m \in [1,M]} N_{m,\max}$. As usual in OFDM, we suppose $K \gg N_{\max}$. For notational convenience, we collect the M impulse response vectors in a $K \times M$ matrix $\mathbf{H} := [\mathbf{h}_1, \ldots, \mathbf{h}_M]$, where the length of each column is increased to K by padding an appropriate number of zeros.

Each user applies a K-point inverse fast Fourier transform (I-FFT) to each snapshot of K-symbol streams, and subsequently inserts a cyclic prefix (CP) of size $N_{\max}$ to obtain a block of $K + N_{\max}$ symbols (i.e., one OFDM symbol), which are subsequently multiplexed and digital to analog (D/A) converted for transmission. These operations along with the corresponding FFT and CP removal at the receiver convert each user's frequency-selective channel to a set of K parallel flat-fading subchannels, each with fading coefficient given by the frequency response of this user's channel evaluated on the corresponding subcarrier. Consider the $K \times M$ matrix $\tilde{\mathbf{H}} := (1/\sqrt{K})\mathbf{F}_K\mathbf{H}$, whose mth column comprises the frequency response of user m's channel.

With the multi-user channel matrix $\tilde{\mathbf{H}}$ acquired (via training symbols), the receiver has available a noise-normalized channel power gain matrix $\mathbf{G}$, where $[\mathbf{G}]_{k,m} := |[\tilde{\mathbf{H}}]_{k,m}|^2/\sigma_{k,m}^2$, with $\sigma_{k,m}^2$ denoting the known variance of the zero-mean additive white Gaussian noise (AWGN) at the receiver. We will use $g_{k,m} := [\mathbf{G}]_{k,m}$ to denote the instantaneous noise-normalized channel power gain for the kth subchannel of the mth user. Likewise, letting $\bar{\mathbf{G}} := \mathbb{E}_{\mathbf{G}}[\mathbf{G}]$, its generic entry $\bar{g}_{k,m} := [\bar{\mathbf{G}}]_{k,m}$ shall denote the average gain of the (k, m) subcarrier-user pair. Having (practically perfect) knowledge of each $\mathbf{G}$ realization, the access point (AP) allocates subcarriers to users after assigning entries of $\mathbf{G}$ to appropriate quantization regions they fall into. Using the indices of these regions, the receiver feeds back the codeword $\mathbf{c} = \mathbf{c}(\mathbf{G})$ for the users to adapt their transmission modes (power, rate and subcarriers) from a finite set of mode triplets.

Our work relies on the following assumptions:

(as1) *Different user channels are uncorrelated; i.e., the columns of* $\mathbf{G}$ *are uncorrelated.*

(as2) *Each user's subchannels are allowed to be correlated, and complex Gaussian distributed; i.e.,* $g_{k,m}$ *obeys an exponential PDF* $f_{g_{k,m}}(g_{k,m}) = (1/\bar{g}_{k,m})\exp(-g_{k,m}/\bar{g}_{k,m})$.

(as3) *Subchannel states (regions) remain invariant over at least two consecutive OFDM symbols.*

(as4) *The feedback channel is error-free and incurs negligible delay.*

(as5) *Symbols are drawn from quadrature amplitude modulation (QAM) constellations so that the resulting instantaneous BER can be approximated as (*$\kappa_1 = 0.2$, $\kappa_2 = 1.5$*)*

$$\epsilon(p_{k,m}, g_{k,m}, r_{k,m}) \simeq \kappa_1 \exp\left(\frac{-p_{k,m}\kappa_2 g_{k,m}}{(2^{r_{k,m}} - 1)}\right). \quad (1)$$

(as6) *A realization of each* $g_{k,m}$ *gain falls into one of* $L_{k,m}$ *disjoint regions* $\{\mathcal{R}_{k,m|l}\}_{l=1}^{L_{k,m}}$.

Since users are sufficiently separated in space (as1) is generally true; (as2) corresponds to fading amplitudes adhering to the commonly encountered Rayleigh model but generalizations are possible; (as3) allows each subchannel to vary from one OFDM symbol to the next so long as the quantization region it falls into remains invariant; error-free feedback under (as4) is guaranteed with sufficiently strong error control codes (especially since data rates in the feedback link are typically low); the accuracy of (as5) is widely accepted; see e.g., [4]; and (as6) represents a practical and low complexity quantization.

The ultimate *goal* in this paper is twofold: (G1) design a channel quantizer to obtain $\mathbf{c}$; and (G2) given $\mathbf{c}$, find appropriate allocation matrices $\mathbf{P}$, $\mathbf{R}$, and $\mathbf{W}$. We want to design $\mathbf{P}$, $\mathbf{R}$, $\mathbf{W}$, and $\{\mathcal{R}_{k,m|l}\}_{l=1}^{L_{k,m}}$ $\forall k, m$, so that the average power $\bar{P}$ is minimized under prescribed average rate $\bar{\mathbf{r}}_0 := [\bar{r}_{0,1}, \ldots, \bar{r}_{0,M}]^T$ and average bit error rate (BER) $\bar{\boldsymbol{\epsilon}}_0 := [\bar{\epsilon}_{0,1}, \ldots, \bar{\epsilon}_{0,M}]^T$ constraints across users.

