

Throughput Maximization in Cognitive Radio Under Peak Interference Constraints With Limited Feedback

YuanYuan He, *Student Member, IEEE*, and Subhrakanti Dey, *Senior Member, IEEE*

Abstract—A spectrum-sharing scenario in a cognitive radio (CR) network where a secondary user (SU) shares a narrowband channel with N primary users (PUs) is considered. We investigate the SU ergodic capacity maximization problem under an average transmit power constraint on the SU and N individual peak interference power constraints at each primary-user receiver (PU-Rx) with various forms of imperfect channel-state information (CSI) available at the secondary-user transmitter (SU-Tx). For easy exposition, we first look at the case when the SU-Tx can obtain perfect knowledge of the CSI from the SU-Tx to the secondary-user receiver link, which is denoted as g_1 , but can only access quantized CSI of the SU-Tx to PU-Rx links, which is denoted as g_{0i} , $i = 1, \dots, N$, through a limited-feedback link of $B = \log_2 L$ b. For this scenario, a locally optimum quantized power allocation (codebook) is obtained with quantized g_{0i} , $i = 1, \dots, N$ information by using the Karush–Kuhn–Tucker (KKT) necessary optimality conditions to numerically solve the nonconvex SU capacity maximization problem. We derive asymptotic approximations for the SU ergodic capacity performance for the case when the number of feedback bits grows large ($B \rightarrow \infty$) and/or there is a large number of PUs ($N \rightarrow \infty$) that operate. For the interference-limited regime, where the average transmit power constraint is inactive, an alternative locally optimum scheme, called the quantized-rate allocation strategy, based on the quantized-ratio $g_1/\max_i g_{0i}$ information, is proposed. Subsequently, we relax the strong assumption of full-CSI knowledge of g_1 at the SU-Tx to imperfect g_1 knowledge that is also available at the SU-Tx. Depending on the way the SU-Tx obtains the g_1 information, the following two different suboptimal quantized power codebooks are derived for the SU ergodic capacity maximization problem: 1) the power codebook with noisy g_1 estimates and quantized g_{0i} , $i = 1, \dots, N$ knowledge and 2) another power codebook with both quantized g_1 and g_{0i} , $i = 1, \dots, N$ information. We emphasize the fact that, although the proposed algorithms result in locally optimum or strictly suboptimal solutions, numerical results demonstrate that they are extremely efficient. The efficacy of the proposed asymptotic approximations is also illustrated through numerical simulation results.

Index Terms—Cognitive radio, interference constraints, limited feedback.

Manuscript received April 29, 2011; revised September 17, 2011 and November 20, 2011; accepted January 3, 2012. Date of publication February 3, 2012; date of current version March 21, 2012. The review of this paper was coordinated by Prof. B. Hamdaoui.

The authors are with the Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville, Vic. 3010, Australia (e-mail: y.he2@student.unimelb.edu.au; sdey@unimelb.edu.au).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TVT.2012.2186597

I. INTRODUCTION

RADIO spectrum is a limited precious natural resource. As the number of wireless communication systems and services grows, traditional spectrum allocation policies that are employed by regulatory authorities become inefficient. To avoid interference, these policies grant the license-holding owner an exclusive right to access the allocated frequency band [1], thus resulting in the severe scarcity of vacant spectrum. However, recent measurements that were performed by the Federal Communications Commission (FCC) revealed that several portions of spectrum are largely under utilized or even unoccupied. This condition led to the idea of cognitive radio (CR) technology, originally introduced by Mitola [2], which promises a dramatic improvement of the efficiency of spectral utilization.

The key idea behind CR is that an unlicensed/secondary user (SU) is allowed to communicate over the frequency band originally licensed to a primary user (PU), as long as the transmission of SUs does not generate an unfavorable impact on the operation of PU in that band. Effectively, the following three categories of CR network paradigms have been proposed: 1) interweave, 2) overlay, and 3) underlay [3]. In the underlay systems, which is the focus of this paper, the SU can simultaneously coexist with the PU, but the transmitted power of the SU should properly be controlled to ensure that the resulting interference does not degrade the received signal quality of the PU to an undesirable level [4] by imposing the so-called interference temperature [1] constraints at the PU, e.g., the average interference power (AIP) or peak interference power (PIP) constraint. This type of CR scenario is also known as the spectrum-sharing [1] model.

Information-theoretic capacity is a very important criterion for analyzing the performance limits of CR systems. In [6], the authors first studied the behavior of the capacities of different additive white Gaussian noise (AWGN) channels under average received-power constraints or, equivalently, AIP constraints at the primary-user receiver (PU-Rx) and showed that, in non-fading AWGN channels, the capacity performances under the transmit and receive power constraints are very similar. In [1], the authors investigated the ergodic capacity of such a dynamic spectrum-sharing model in a narrowband under either the AIP or the PIP constraint at the PU-Rx in various fading environments, illustrating that, in a fading environment, spectrum access opportunity for the SU significantly increases compared to the AWGN case. In [7], the authors extended the work in [1] to an asymmetric fading environment. In [8], the authors studied optimum power allocation for three different capacity notions under both AIP and PIP constraints. The optimal power

allocation strategies for maximizing secondary ergodic capacity and outage capacity under various combinations of secondary transmit power constraints and interference constraints were derived in [4]. For results on resource allocation in CR networks with various other forms of interference constraints, see [5].