III. Quantizer and Transmission Mode Design

A. Problem Formulation

Given (as6), let $\boldsymbol{\mathcal{R}}_{k,m|l} := \{\mathbf{G} : g_{k,m} \in \mathcal{R}_{k,m|l}\}$ denote the set of matrices $\mathbf{G}$ for which $g_{k,m}$ belongs to the region $\mathcal{R}_{k,m|l}$. Furthermore, let $p_{k,m|l}$ and $r_{k,m|l}$ denote[3] respectively, the

[2]Orthogonal access schemes other than TDMA are also possible. But as we will see later, the one chosen is not particulary important because the optimal choice will typically correspond to no sharing; i.e., each subcarrier will be owned by a single user.

[3]The subscript l here will be also written explicitly as $l(\mathbf{G})$ in places that this dependence must be emphasized.

instantaneous power and rate loadings of user m on subcarrier k given that $\mathbf{G} \in \boldsymbol{\mathcal{R}}_{k,m|l}$. Recall that $w_{k,m}(\mathbf{G}) \leq 1$, and thus the expected power and bit loadings for the $\mathbf{G}$ realization over the time between successive feedback updates will be $p_{k,m|l} w_{k,m}(\mathbf{G})$ and $r_{k,m|l} w_{k,m}(\mathbf{G})$, respectively.

Our goal is to minimize the average transmit-power $\mathbb{E}_{\mathbf{G}}[p_{k,m|l(\mathbf{G})} w_{k,m}(\mathbf{G})]$ over all subcarriers and users while satisfying average rate and BER requirements. Specifically, we want the average rate of any user (say the mth) across all subcarriers to satisfy $\sum_{k=1}^{K} \mathbb{E}_{\mathbf{G}}[r_{k,m|l(\mathbf{G})} w_{k,m}(\mathbf{G})] \geq [\bar{\mathbf{r}}_0]_m$. To enforce the average BER requirement we will have the instantaneous BER stay always below a pre-specified BER[4]. Then if ϵ_P^{-1} denotes the inverse function involved when solving (1) w.r.t. $p_{k,m}$ and $g_{k,m|l}^{\min} := \min\{g_{k,m} \mid g_{k,m} \in \mathcal{R}_{k,m|l}\}$ represent the worst channel gain, the power loading

$$p_{k,m|l(\mathbf{G})} := \epsilon_P^{-1}(r_{k,m|l(\mathbf{G})}, g_{k,m|l(\mathbf{G})}^{\min}, [\bar{\boldsymbol{\epsilon}}_0]_m) \tag{2}$$

will automatically fulfill the BER requirement.

Analytically, the constrained optimization problem we wish to solve is:

$$\begin{cases} \min_{\mathbf{R}(\mathbf{G}) \geq \mathbf{0}, \mathbf{W}(\mathbf{G}) \geq \mathbf{0}} \ \bar{P}, \ \text{where } \bar{P} := \\ \ \sum_{k=1}^{K} \sum_{m=1}^{M} \mathbb{E}_{\mathbf{G}}[\epsilon_P^{-1}(r_{k,m|l(\mathbf{G})}, g_{k,m|l(\mathbf{G})}^{\min}, [\bar{\boldsymbol{\epsilon}}_0]_k) w_{k,m}(\mathbf{G})] \\ \text{subject to}: \\ C1. - \sum_{k=1}^{K} \mathbb{E}_{\mathbf{G}}[r_{k,m|l(\mathbf{G})} w_{k,m}(\mathbf{G})] + [\bar{\mathbf{r}}_0]_m \leq 0, \forall m, \\ C2. \ \sum_{m=1}^{M} w_{k,m}(\mathbf{G}) - 1 \leq 0, \ \ \forall k, \mathbf{G}, \\ C3. - r_{k,m|l} \leq 0, \forall k, m, l, \ \ C4. - w_{k,m}(\mathbf{G}) \leq 0, \forall k, m, \mathbf{G}, \end{cases} \tag{3}$$

where through the second constraint we enforce the total utilization of any subcarrier by all users not to exceed one, per $\mathbf{G}$ realization.

Although the optimization problem in (3) is not convex, defining the auxiliary variable $\bar{r}_{k,m|l} = r_{k,m|l} \bar{w}_{k,m|l}$, with $\bar{w}_{k,m|l} := \int_{\mathbf{G} \in \boldsymbol{\mathcal{R}}_{k,m|l}} w_{k,m}(\mathbf{G}) f_{\mathbf{G}}(\mathbf{G}) d\mathbf{G}$ we can render it convex. Since $\bar{w}_{k,m|l} \neq 0$ the resultant convex optimization problem using globally convergent interior point algorithms [1]. The final formulation now enjoys both reduced complexity and global convergence in polynomial time thanks to convexity.

The objective in (3) is to minimize the average power over all possible channel realizations. However, the constraints involve different forms of CSI: while $C2$ and $C4$ needs to be satisfied per channel realization (and thus will change depending on $\mathbf{G}$); $C1$ is an average requirement and $C3$ is unique per region, and therefore their values do not depend of the concrete channel realization $\mathbf{G}$. In the following subsection, we will derive the Karush-Kuhn-Tucker (KKT) conditions associated with (3). These will lead us not only to the expressions determining the optimal loading variables but will also provide valuable insights about the structure of the power-efficient resource allocation policies. At this point, a remark is due on another aspect related to power efficiency in OFDMA.

Remark 1: Although the peak-to-average-power-ratio (PAPR) plays an important role in power (battery) consumption of OFDM systems, in (3) we did not impose PAPR constraints. The underlying reason is that available digital predistortion schemes can be applied to the users' OFDM symbols to meet such constraints, see e.g., [8].

B. Optimal Policies

Let β_m^r, β_k^w, $\alpha_{k,m|l}^p$, $\alpha_{k,m|l}^r$, $\alpha_{k,m}^w$ denote the positive Lagrange multipliers associated with $C1$-$C4$, respectively. Specifically, upon defining $\kappa_{3,m} := \kappa_2^{-1} \ln(\kappa_1 / [\bar{\epsilon}_0]_m)$ and setting the derivative of the dual Lagrangian function in (3) with respect to (w.r.t.) the auxiliary variable $\bar{r}_{k,m|l}$ equal to zero, yields after tedious but straightforward manipulations the following KKT condition expressed in terms of the original variable[5]

$$r_{k,m|l(\mathbf{G})}^* = \log_2 \left(\frac{(\beta_m^{r*} + \alpha_{k,m|l}^{r*}) g_{k,m|l(\mathbf{G})}^{\min}}{\ln(2) \kappa_{3,m}} \right). \tag{4}$$

Because KKT conditions for $C3$ dictate $r_{k,m|l}^* \alpha_{k,m|l}^{p*} = 0$ [1], for $r_{k,m|l}^* > 0$, we need $\alpha_{k,m|l}^{r*} = 0$ in (4) and therefore the condition $g_{k,m|l(\mathbf{G})}^{\min} > \ln(2) \kappa_{3,m} (\beta_m^{r*})^{-1}$ has to be satisfied in order for the optimum rate loading in the region to be nonzero. Intuitively, this condition eliminates from the optimum allocation set the regions $\boldsymbol{\mathcal{R}}_{k,m|l}$ with very poor channel conditions ($\alpha_{k,m|l}^{r*} > 0$), while for the remaining regions $\alpha_{k,m|l}^{r*} = 0$. Furthermore, it is worth noting that due to the logarithmic expression of ϵ_P^{-1} [c.f. (1)], the rate loading is reminiscent of the classical capacity water-filling solution.

Before analyzing the optimality condition for $w_{k,m}(\mathbf{G})$, let us define the power cost of user m utilizing subcarrier k as

$$\mathcal{P}_{k,m}(\mathbf{G}) := \frac{(2^{r_{k,m|l(\mathbf{G})}^*} - 1) \kappa_{3,m}}{g_{k,m|l(\mathbf{G})}^{\min}} - \beta_m^{r+*} r_{k,m|l(\mathbf{G})}^*. \tag{5}$$

Supposing that $\boldsymbol{\mathcal{R}}_{k,m|l}$ is active, using (5), and differentiating the Lagrangian of (3) w.r.t. $w_{k,m}(\mathbf{G})$, we find at the optimum

$$\mathcal{P}_{k,m}(\mathbf{G}) f_{\mathbf{G}}(\mathbf{G}) + \beta_k^{w*}(\mathbf{G}) - \alpha_{k,m}^{w*}(\mathbf{G}) = 0, \ \forall \mathbf{G}, \ \forall m \in [1, M]. \tag{6}$$

It is useful to check three things: (i) the $LHS(6)$ does not depend explicitly on $w_{k,m}^*(\mathbf{G})$ but only through the associated multipliers $\beta_k^{w*}(\mathbf{G})$ and $\alpha_{k,m}^{w*}(\mathbf{G})$; (ii) the multiplier $\beta_k^{w*}(\mathbf{G})$ is common $\forall m$; and (iii) for the same subcarrier k and a given realization $\mathbf{G}$, the power cost $\mathcal{P}_{k,m}(\mathbf{G})$ is fixed and in general different for each user m. Furthermore, for each k, the KKT condition corresponding to $C4$ also dictates

$$w_{k,m}^*(\mathbf{G}) \alpha_{k,m}^{w*}(\mathbf{G}) = 0, \ \forall \mathbf{G}, \ \forall m \in [1, M]. \tag{7}$$

Proposition 1: *In the set $\{w_{k,m}^*(\mathbf{G})\}_{m=1}^{M}$, we have $w_{k,m_k}^*(\mathbf{G}) = 1$ for a unique user m_k, and $w_{k,m}^*(\mathbf{G}) = 0$ for $m \neq m_k$. Moreover, the user m_k assigned to*

[4]Although this relaxation may lead to a suboptimal solution, it leads to a less complex optimization problem that turns out to be convex. Numerical results will show that solution of the relaxed problem attains comparable performance to that reached by the optimum solution.

[5]Henceforth, x^* will denote the optimal value of x.

utilize the kth subchannel is the one that satisfies: $m_k = \arg\min_m\{\mathcal{P}_{k,m}(\mathbf{G})\}_{m=1}^M$.