Most of the aforementioned results assume perfect knowledge of the full CSI of the fading channels at the secondary-user transmitter (SU-Tx), which is very difficult to implement in practice, particularly for the channel information from the SU-Tx to the PU-Rx. A few recent papers have addressed this concern by investigating capacity analysis with imperfect CSI. The effect of imperfect channel information for the secondary to primary channels under the AIP or PIP constraint has been investigated in [9] by assuming perfect knowledge of the CSI from SU-Tx to the secondary-user receiver (SU-Rx) channel, considering the channel information from the SU-Tx to the PU-Rx as a noisy estimate of the true CSI, and employing the so-called truncated channel inversion with fixed rate transmission policy. In [10], the authors proposed a practical design paradigm for cognitive beamforming based on finite-rate cooperative feedback from the PU-Rx to the SU-Tx and cooperative feedforward from the SU-Tx to the PU-Rx. In another recent work [12], the authors considered imperfect CSI for the SU-Tx to PU-Rx channel in the form of noisy channel estimates (ranging from near-perfect estimates to seriously flawed estimates) and studied the effect of using a midrise uniformly quantized CSI, assuming full knowledge of the SU-Tx to SU-Rx channel at the SU-Tx. A robust cognitive beamforming scheme was also analyzed in [11], where full CSI on the SU-Tx to SU-Rx channel was assumed, and the imperfect channel information on the SU-Tx to the PU-Rx was modeled using an uncertainty set. Related work also includes [13], where a stochastic resource allocation scheme for a multiuser CR network with a probabilistic primary interference constraint is considered with imperfect information on the primary activity, and [14], where separate channel quantization (SCQ) of secondary and primary links is considered to maximize secondary spectral efficiency over fading channels. Finally, [15] studies the issue of channel quantization for resource allocation through the framework of utility maximization in CR networks based on orthogonal frequency-division multiple access but does not investigate the joint channel partitioning and rate/power codebook design problem. Indeed, the lack of a rigorous and systematic design methodology for quantized resource allocation algorithms in the context of CR networks formed the key motivation for our recent work [19], where we addressed an SU ergodic capacity maximization problem in a wideband spectrum-sharing scenario with quantized information about the vector channel space that involves the SU-Tx to SU-Rx channel and the SU-Tx to PU-Rx channel over all bands. In [19], we considered an average transmit power (ATP) constraint at the SU-Tx and an average interference constraint at the PU-Rx. However, the techniques that were used for designing an optimal quantized power codebook in [19] cannot be directly extended to the case of peak interference constraints at the PU-Rx, which is the topic of interest in this paper.

In this paper, we aim at designing locally optimum power allocation algorithms in the narrowband spectrum-sharing system

of an *infrastructure-based* CR network similar to [19]. In this setting, an SU communication uplink shares a randomly fading frequency band with N PUs. The objective is to maximize the SU throughput (ergodic capacity), under an ATP constraint on the SU-Tx and the N individual peak interference constraints at each PU-Rx [contained within the primary-user base station (PU-BS)], with various forms of imperfect CSI knowledge at the SU-Tx. We first consider the throughput maximization problem with full knowledge of the CSI for the SU-Tx to SU-Rx [contained within the secondary-user base station (SU-BS)] link, which is denoted as g_1 , and quantized information about the CSI from the SU-Tx to each PU-Rx link, which is denoted as $\mathbf{g}_0 = \{g_{01}, \dots, g_{0N}\}$ and is available at the SU-Tx through a limited-feedback link of $B = \log_2 L$ b. We derive the structure of the optimal quantization regions, and a locally optimal power codebook is then obtained by solving the non-convex throughput maximization problem using the associated Karush–Kuhn–Tucker (KKT) necessary optimality conditions. Asymptotic approximations for the SU ergodic capacity when B and/or N grows large are derived. For the interference-limited regime, where only the PIP constraints are active, a locally optimum quantized transmission rate codebook is also designed by quantizing the ratio $g_1 / \max_i g_{0i}$. To this end, similar to [19], we assume the availability of a central controller, called the CR network manager, which can obtain full CSI of g_1 and g_{0i} , $i = 1, 2, \dots, N$, from the SU-BS and the PU-BS, respectively, through (possibly) fiber-optic backhaul links. This controller is essentially needed for assigning the joint real-time channel information $(g_1, g_{01}, \dots, g_{0N})$ to the optimal channel partition (designed offline) so that the corresponding power codebook index can be fed back to the SU-Tx (or the SU-Rx) through the limited-feedback link. Further details on the assumption with regard to the availability of a CR network manager, its justification, and the benefits that it brings can be found in [19]. We also investigate the combined effect of imperfect g_1 and \mathbf{g}_0 knowledge at the SU-Tx in designing a locally optimal power codebook. When a noisy estimate of g_1 is available together with quantized \mathbf{g}_0 at the SU-Tx, it can be shown that it is not possible to guarantee that the actual PIP constraints will be satisfied with probability one, particularly if the channel estimation error has an unbounded support. Thus, a more appropriate approach for this case is to allow the PIP constraints to be violated with a certain small probability, called the interference violation probability (IVP). The relationship between SU capacity loss due to the effect of noisy estimated g_1 and the IVP is studied. For the case where both quantized g_1 and \mathbf{g}_0 are available at the SU-Tx, due to the difficulty and complexity of the associated optimal quantized power allocation (QPA) analysis, we design two different types of suboptimal quantized power codebooks. The efficacy of the various proposed algorithms and asymptotic approximations is evaluated through numerical simulations, which illustrate that, in general, a small number (4–6) of feedback bits is often enough to achieve SU ergodic capacity very close to the full-CSI-based performance.

The rest of this paper is organized as follows. Section II describes the system model and briefly describes the optimal power allocation scheme for the throughput maximization

problem with full-CSI assumption. In Section III, we design a locally optimal power codebook for the throughput maximization problem with perfect g_1 information and quantized \mathbf{g}_0 knowledge at the SU-Tx. The asymptotic SU ergodic capacity performance analysis for a large number of quantization levels and/or a large number of PUs is provided. For the interference-limited regime, a locally optimum quantized rate allocation (QRA) scheme for solving the throughput maximization problem is studied. In Section IV, the throughput maximization problem with both imperfect g_1 and \mathbf{g}_0 information at the SU-Tx is investigated. A discussion on how we can extend the analysis to the case of multiple SUs is provided in Section V. Numerical results are presented in Section VI, and Section VII contains some concluding remarks. All proofs (unless otherwise specified) are relegated to the Appendices.