Proof: Assume that $w^*_{k,m_k}(\mathbf{G}) > 0$, then $\alpha^{w*}_{k,m_k}(\mathbf{G}) = 0$ and (6) implies that $\beta^{w*}_k(\mathbf{G}) = -\mathcal{P}_{k,m_k}(\mathbf{G})f_{\mathbf{G}}(\mathbf{G})$. If also $w^*_{k,m'_k}(\mathbf{G}) > 0$ with $m'_k \neq m_k$ then $\alpha^{w*}_{k,m'_k}(\mathbf{G}) = 0$, and (6) for m'_k becomes $\mathcal{P}_{k,m_k}(\mathbf{G}) = \mathcal{P}_{k,m'_k}(\mathbf{G})\ \forall\mathbf{G}$; which is not true and thus $w^*_{k,m'_k}(\mathbf{G}) > 0$ is not either (uniqueness). If now $w^*_{k,m'_k}(\mathbf{G}) = 1$ for a $m'_k \neq m_k$, then $\beta^{w*}_k(\mathbf{G}) = -\mathcal{P}_{k,m'_k}(\mathbf{G})f_{\mathbf{G}}(\mathbf{G})$ and using (6) for m_k we can write $\alpha^{w*}_{k,m_k}(\mathbf{G}) = [\mathcal{P}_{k,m_k}(\mathbf{G}) - \mathcal{P}_{k,m'_k}(\mathbf{G})]f_{\mathbf{G}}(\mathbf{G}) \geq 0$, which is not true if $\mathcal{P}_{k,m_k}(\mathbf{G}) < \mathcal{P}_{k,m'_k}(\mathbf{G})$ (minimum). ■

Notice that if for a given $\mathbf{G}$ the minimum value of $\{\mathcal{P}_{k,m}(\mathbf{G})\}_{m=1}^M$ is attained by more than one user, any arbitrary time sharing of the subcarrier k among them is optimum; or we can simply pick one of them at random without altering the power cost. Finally, if there exists a realization $\mathbf{G}$ such that $\mathcal{P}_{k,m}(\mathbf{G}) > 0\ \forall m$, then since $\beta^{w*}_k(\mathbf{G}) \geq 0$ it follows from (6) that $\alpha^{w*}_{k,m}(\mathbf{G}) \neq 0,\ \forall m$; and the optimal solution will not allocate this subcarrier to any user. Therefore, introducing a fictitious $\mathcal{P}_{k,0}(\mathbf{G}) = 0\ \forall k$ and $\mathbf{G}$, we can express $w^*_{k,m}(\mathbf{G})$ in compact form using the indicator function as

$$w^*_{k,m}(\mathbf{G}) = \mathbf{I}_{\{m=\arg\min_{m'}\{\mathcal{P}_{k,m'}(\mathbf{G})\}_{m'=0}^M\}}. \quad (8)$$

Since so far we have obtained conditions that the optimal $r^*_{k,m|l}$ and $w^*_{k,m}(\mathbf{G})$ should satisfy, what remains is a condition for $p^*_{k,m|l}$. Although in principle $p^*_{k,m|l}$ can be calculated by substituting $r^*_{k,m|l}$ and $w^*_{k,m}(\mathbf{G})$ into (2), recall that to render the optimization problem tractable we had enforced via (2) an instantaneous BER constraint that is a more restrictive than the original average BER constraint. Capitalizing on this fact, a different approach to obtain the power per region (denoted by $p^{+*}_{k,m|l}$), given $r^*_{k,m|l}$ and $w^*_{k,m}(\mathbf{G})$, consists of finding $p^{+*}_{k,m|l}$ to satisfy

$$\begin{aligned}\mathbb{E}_{\mathbf{G}\in\boldsymbol{\mathcal{R}}_{k,m|l}}[\epsilon(p^{+*}_{k,m|l}, g_{k,m}, r^*_{k,m|l})w^*_{k,m}(\mathbf{G})] \\ = [\bar{\boldsymbol{\epsilon}}_0]_m \mathbb{E}_{\mathbf{G}\in\boldsymbol{\mathcal{R}}_{k,m|l}}[w^*_{k,m}(\mathbf{G})]. \quad (9)\end{aligned}$$

Using this approach we, guarantee that the expected BER averaged over all channel realizations inside the region $\mathcal{R}_{k,m|l}$ satisfies the pre-specified requirement $[\bar{\boldsymbol{\epsilon}}_0]_m$. Since $p^{+*}_{k,m|l} < p^*_{k,m|l}$, we will rely on (9) to calculate the final power loading. Note also that, although (9) involves an integration and $p^{+*}_{k,m|l}$ can not be found in closed-form, the monotonicity of the exponential inside the instantaneous BER in (1) allows obtaining $p^{+*}_{k,m|l}$ through line search (using e.g., the bisection method).

The following algorithm summarizes the main steps for finding the optimal allocation of $\mathbf{R}^*$, $\mathbf{W}^*$ and $\mathbf{P}^*$.

Algorithm 1: *Resource Allocation*

(S1.0) Select a small positive number δ, and initialize $\boldsymbol{\beta^r}$ using an arbitrary non-negative vector.

(S1.1) For each (k, m, l) triplet per iteration:
(S2.1.1) Use $[\boldsymbol{\beta^r}]_m$ to determine $r_{k,m|l}$ via (4)($\alpha^r_{k,m|l} = 0$).
(S2.1.2) Use $[\boldsymbol{\beta^r}]_m$ to determine $w_{k,m}(\mathbf{G})$ via (8).