II. SYSTEM MODEL AND PROBLEM FORMULATION

We consider an infrastructure-based spectrum-sharing scenario where a SU communicates to its base station using a narrowband randomly fading channel shared with multiple PUs for transmission. Regardless of the ON/OFF status of PU_s , the SU is allowed to use the band that is licensed to PU_1, \dots, PU_N , as long as the impact of the secondary transmission does not substantially degrade the received signal quality of the PU_s . Let $g_1 = |h_1|^2$ and $g_{0i} = |h_{0i}|^2$ ($i = 1, \dots, N$) denote the non-negative real-valued instantaneous channel power gains for the links from the SU-Tx to the secondary receiver (SU-Rx) and from the SU-Tx to the receiver of PU_i ($i = 1, \dots, N$), respectively, where h_1 and h_{0i} are the corresponding complex channel amplitude gains. These channels are assumed to be Rayleigh block-fading channels such that all g_1 and g_{0i} ($i = 1, \dots, N$) are statistically and mutually independent and, without loss of generality (*w.l.o.g.*), are exponentially distributed with unity mean. Similarly, additive noises for each channel are independent Gaussian random variables with zero mean and unit variance *w.l.o.g.* Note that extensions to other distributions such as Nakagami and Rician can also be handled by our techniques. For analytical simplicity, the interference from the primary transmitter to the SU-Rx is neglected following previous work such as [1] and [4]. This assumption is justified when either the SU is outside the PU's transmission range or the SU-Rx is equipped with interference cancellation capability, particularly when the PU signal is strong.

Given channel realization $\mathbf{g}_0 \triangleq \{g_{01}, \dots, g_{0N}\}$ and g_1 , assume that the channel-state information (CSI; \mathbf{g}_0, g_1) is available at the SU-Tx and the power that is allocated at the SU-Tx is represented by $p(\mathbf{g}_0, g_1)$; then, the ergodic capacity of the SU for this spectrum-sharing system can be expressed as

$$C = E [\log (1 + g_1 p(\mathbf{g}_0, g_1))] \quad (1)$$

where \log represents the natural logarithm. One common way of protecting the PU's received signal quality is by imposing either an AIP or a PIP constraint at the PU-Rx. We studied the optimization problem of maximizing the SU capacity under both the ATP constraint at the SU and AIP constraints at PUs with quantized CSI in [19]. Although the AIP constraint

is more favorable, particularly in the context of transmission over fading channels [16], there are other applications where it is desirable to impose a PIP constraint [7]. Thus motivated, we consider the following optimal power allocation problem that maximizes the ergodic capacity of SU in a narrowband spectrum sharing with multiple PUs, under an ATP constraint at the SU-Tx and a PIP constraint at each PU_i -Rx, given by

$$\begin{aligned} \max_{p(\mathbf{g}_0, g_1) \geq 0} \quad & E [\log (1 + g_1 p(\mathbf{g}_0, g_1))] \\ \text{s. t.} \quad & E [p(\mathbf{g}_0, g_1)] \leq P_{av} \\ & g_{0i} p(\mathbf{g}_0, g_1) \leq Q_{pk} \quad \forall i \quad \text{almost surely} \quad (2) \end{aligned}$$

where P_{av} is the ATP of the SU, and Q_{pk} is the maximum peak interference power tolerated by each PU-Rx. It is easy to show that the aforementioned PIP constraints can be reformulated as [1], [7]

$$p(\mathbf{g}_0, g_1) \leq \frac{Q_{pk}}{\max_i g_{0i}} \quad i = 1, \dots, N. \quad (3)$$

For the case of $N = 1$, the optimal power allocation for (2), assuming that full CSI is available at the SU-Tx, can be found in [4]. A trivial extension of this result for $N > 1$ shows that the optimal solution for (2) with full CSI is given by

$$p(\mathbf{g}_0, g_1) = \begin{cases} 0, & g_1 \leq \lambda_f \\ \frac{1}{\lambda_f} - \frac{1}{g_1}, & g_1 > \lambda_f, g_m < \frac{Q_{pk}}{\frac{1}{\lambda_f} - \frac{1}{g_1}} \\ \frac{Q_{pk}}{g_m}, & g_1 > \lambda_f, g_m \geq \frac{Q_{pk}}{\frac{1}{\lambda_f} - \frac{1}{g_1}} \end{cases} \quad (4)$$

where $g_m = \max_i g_{0i}$, $i = 1, \dots, N$, and λ_f is the nonnegative Lagrange multiplier that is associated with the ATP constraint and can be obtained by solving $\lambda_f (E[p(\mathbf{g}_0, g_1)] - P_{av}) = 0$. In fact, (4) can also be written as (below $[x]^+ = \max(x, 0)$)

$$p(\mathbf{g}_0, g_1) = \min \left(\left[\frac{1}{\lambda_f} - \frac{1}{g_1} \right]^+, \frac{Q_{pk}}{g_m} \right). \quad (5)$$

Remark 1: Note that it is standard to consider an ATP constraint for transmission over fading channels, such as in (2). The justification is given by the fact that an ATP constraint allows the system designer to adapt transmission power to the channel conditions while optimizing some performance measure. For example, when maximizing ergodic capacity for a single user under an average power constraint, the optimal power allocation policy is the celebrated "water-filling" policy [28]. However, motivated by the peak power limitations imposed by practical power amplifiers, it becomes necessary to consider an additional peak power constraint, such as in [29]. For low average power values, the peak power constraint becomes inactive and, therefore, can be ignored altogether. In fact, a more recent study [30] showed how the optimal single-user power allocation policy for maximizing the ergodic capacity under both average and peak power constraints can be characterized in terms of the average power and the peak-to-average-power ratio (PAPR). It turns out, however, that, for most practical fading distributions, the maximum FCC-prescribed PAPR value of 13 dB [31] can

still be met for average power greater than -10 dB with the water-filling-based power allocation, i.e., by completely ignoring the peak power constraint. Although these results are derived for standard noncognitive (or primary) radio networks, they provide a justification for us to also ignore peak transmit power constraints at the SU-Tx's for our CR network setting. Note that we can derive similar conditions as in [30] to justify for what range of values of P_{av} and Q_{pk} we can ignore the peak power constraint while still satisfying the aforementioned FCC regulations. However, these conditions involve far more complicated expressions than the expressions in [30] due to the presence of the peak interference constraint in (2), particularly for $N > 1$ PUs. Hence, we do not go into further details about these results to maintain readability.