(S1.2) Check $C1$ (3) $\forall m$; if $|LHS(C1)| < \delta[\bar{\mathbf{r}}_0]_m\ \forall m$, then go to (S1.3); otherwise, increase $[\boldsymbol{\beta^r}]_m$ for the users m whose $LHS(C1) > 0$; decrease $[\boldsymbol{\beta^r}]_m$ for the users m whose $LHS(C1) < 0$; and go to (S1.1).

(S1.3) Once $\mathbf{R}^*, \boldsymbol{\beta}^{r*}, \mathbf{W}^*$ are obtained, use (9) to calculate the finally allocated power.

The convexity of (3) enables efficient methods to update $\boldsymbol{\beta^r}$ [1]. For example, we can set the initial value of $\boldsymbol{\beta^r}$ equal to any small number and update each component $[\boldsymbol{\beta^r}]_m$ separately $\forall m$ by fixing $[\boldsymbol{\beta^r}]_{m'},\ \forall m' \neq m$ from the previous iteration. The adaptation of each $[\boldsymbol{\beta^r}]_m$ is then performed using line search until the rate constraint for the mth user is tightly satisfied. This simple algorithm has guaranteed convergence and facilitates computation *distributed* across users.

C. Codeword Structure

Given the quantizer design, we developed so far resource allocation policies to assign rate, power and subcarriers across users. Once the quantizer and resource allocation strategy are designed, the AP quantizes each fading state and feeds back a codeword that identifies the user-subcarrier assignment and the region index each subchannel falls into per fading realization $\mathbf{G}$. Based on this form of Q-CSIT, each user node is informed about its own subset of subcarriers (if any) and relies on the region indices to retrieve the corresponding power and rate levels from a lookup table. The following proposition describes the construction of this codeword.

Proposition 2: *Given the quantizer design and the optimal allocation parameters* ($\mathbf{P}^*$, $\mathbf{R}^*$, $\boldsymbol{W}^*(\boldsymbol{\beta}^{r*})$) *returned by Algorithm 1, the AP broadcasts to the users the codeword* $\mathbf{c}^*(\mathbf{G}) = [\mathbf{c}^*_1(\mathbf{G}), \ldots, \mathbf{c}^*_K(\mathbf{G})]$ *specifying the optimal resource allocation for the current fading state, where* $\mathbf{c}^*_k(\mathbf{G}) = [m^*_k(\mathbf{G}), l^*_k(\mathbf{G})]^T$ *is determined* $\forall k$ *as:*

1) $m^*_k(\mathbf{G}) = \arg\min_m\{\mathcal{P}_{k,m}(\mathbf{G}, \mathbf{P}^*, \mathbf{R}^*, \boldsymbol{\beta}^{r*}, \{\beta^{\epsilon*}\}_{l=1}^{L_{k,m}})\}_{m=1}^M$ *(pick randomly any user* m^*_k *when multiple minima occur); and*
2) $l^*_k(\mathbf{G}) = \{\ l \mid \mathbf{G} \in \boldsymbol{\mathcal{R}}_{k,m^*_k(\mathbf{G}),l},\ \ l = 1, \ldots, L_k\}$.

The structure of $\mathbf{c}^*(\mathbf{G})$ in Proposition 2 encodes information pertinent to each subcarrier (namely, its region and assigned user) which is more efficient in terms of the number of feedback bits relative to encoding each user's individual information, yielding a codeword length of $\left\lceil \sum_{k=1}^K \log_2\left(\sum_{m=1}^M (L_{k,m})\right)\right\rceil$ bits.

We conclude this section by emphasizing that $\mathbf{R}$ (and thus $\mathbf{P}$) in (3) are involved only in average quantities. Hence, $\mathbf{R}^*$ and $\mathbf{P}^*$ are computed off-line and only the subcarrier-user assignment (involved in instantaneous constraints) and the indexing of the corresponding entries of these matrices need to be fed back on-line. Thus, almost all the complexity is carried out off-line (Algorithm 1), while only a light computation (Proposition 2) has to be carried out on-line.

IV. Quantizer Design

In the previous section we addressed our objective (G2) to derive optimum subcarrier, rate, and power allocation policies assuming the quantization regions, $\boldsymbol{\mathcal{R}}_{k,m|l}$, are given. In this section, we will address (G1) by deriving a quantizer that enforces equally probable quantization regions.

A. Equally Probable Region Quantizer

To design the quantizer, we first solve the optimal resource allocation problem supposing CSI is available without quantization (i.e., $L_{k,m} = \infty$), and subsequently calculate $\tau_{k,m|l}$ to satisfy

$$\int_{\tau_{k,m|l}}^{\tau_{k,m|l+1}} \Pr(w_{k,m} = 1, g_{k,m}) dg_{k,m} = \int_{\tau_{k,m|1}}^{\tau_{k,m|L_{k,m}}} \Pr(w_{k,m} = 1, g_{k,m}) dg_{k,m}/L_{k,m}, \quad (10)$$

with $\tau_{k,m|1} = 0$ and $\tau_{k,m|L_{k,m}+1} = \infty$. If the joint probabilities can be computed, solving (10) yields thresholds $\{\tau_{k,m|l}\}_{l=1}^{L_{k,m}}$ per subcarrier k and user m that divide the joint probability $\Pr(w_{k,m} = 1, g_{k,m})$ into regions of equal area; hence the term *equally probable region quantizer*. Intuitively speaking, this quantizer design tries to maximize the entropy in the feedback link.[6]