Special case. When P_{av} is sufficiently large¹ to make the ATP constraint inactive and, thus, only the PIP constraints are active, called the interference-limited regime (ILR), (2) with full CSI reduces to the problem considered in [1] and [7], where the ergodic capacity maximization problem is studied under PIP constraints only and gives the maximum ergodic capacity of SU as

$$C = E \left[\log \left(1 + g_1 \frac{Q_{pk}}{g_m} \right) \right] = E [\log(1 + zQ_{pk})] \quad (6)$$

where $Z = g_1/g_m$, and the probability density function (pdf) of Z is given by [1], [7]

$$f(z) = N \sum_{k=0}^{N-1} (-1)^k \binom{N-1}{k} \frac{1}{(1+k+z)^2}. \quad (7)$$

However, the assumption of full CSI at the SU-Tx (particularly of g_0) is usually unrealistic in practical systems. In the next section, we are therefore interested in designing power allocation schemes of the SU ergodic capacity maximization problem (2) based on the quantized information of g_0 , which is acquired through a no-delay error-free feedback link with a limited rate, where we assume that the SU-Tx has perfect knowledge of g_1 but can only access partial knowledge of g_0 . Later, we will relax the assumption of perfect g_1 at the SU-Tx. Finally, with regard to the usage of the various notions of CSI, we interchangeably use the term “perfect” and “full” CSI to denote the availability of exact instantaneous channel information. Similarly, “imperfect” or “partial” CSI is interchangeably also used to denote some loss of CSI information. In particular, imperfect or partial CSI can be available in the following two forms: 1) as a noisy estimated (noisy instantaneous) CSI or 2) a quantized CSI (through a finite-rate limited-feedback link).

III. QUANTIZED POWER ALLOCATION WITH PERFECT g_1 AND IMPERFECT g_0 AT THE SECONDARY-USER TRANSMITTER

In this section, we design and analyze a QPA strategy when the SU-Tx has exact knowledge of the instantaneous channel

¹For $N = 1$, it can be shown that we need infinite P_{av} for this condition to hold, but for $N > 1$, a sufficiently large but finite P_{av} will suffice.

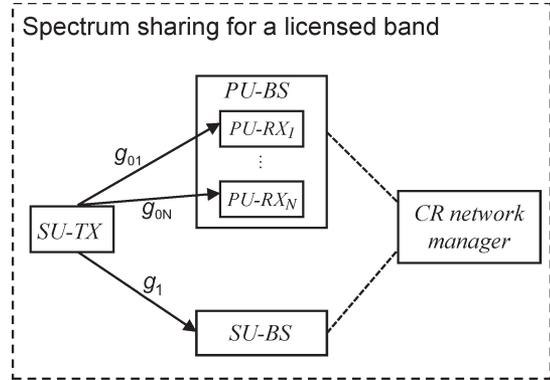


Fig. 1. System model for the QPA strategy.

information of g_1 and quantized information of the channel vector g_0 , which is available through a finite-rate limited-feedback link. In Section III-A, we first present an iterative algorithm to find a locally optimum solution to the associated nonconvex optimization problem, followed by some asymptotic analysis for the case when the number of feedback bits becomes large. We also present a simple suboptimal algorithm for comparison, because our algorithm can attain only a local optimum. In Section III-B, we provide an asymptotic analysis for the case when the number of PUs N becomes large. Finally, in Section III-C, we present an alternative scheme for the ILR, called the QRA strategy, based on the rate-constrained information of the ratio $g_1/\max_i g_{0i}$ and discuss the relative advantages and disadvantages of QPA and QRA.

A. QPA With the Limited-Feedback Strategy

Following [19] and as shown in Fig. 1, here, we assume that there is a central controller, called the CR network manager, who can obtain perfect information on g_1 from the SU-BS and perfect information of g_0 from the PU-BS, possibly over fiber-optic links, and then forward some appropriately quantized $V = Q_{pk}/g_m$ CSI information to the SU-Tx through a finite-rate feedback link. Note that the existence of such central controllers is also assumed quite commonly in the literature on multicell multiple-input–multiple-output or macrodiversity-based systems with cooperative base stations in a primary network, where several base stations are assumed to be connected to a central controller through a backhaul link so that information about out-of-cell interference can be obtained, resulting in higher capacity [17], [18]. Under such a network modeling assumption, given B b of feedback, a power codebook $\mathcal{P} = \{p_1, \dots, p_L\}$ of cardinality, $L = 2^B$, is designed offline based on the statistics of V and g_1 . This codebook is made available *a priori* by the SU-Tx, the SU-Rx, and the CR network manager. Given a channel realization (g_0, g_1) , the CR network manager applies a deterministic mapping $\mathcal{I}(V, g_1)$ from the current instantaneous (V, g_1) information to one of the L integer indices, which partitions the vector space of (V, g_1) into L regions $\mathcal{R}_1, \dots, \mathcal{R}_L$. This mapping is defined as $\mathcal{I}(V, g_1) = j$, if $(V, g_1) \in \mathcal{R}_j$, $j = 1, \dots, L$, and then sends the corresponding index $j = \mathcal{I}(V, g_1)$ to the SU-Tx (and also the SU-Rx) through the feedback link. The SU-Tx then uses

the associated power codebook element (for example, if the feedback signal is j , then p_j will be used as the transmission power) to adapt its transmission strategy.

Let $\Pr(\mathcal{R}_j)$, $E[\bullet|\mathcal{R}_j]$ indicate $\Pr((V, g_1) \in \mathcal{R}_j)$ (the probability that (V, g_1) falls in the region \mathcal{R}_j) and $E[\bullet|(V, g_1) \in \mathcal{R}_j]$, respectively. Then, the SU ergodic capacity maximization problem (2) with limited feedback can be formulated as

$$\begin{aligned} \max_{\{p_1, \dots, p_L\}} C_L(\mathcal{P}) &= \sum_{j=1}^L E[\log(1 + p_j g_1) | \mathcal{R}_j] \Pr(\mathcal{R}_j) \\ \text{s.t.} \quad \sum_{j=1}^L p_j \Pr(\mathcal{R}_j) &\leq P_{av} \\ 0 \leq p_j &\leq \min(V | V \in \mathcal{R}_j) \quad \forall j = 1, \dots, L. \end{aligned} \quad (8)$$

Thus, we need to jointly optimize the channel partition regions and the power codebook such that the ergodic capacity of SU is maximized under the aforementioned constraints. Note that the aforementioned joint optimization problem, in general, is not convex, and hence, any optimum solution to the aforementioned problem that we can find by solving the KKT necessary conditions should be interpreted as a local optimum only. In the following discussion, we have also made it explicitly clear that the solutions that we obtained are locally optimum.