To evaluate $\Pr(w_{k,m} = 1, g_{k,m})$ needed in (10), we apply Bayes' rule to re-write it as $\Pr(w_{k,m} = 1|g_{k,m}) f_{g_{k,m}}(g_{k,m})$ and recall that $f_{g_{k,i}}(g_{k,m})$ is known per (as2). To calculate $\Pr(w_{k,m} = 1|g_{k,m})$, we will need to first solve the optimal resource allocation problem assuming no quantization. Clearly, as $L_{k,m} \to \infty$, we have $g_{\min} \to g$ in (5). Letting $\mathcal{P}_{k,m}^{\infty}$ and $\beta_m^{r\infty*}$ denote, respectively, the cost indicator and the Lagrange multiplier when $L_{k,m} \to \infty$, we can write

$$\mathcal{P}_{k,m}^{\infty}(g_{k,m}) = \frac{\beta_m^{r\infty*}}{\ln(2)} - \frac{\kappa_{3,m}}{g_{k,m}} - \beta_m^{r\infty*} \log_2\left(\frac{g_{k,m}\beta_m^{r\infty*}}{(\kappa_{3,m}\ln(2))}\right). \quad (11)$$

Because equation (8) establishes that $w_{k,m}(\mathbf{G}) = \mathbf{I}_{\{m=\arg\min_{m'}\{\mathcal{P}_{k,m'}^{\infty}(\mathbf{G})\}_{m'=0}^{M}\}}$, if $\mathcal{P}_{k,m}^{\infty}(g_{k,m}) < \mathcal{P}_{k,0}^{\infty} = 0$, then $g_{k,m} > \ln(2)\kappa_{3,m}/\beta_m^{r\infty*}$ is a necessary condition for the user m to be active. Taking also into account that channels of different users are uncorrelated [cf. (as1)], we can write

$$\Pr(w_{k,m} = 1|g_{k,m}) = \mathbf{I}_{\{g_{k,m} > \frac{\ln(2)\kappa_{3,m}}{\lambda_m^{r*}}\}} \times \prod_{\mu=1,\mu\neq m}^{M} \Pr(\mathcal{P}_{k,m}^{\infty} < \mathcal{P}_{k,\mu}^{\infty} \mid g_{k,m}). \quad (12)$$

Interestingly, for the active users it holds that $\mathcal{P}_{k,m}^{\infty}(g_{k,m}) = \mathbf{I}_{\{g_{k,m} > \frac{\ln(2)\kappa_{3,m}}{\beta_m^{r\infty*}}\}}\left(\frac{\beta_m^{r\infty*}}{\ln(2)} - \frac{\kappa_{3,m}}{g_{k,m}} - \beta_m^{r\infty*}\log_2\left(\frac{g_{k,m}\beta_m^{r\infty*}}{(\ln(2)\kappa_{3,m})}\right)\right)$, and therefore $\frac{\partial\mathcal{P}_{k,m}^{\infty}(g_{k,m})}{\partial g_{k,m}} = \mathbf{I}_{\{g_{k,m} > \frac{\ln(2)\kappa_{3,m}}{\beta_m^{r\infty*}}\}}\left(\frac{\kappa_{3,m}}{g_{k,m}} - \frac{\beta_m^{r\infty*}}{\ln(2)}\right)\frac{\kappa_{3,m}}{g_{k,m}} \leq 0$, which implies that $\mathcal{P}_{k,m}^{\infty}(g_{k,m})$ is monotonically decreasing. Therefore, we can find unique channel gains $\gamma_{k,\mu}$, $\forall\mu \neq m$, such that

$$\mathcal{P}_{k,m}^{\infty}(g_{k,m}) = \mathcal{P}_{k,\mu}^{\infty}(\gamma_{k,\mu}). \quad (13)$$

It is then clear that $\mathcal{P}_{k,m}^{\infty}(g_{k,m}) \leq \mathcal{P}_{k,\mu}^{\infty}(g_{k,\mu})$ if $g_{k,\mu} \leq \gamma_{k,\mu}$, and $\mathcal{P}_{k,m}^{\infty}(g_{k,m}) > \mathcal{P}_{k,\mu}^{\infty}(g_{k,\mu})$ if $g_{k,\mu} > \gamma_{k,\mu}$. And consequently, $\Pr(\mathcal{P}_{k,m}^{\infty} < \mathcal{P}_{k,\mu}^{\infty} \mid g_{k,m}) = \Pr(g_{k,\mu} < \gamma_{k,\mu} \mid g_{k,m})$.

[6]The simplicity and efficiency of using an equally probable quantizer were first pointed out in [7] for the basic case of a single user and subcarrier. We here largely expand the scope of this quantizer and apply it to the challenging multiple-user/multiple-subcarrier scenario.