Lemma 1: Let $\mathcal{P} = \{p_1, \dots, p_L\}$ and the corresponding channel partitioning $\mathcal{R}_1, \dots, \mathcal{R}_L$ denote a locally optimal solution to the optimization problem (8). Let $p(V, g_1)$ represent the mapping from the instantaneous (V, g_1) to the allocated power level; then, the following cases hold.

- 1) When $\lambda > 0$, let $\{v_1, \dots, v_{L-1}\}$ denote the corresponding quantization thresholds on the V -axis ($0 = v_1 < \dots < v_{L-1} < (1/\lambda)$) and let $v_L = 1/\lambda$. We have (9), shown at the bottom of the page.
- 2) When $\lambda = 0$, let $\{v_1, \dots, v_L\}$ denote the corresponding quantization thresholds on the V -axis ($0 = v_1 < \dots < v_L < \infty$) and let $v_{L+1} = \infty$. We have

$$p(V, g_1) = p_j = v_j \quad \text{if} \quad v_j \leq V < v_{j+1}, \quad j = 1, \dots, L \quad (10)$$

where λ is the nonnegative Lagrange multiplier associated with the ATP constraint of (8).

Proof: See Appendix A. ■

Based on Lemma 1, when $\lambda = 0$, i.e., the ATP constraint is inactive, the quantization structure is pretty straightforward, because it involves only the quantization of the V -axis. Fig. 2 illustrates the optimum partition region structure for the non-trivial case when $\lambda > 0$. Note that, here, the first region \mathcal{R}_1

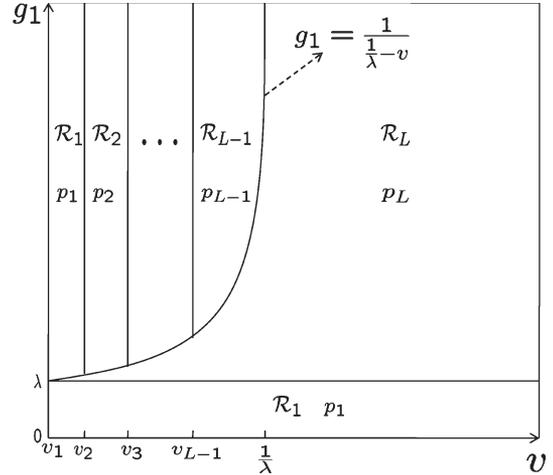


Fig. 2. Structure of optimum quantization regions when $\lambda > 0$.

includes two parts— $\{g_1 \leq \lambda\}$ and $\{V < (1/\lambda - (1/g_1)), v_1 \leq V < v_2\}$ —and $p_1 = 0$ implies that the first region is in outage.

We separately consider the following two cases.

Case I: $\lambda > 0$ (the ATP constraint is active).

Let $F(v)$, $f(v)$ indicate the cumulative distribution function (cdf) and pdf of V , respectively, based on Lemma 1. Then, (8) becomes

$$\begin{aligned} \max_{\{v_2, \dots, v_{L-1}\}} C_L(\mathcal{P}) &= \sum_{j=2}^{L-1} E[\log(1 + v_j g_1) | \mathcal{R}_j] \Pr(\mathcal{R}_j) \\ &\quad + E\left[\log\left(1 + \left(\frac{1}{\lambda} - \frac{1}{g_1}\right) g_1\right) | \mathcal{R}_L\right] \Pr(\mathcal{R}_L) \\ \text{s.t.} \quad \sum_{j=2}^{L-1} v_j \Pr(\mathcal{R}_j) &+ E\left[\left(\frac{1}{\lambda} - \frac{1}{g_1}\right) | \mathcal{R}_L\right] \Pr(\mathcal{R}_L) \leq P_{av}. \end{aligned} \quad (11)$$

Although the aforementioned optimization problem may be verified to be nonconvex, under some regularity conditions, we can employ the KKT necessary conditions to find local maxima for (11). These regularity conditions (see Section III-C1) essentially state that, for a feasible solution to be a local maximum that satisfies the KKT necessary conditions, the gradient vectors of the active inequality constraints with respect to the variables of optimization, evaluated at the solution, must be linearly independent. Note that, although such regularity conditions cannot be verified, in general, before solving the KKT conditions, they can be checked after finding the solutions

$$p(V, g_1) = \begin{cases} p_1 = 0, & \text{if } g_1 \leq \lambda \text{ or } V < \left(\frac{1}{\lambda} - \frac{1}{g_1}\right), v_1 \leq V < v_2 \\ p_j = v_j, & \text{if } V < \left(\frac{1}{\lambda} - \frac{1}{g_1}\right), v_j \leq V < v_{j+1}, j \in [2, L-1] \\ p_L = \left(\frac{1}{\lambda} - \frac{1}{g_1}\right), & \text{if } g_1 > \lambda, V \geq \left(\frac{1}{\lambda} - \frac{1}{g_1}\right) \end{cases} \quad (9)$$

that satisfy the KKT conditions. Taking the first derivative of the Lagrangian function of (11) by using the Leibniz integral rule and setting it to zero, we can obtain

$$\int_{v_j}^{v_{j+1}} \left(\frac{-e^{\frac{1}{v_j}} E_1 \left(\frac{1}{\lambda - v} + \frac{1}{v_j} \right)}{v_j^2} + e^{-\frac{1}{\lambda - v}} \left(\frac{1}{v_j} - \lambda \right) \right) f(v) dv$$

$$= \left(\hat{f}_j(v_j) - \hat{f}_j(v_{j-1}) \right) f(v_j) \quad j = 2, \dots, L-1 \quad (12)$$

where $E_1(x)$ is the exponential integral, defined as $E_1(x) = \int_x^\infty (e^{-z} dz/z)$, and $\hat{f}_2(v_1) = 0$. In addition

$$\hat{f}_j(v) = e^{\frac{1}{v}} E_1 \left(\frac{1}{\lambda - v_j} + \frac{1}{v} \right) + e^{-\frac{1}{\lambda - v_j}} \log \left(1 + v \frac{1}{\lambda - v_j} \right) - \lambda v e^{-\frac{1}{\lambda - v_j}} \quad (13)$$

and the λ can be obtained by solving

$$\lambda \left(\sum_{j=2}^{L-1} v_j \Pr(\mathcal{R}_j) + E \left[\frac{1}{\lambda} - \frac{1}{g_1} |\mathcal{R}_L| \right] \Pr(\mathcal{R}_L) - P_{av} \right) = 0. \quad (14)$$