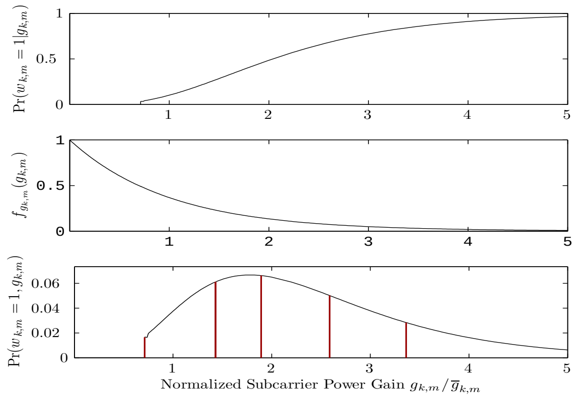

Fig. 2. Equally Probable Regions Quantizer ($K = 64$, $M = 6$, $L = 5$).

Excluding the case $\mathcal{P}_{k,m}^{\infty}(g_{k,m}) > 0$, which amounts to $w_{k,m} = 0$ and solving (13) w.r.t. $\gamma_{k,\mu}$ yields

$$\gamma_{k,\mu}(g_{k,m}) = -\frac{\kappa_{3,\mu}\ln(2)/\beta_\mu^{r\infty*}}{f^W\left[-2^{-\frac{\kappa_{3,m}}{g_{k,m}\beta_\mu^{r\infty*}}} e^{-1}\left(\frac{e\ln(2)\kappa_{3,m}}{g_{k,m}\beta_m^{r\infty*}}\right)^{\frac{\beta_m^{r\infty*}}{\beta_\mu^{r\infty*}}}\right]}, \quad (14)$$

where $f^W[x] = y$ is the real-valued Lambert's f^W function which solves the equation $ye^y = x$ for $-1 \leq y \leq 0$ and $-1/e \leq x \leq 0$ [3].

Using (12)-(14), we express $\Pr(w_{k,m} = 1, g_{k,m})$ in (10) as

$$\Pr(w_{k,m} = 1, g_{k,m}) = \mathbf{I}_{\{g_{k,m} > \frac{\ln(2)\kappa_{3,\mu}}{\beta_m^{r\infty*}}\}} \frac{e^{-g_{k,m}/\bar{g}_{k,m}}}{\bar{g}_{k,m}} \times \prod_{\mu=1,\mu\neq m}^{M}\left(1 - e^{-\gamma_{k,\mu}(g_{k,m})/\bar{g}_{k,\mu}}\right). \quad (15)$$

Since (15) depends on $\{\beta_\mu^{r\infty*}\}_{\mu=1}^{M}$, we need to solve the optimal allocation problem as $L_{k,m} \to \infty$. The thresholds $\{\tau_{k,m|l}\}_{l=2}^{L_{k,m}}$ are obtained by solving (10) using a line search. Notice that we can also take advantage of the condition $\mathcal{P}_{k,m}^{\infty}(g_{k,m}) < 0$, by setting $\tau_{k,m,2} \geq \kappa_{3,m}\ln(2)/\beta_m^{r\infty*}$.

An example illustrating this quantizer is given in Fig. 2 which depicts $\Pr(w_{k,m} = 1|g_{k,m})$, $f_{g_{k,m}}(g_{k,m})$ and $\Pr(w_{k,m} = 1, g_{k,m})$ versus $g_{k,m}/\bar{g}_{k,m}$ for $M = 6$, $L_{k,m} = 5$, equal average subcarrier gains, and equal rate constraints. The first subplot in this figure, $\Pr(w_{k,m} = 1|g_{k,m})$, reveals that the better the channel the more likely the corresponding user is to be selected. Coupling this observation with the exponential behavior of $f_{g_{k,m}}(g_{k,m})$ in the second subplot, the bell-ring characteristic of the joint PDF, $\Pr(w_{k,m} = 1, g_{k,m})$, results naturally in the third subplot (after pair-wise multiplication of the functions in the first two subplots), where the quantization thresholds (and regions) resulting from (10) are also identified.

Remark 2: With the method presented in this section, calculation of $\tau_{k,m|l}$ based on (10) has to be executed only once (to solve (G1)). With the thresholds available, the resource allocation can be easily obtained through Algorithm 1 (to solve (G2)). Numerical results in the next section will show that this low complexity non-iterative design exhibits power consumption similar to that of the P-CSIT solution.

TABLE I

TOTAL AVERAGE TRANSMIT-POWER (IN dB_W) FOR P-CSIT, Q-CSIT-EP AND Q-CSIT-UN SCHEMES.

CASE	Q-CSIT-UN	Q-CSIT-EP	P-CSIT
Reference Case	36.1	28.6	28.2
$\bar{\epsilon}_0 = 10^{-4}$	37.2	30.5	30.1
$\bar{\mathbf{r}} = [30,\ 30,\ 30]^T$	29.0	23.8	23.2
$K = 128$	32.2	27.0	26.3
$M = 6$	43.5	36.8	36.2
$\overline{\text{SNR}} = 0$	39.2	31.7	31.2
$\overline{\text{SNR}} = 10$	29.5	21.7	21.3

TABLE II

TOTAL AVERAGE TRANSMIT-POWER (IN dB_W) AS $L_{k,m}$ VARIES.

$L_{k,m}$	2	3	4	5	∞ (P-CSIT)
$\overline{P}$	27.9	25.8	25.1	24.7	24.2

V. NUMERICAL EXAMPLES

To numerically test our power-efficient designs, we consider an adaptive OFDMA system with $M = 3$ users, $K = 64$ subcarriers, noise power per user and subcarrier at 0 dB_W, $L_{k,m} = 5$ regions (i.e., 4 active regions) per subcarrier, and $[\bar{\boldsymbol{\epsilon}}_0]_m = \bar{\epsilon}_0 = 10^{-3}\ \forall m$. The average signal-to-noise-ratio ($\overline{\text{SNR}}$) considered was 3 dB; and three uncorrelated Rayleigh taps were simulated per user.