It is shown in Appendix B that

$$F(v) = 1 - \left(1 - e^{-\frac{Q_{pk}}{v}} \right)^N$$

$$f(v) = \frac{NQ_{pk}}{v^2} e^{-\frac{Q_{pk}}{v}} \left(1 - e^{-\frac{Q_{pk}}{v}} \right)^{N-1}. \quad (15)$$

Thus, for fixed λ , given a v_2 , based on (12), we can successively compute v_3, \dots, v_{L-1} numerically, and then, (12) with $j = L-1$, which has thus only one unknown variable v_2 , can numerically be solved for v_2 . In principle, the optimal value of the Lagrange multiplier λ (note that, although the primal problem is nonconvex for a fixed λ , the dual problem is convex in λ [32]) can be obtained by solving (14) with a subgradient method, i.e., by updating λ until convergence using

$$\lambda^{l+1} = \left[\lambda^l - \alpha^l \left(P_{av} - \sum_{j=2}^{L-1} v_j \Pr(\mathcal{R}_j) - E \left[\left(\frac{1}{\lambda} - \frac{1}{g_1} \right) |\mathcal{R}_L| \right] \Pr(\mathcal{R}_L) \right) \right]^+ \quad (16)$$

where l is the iteration number, and α^l is the positive scalar step size for the l th iteration that satisfies $\sum_{l=1}^\infty \alpha_l = \infty$ and $\sum_{l=1}^\infty (\alpha_l)^2 < \infty$. One alternative method is to use an iterative bisection search method to find the optimal λ , because the average power consumption function in the constraint of (11) can be shown to be a monotonically decreasing and continuous function of λ , following a proof identical to [34]. In our implementation, however, due to the nonconvexity of the primal problem, we cannot guarantee the global optimality of the variables $\{v_j\}$, and hence, neither the subgradient- nor

the bisection-based method can theoretically be strictly guaranteed to converge. It is common practice, however, to use such methods, even in the case of nonconvex primal problems (for example, see [32] for a justification on the subgradient method), and indeed, our simulation studies confirm that both the subgradient- and bisection-based methods for finding λ converge to almost-identical solutions. Note that the subgradient in this case can be considered an approximate subgradient (when only locally optimal solutions to the primal problem can be found), and for a discussion on related convergence issues, see [33, Sec. VI-C2, pp. 614-615].

We can thus iteratively repeat the aforementioned two steps (12) and (16) until a satisfactory convergence criterion is met.

Case 2: $\lambda = 0$ (only the PIP constraints are active, i.e., ILR).

In this case, we can see based on Lemma 1 that none of the power levels depends on the g_1 information; therefore, in this case, the SU-Tx or the CR network manager does not require any knowledge of g_1 , and (8) becomes

$$\max_{\{v_2, \dots, v_L\}} C_L(\mathcal{P}) = \sum_{j=2}^L E [\log(1 + v_j g_1)] (F(v_{j+1}) - F(v_j)). \quad (17)$$

By using the KKT necessary conditions, we can obtain

$$F(v_{j+1}) = F(v_j) + f(v_j) \frac{\tilde{f}_1(v_j) - \tilde{f}_1(v_{j-1})}{\tilde{f}_2(v_j)}, \quad j \in [2, L] \quad (18)$$

where $\tilde{f}_1(v_1) = 0$, and

$$\tilde{f}_1(v) = E [\log(1 + v g_1)] = e^{\frac{1}{v}} E_1 \left(\frac{1}{v} \right) \quad \tilde{f}_2(v) = \frac{\partial \tilde{f}_1(v)}{\partial v}$$

$$= \frac{-e^{\frac{1}{v}} E_1 \left(\frac{1}{v} \right)}{v^2} + \frac{1}{v}. \quad (19)$$

Based on the expression of $F(v)$ in (15), (18) can be rewritten as

$$v_{j+1} = - \frac{Q_{pk}}{\log \left(1 - \left(1 - \left(F(v_j) + f(v_j) \frac{\tilde{f}_1(v_j) - \tilde{f}_1(v_{j-1})}{\tilde{f}_2(v_j)} \right) \right)^{\frac{1}{N}} \right)}$$

$$j = 2, \dots, L-1$$

$$F(v_L) + f(v_L) \frac{\tilde{f}_1(v_L) - \tilde{f}_1(v_{L-1})}{\tilde{f}_2(v_L)} = 1. \quad (20)$$

Thus, given a specific value of v_2 , we can successively compute v_3, \dots, v_L using (20), and then, the last equation in (20), which has thus only one unknown variable v_2 , can numerically be solved for v_2 .

As the number of feedback bits $B = \log_2(L) \rightarrow \infty$, the length of quantization interval $[v_j, v_{j+1})$ approaches zero, and by using the mean value theorem [21], we can get

$$\frac{\tilde{f}_1(v_j) - \tilde{f}_1(v_{j-1})}{\tilde{f}_2(v_j)} \approx v_j - v_{j-1}. \quad (21)$$

Substituting (21) into (18), for $j = 2, \dots, L$, we have

$$F(v_{j+1}) - F(v_j) \approx f(v_j)(v_j - v_{j-1}) \approx F(v_j) - F(v_{j-1}). \quad (22)$$

Based on (22), we have

$$F(v_{L+1}) - F(v_L) \approx \dots \approx F(v_2) - F(v_1) = \frac{1}{L}. \quad (23)$$

This expression implies that we can apply an equal probability per region (EPrPR) approximation that yields

$$F(v_j) \approx \frac{j-1}{L} \quad j = 2, \dots, L. \quad (24)$$

Therefore, based on the expression of $F(v)$, we have

$$v_j = F^{-1}(v) \approx -\frac{Q_{pk}}{\log\left(1 - \left(1 - \frac{j-1}{L}\right)^{\frac{1}{N}}\right)}, \quad j = 2, \dots, L. \quad (25)$$

Then, we can obtain an approximate expression for the maximum ergodic capacity of SU of (17) with large L as

$$C_L \approx \frac{1}{L} \sum_{j=2}^L e^{-\frac{\log\left(1 - \left(1 - \frac{j-1}{L}\right)^{\frac{1}{N}}\right)}{Q_{pk}}} E_1\left(-\frac{\log\left(1 - \left(1 - \frac{j-1}{L}\right)^{\frac{1}{N}}\right)}{Q_{pk}}\right). \quad (26)$$

Remark 2: Note that the aforementioned approximation based on the EPrPR (although mathematically somewhat justified) approach is only a heuristic-based technique, and therefore, the asymptotic SU capacity expression is only approximately valid for large L . The justification of this heuristic is, however, shown through the effectiveness of these asymptotic results, as illustrated in Section VI, *Numerical Results* (see Fig. 6 and the associated explanations in Section VI).

Special case. When, in addition to $L \rightarrow \infty$, $N \rightarrow \infty$ (i.e., in the case of large number of PUs), applying an asymptotic approximation to the cdf of v given by $F(v) \approx 1 - e^{-Ne^{-(Q_{pk}/v)}}$, which is derived as follows based on (27) and (28), shown below, to (24), we have $v_j \approx -(Q_{pk}/\log(-(\log(1 - (j-1)/L))/N))$, $j = 2, \dots, L$. Therefore, we can obtain $C_L \approx (1/L) \sum_{j=2}^L e^{-(\log(-(\log(1 - (j-1)/L))/N))/Q_{pk}} E_1(-(\log(-(\log(1 - (j-1)/L))/N))/Q_{pk})$. For further details and numerical results on the efficacy of this approximation, see [20].

So far, we have discussed how we can solve the quantization problem (8) for the $\lambda > 0$ and $\lambda = 0$ cases, respectively. We can now combine these two procedures to define the following two steps for finding a locally optimal solution for problem (8) (we call this QPA with perfect g_1 and quantized g_0 scheme as QPA- g_0).

1) Let $\lambda = 0$; then, solving (20) gives a power codebook $\{p_1, \dots, p_L\}$. With this codebook, if $\sum_{j=1}^L p_j \Pr(\mathcal{R}_j) \leq P_{av}$, then it is a locally optimal power codebook for problem (8). Stop; otherwise, go to step 2.

2) If step 1 is not satisfied, we must have $\lambda > 0$. Starting with a random initial value for λ , we can solve (12) to obtain the corresponding power codebook $\{p_1, \dots, p_L\}$ and then update λ by (16). Repeat these steps until convergence, and the final codebook will be a locally optimal power codebook for (8).

Suboptimal algorithm. Because we derive results in a locally optimum solution in the aforementioned scheme, for comparison, we also propose a suboptimal QPA algorithm with a simple intuitive channel quantizer described as follows. Using the celebrated Lloyd algorithm (LA), we quantize V by minimizing its distortion $\sum_{n=1}^{L'} E[(v - v_n)^2 | \mathcal{R}_n] \Pr(\mathcal{R}_n)$, where v_n is the reconstruction point for v in region \mathcal{R}_n . When $\lambda = 0$, we partition the V -axis into L regions with the LA. With the resulting channel quantization regions, the corresponding optimal power allocation is given by $p_n^* = \min(v | \mathcal{R}_n)$, $n = 1, \dots, L$. For $\lambda > 0$, based on the optimal quantization structure in Fig. 2, we quantize the finite set $\{V | 0 \leq V \leq 1/\lambda\}$ into $L-1$ regions with the LA and obtain power allocation $p_n^* = \min(v | \mathcal{R}_n)$, $n = 1, \dots, L-1$ for the first $L-1$ regions and then use $p_L^* = ((1/\lambda) - (1/g_1))$ as the optimal power for the last region. We call this suboptimal method suboptimal scalar quantization (SSQ). Numerical results show that our locally optimum QPA- g_0 significantly outperforms this suboptimal method.

B. Asymptotic Analysis With a Large Number of PUs for QPA (ALNPs-QPA)

In the previous analysis, we have considered one SU and N PUs. As shown in Section III-A, a change in the value of N affects only the distribution of V in (15). Instead of using the exact cdf and pdf for a particular value of N , in this section, we are interested in finding the asymptotic distribution of V as $N \rightarrow \infty$ so that we can significantly reduce the computational complexity of solving (8) for a large number of PUs.

As shown, $V = Q_{pk}/g_m$. Then, the cdf of V can be expressed as

$$F(v) = \Pr\left(\frac{Q_{pk}}{g_m} < v\right) = \int_{\frac{Q_{pk}}{v}}^{\infty} f_{g_m}(g_m) dg_m \quad (27)$$

where $f_{g_m}(g_m)$ denotes the pdf of g_m . It is shown in Appendix C that, as $N \rightarrow \infty$, the limiting asymptotic pdf of g_m is given by

$$f_{g_m}(g_m) \sim Ne^{-g_m} e^{-Ne^{-g_m}} \quad (28)$$

where the notation $y(N) \sim x(N)$ implies that $\lim_{N \rightarrow \infty} y(N)/x(N) = 1$, where $y(N)$, $x(N)$ are two functions of N . Substituting (28) into (27), we can use the approximation $1 - e^{-Ne^{-(Q_{pk}/v)}}$ for the cdf $F(v)$ as N becomes large. Then, after differentiation, the limiting asymptotic pdf of V given by $(NQ_{pk}/v^2)e^{-(Q_{pk}/v)}e^{-Ne^{-(Q_{pk}/v)}}$ can be used as an approximation for $f(v)$ when N becomes large, and the same techniques as the previous section can be used to find a locally optimum QPA.

C. QRA for the ILR

For the ILR case, when the ATP constraint is inactive and only the PIP constraints are active, we can use the QPA strategy with $\lambda = 0$, as discussed in Section III-A. In this section, we propose an alternative limited-feedback-based strategy called the QRA scheme with quantized information of the ratio $Z = g_1/g_m$. Note that, unlike QPA with $\lambda = 0$ (here, we call it QPA0), the QRA scheme requires the assumption that the SU-Tx has full knowledge of g_1 .