Test Case 1 *(Comparison of allocation schemes):* For comparison purposes we will contrast the power consumption of our proposed quantization and allocation scheme (denoted by Q-CSIT-EP) with the power consumption of the allocation based on P-CSIT (that represents a lower bound) as well as with the power consumption of a heuristic scheme based on Q-CSIT with uniform quantization and subcarrier allocation but optimum rate and power loading (denoted by Q-CSIT-UN). Numerical results assessing the performance of P-CSIT, Q-CSIT-EP and Q-CSIT-UN schemes over a wide range of parameter values are summarized in Table I, where the reference case is $K = 64$, $M = 3$, $\bar{\mathbf{r}}_0 = [60,\ 60,\ 60]^T$, $\overline{\text{SNR}} = 3\ dB$; and the "column case" entails only a single variation w.r.t. the reference case. The striking observation here is the almost equivalent performance of Q-CSIT-EP and P-CSIT schemes. These results certify the usefulness of the proposed quantizer based on equally probable regions and validates the optimality of Algorithm 1. Moreover, the power loss of at least 6 dB with respect to Q-CSIT-UN shows the important role of subcarrier allocation and quantization in terms of minimizing the transmit-power.

Test Case 2 *(Number of quantization regions):* Table II lists the average transmit-power when varying the number of regions per subcarrier for $\bar{\mathbf{r}}_0 = [20,\ 40,\ 60]^T$. Notice that usually the number of active regions is equal to $L_{k,m} - 1$, since the optimum solution yields one outage region (for the cases when the channel gain is very poor). Simulation results demonstrate that our proposed equally probable quantization and Algorithm 1 used together lead to a power loss no greater than 3-5 dB with respect to the P-CSIT case ($L_{k,m} = \infty$). Moreover, the resulting power gap shrinks as the number of regions increases reaching a power loss of approximately only 0.5 dB in the case of four active regions.

VI. CONCLUDING SUMMARY AND FUTURE RESEARCH

Based on Q-CSIT, we devised a power efficient OFDMA scheme under prescribed individual average rate and BER constraints. In this setup, an access point quantizes the subcarrier gains and feeds back to the users a codeword conveying the optimum power, rate, and subcarrier policy. The resulting near-optimal transceivers are attractive because they only incur a power loss as small as 1 dB relative to the benchmark design based on perfect CSIT which requires feedback information which may be unrealistic in wireless systems.

Our novel design separates resource allocation from channel quantization. The first was obtained as the solution of a convex constrained optimization problem, while the second was designed to ensure equally probable regions per user and subcarrier. We ended up with a lightweight resource allocation protocol where both rate and power are available at the transmitter through a lookup table and only the subcarrier assignment needs be determined on-line.

To build on the presented framework, future directions will exploit the possible correlation across subcarriers to group subcarriers and then index each group.[7]

REFERENCES

[1] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1999.

[2] M. Cho, W. Seo, Y. Kim, D. Hong, "A Joint Feedback Reduction Scheme Using Delta Modulation for Dynamic Channel Allocation in OFDMA Systems", *Proc. of IEEE Int. Symposium on Personal, Indoor and Mobile Radio Commun.*, Berlin, Germany, vol. 4, pp. 2747–2750, Sep. 2005.

[3] 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*, pp. 5:329–359, 1996.

[4] A. J. Goldsmith and S. -G. Chua, "Variable-rate variable-power M-QAM for fading channels," *IEEE Trans. on Commun.*, vol. 45, pp. 1218–1230, Oct. 1997.

[5] J. H. Kwon, D. Rhee, I. M. Byun, Y. Whang, K. S. Kim, "Adaptive Modulation Technique with Partial CQI for Multiuser OFDMA Systems," *Proc. of Int. Conf. on Advanced Commun. Technology*, Phoenix Park, Korea, vol. 2, pp. 1283–1286, Feb. 2006.

[6] A. Lapidoth and S. Shamai, "Fading channels: how perfect need 'perfect side information' be?," *IEEE Trans. on Info. Theory*, vol. 48, no. 5, pp. 1118-1134, May 2002.

[7] A. G. Marques, F. F. Digham, and G. B. Giannakis, "Optimizing power efficiency of OFDM using quantized channel state information," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1581–1592, Aug. 2006.

[8] H. Qian, C. Xiao, N. Chen, and G. T. Zhou, "Dynamic selected mapping for OFDM," in *Proc. of Intl. Conf. on Acoustics, Speech, and Signal Processing*, Philadelphia, PA, pp. 325–328, Mar. 2005.

[9] C.Y. Wong, R.S. Cheng, K.B. Lataief, R.D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 10, pp. 1747–1758, Oct. 1999.

[10] S. Zhou and B. Li, "BER criterion and codebook construction for finite-rate precoded spatial multiplexing with linear receivers," *IEEE Trans. on Signal Proc.*, vol. 54, no. 5, pp. 1653–1665, May 2006.

[7]The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U. S. Government.