The limited-feedback strategy for QRA is similar to the QPA case, because given B b of feedback, a locally optimal operating rate codebook $\mathbf{r} = \{r_1, \dots, r_L\}$ is designed purely offline based on the statistics of ratio Z information. Again, the codebook is known *a priori* by both the SU-Tx and the CR network manager. Given a channel realization (\mathbf{g}_0, g_1) , the CR network manager applies a deterministic mapping $\mathcal{I}(Z)$ from the current instantaneous ratio Z information to one of the L integer indices, which partitions the nonnegative ratio information (scalar space) into L regions $\mathcal{R}_1, \dots, \mathcal{R}_L$, i.e., $\mathcal{I}(Z) = j$, if $Z \in [z_j, z_{j+1})$, $j = 1, \dots, L$, where z_j represents the boundary point between \mathcal{R}_{j-1} and \mathcal{R}_j , and $z_1 = 0$, $z_{L+1} = \infty$. It then sends the corresponding index $j = \mathcal{I}(Z)$ to the SU-Tx through the feedback link. The SU-Tx then uses the associated rate codebook element to adapt its transmission strategy. We will show later that $r_j = \log(1 + z_j Q_{pk})$; thus, with perfect knowledge of g_1 , the actual transmission power at the SU is $p_j = z_j Q_{pk}/g_1$, and then, the actual received interference power at PU_{*i*} is

$$p_j g_{0i} = \frac{z_j Q_{pk}}{g_1} \leq \frac{z_j Q_{pk}}{\max_i g_{0i}} \leq Q_{pk} \quad (29)$$

because the current ratio CSI $g_1/\max_i g_{0i}$ falls in \mathcal{R}_j , $g_1/\max_i g_{0i} \geq z_j$. (29) confirms that this limited-feedback strategy can guarantee that all PIP constraints are satisfied at the PU-Rx's.

For any given ratio-state information $Z = z$, the corresponding maximum mutual information of the SU is given by $R(z) = \log(1 + z Q_{pk})$. Thus, for any $z \in \mathcal{R}_j$, with the rate level being r_j , reliable transmission can be guaranteed only if $R(z) \geq r_j$. In other words, when $R(z) < r_j$, outage will occur. Let $\Pr(\mathcal{R}_j)$, $\Pr(\bullet|\mathcal{R}_j)$ denote $\Pr(Z \in \mathcal{R}_j)$ and $\Pr(\bullet|Z \in \mathcal{R}_j)$, respectively. Then, the ergodic capacity of the SU can be expressed as

$$C_L(\mathbf{r}) = \sum_{j=1}^L r_j \Pr(R(z) \geq r_j | \mathcal{R}_j) \Pr(\mathcal{R}_j). \quad (30)$$

Lemma 2: Let z_j^* be the unique solution for $r_j = \log(1 + z_j^* Q_{pk})$. Then, we have $z_j^* \in [z_j, z_{j+1})$.

Proof: See Appendix D. \blacksquare

With a slight abuse of notation, let $F(z)$ indicate the cdf of ratio Z , and using a result in [7], we obtain

$$F(z) = 1 - N \sum_{k=0}^{N-1} (-1)^k \binom{N-1}{k} \frac{1}{1+k+z} \quad (31)$$

which can be rewritten as

$$F(z) = 1 - NB(N, z + 1) \quad (32)$$

where $B(a, b)$ is the beta function, defined by $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$, where $\Gamma(x)$ is defined by $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$. Note that the pdf $f(z)$ (again, with a slight abuse of notation) of ratio Z is given in (7). Then, the secondary ergodic capacity maximization problem (6) with QRA can be formulated as

$$\begin{aligned} & \max_{\{z_2, \dots, z_L, z_1^*, \dots, z_L^*\}} \sum_{j=1}^L \log(1 + z_j^* Q_{pk}) (F(z_{j+1}) - F(z_j^*)) \\ \text{s.t. } & z_j \leq z_j^* \leq z_{j+1} \quad \forall j. \end{aligned} \quad (33)$$

Lemma 3: $z_j = z_j^* \quad \forall j = 2, \dots, L$.

Proof: See Appendix E. \blacksquare

Using this result, (33) becomes

$$\max_{z_j^*} \sum_{j=1}^L \log(1 + z_j^* Q_{pk}) (F(z_{j+1}^*) - F(z_j^*)). \quad (34)$$

Applying the KKT necessary condition to (34), we have

$$\log\left(\frac{1+z_{j-1}^* Q_{pk}}{1+z_j^* Q_{pk}}\right) f(z_j^*) + \frac{Q_{pk}}{1+z_j^* Q_{pk}} (F(z_{j+1}^*) - F(z_j^*)) = 0 \quad (35)$$

where $z_0^* = 0$, and $z_{L+1}^* = \infty$. Based on (35), we have, $j = 2, \dots, L$

$$z_{j-1}^* = \frac{1}{Q_{pk}} \left(e^{\left\{ -\frac{Q_{pk}}{1+z_j^* Q_{pk}} \frac{F(z_{j+1}^*) - F(z_j^*)}{f(z_j^*)} + \log(1+z_j^* Q_{pk}) \right\}} - 1 \right). \quad (36)$$

Thus, given a z_L^* , based on (36), we can successively compute z_{L-1}^*, \dots, z_1^* , and then, (35) with $j = 1$ becomes an equation with only one unknown variable z_L^* , which can numerically be solved.

QRA has a similar asymptotic behavior in the high-resolution quantization regime as QPA0, and we use an EPrPRL-like (EPrPRL) approximation to obtain an asymptotic expression for the SU ergodic capacity. Based on the KKT conditions (35), for $j = 1, \dots, L$, we have

$$\frac{F(z_{j+1}^*) - F(z_j^*)}{f(z_j^*)} = \frac{\log(1 + z_j^* Q_{pk}) - \log(1 + z_{j-1}^* Q_{pk})}{\frac{Q_{pk}}{1+z_j^* Q_{pk}}}. \quad (37)$$

Again, as the number of feedback bits $B = \log_2(L) \rightarrow \infty$, by using the mean-value theorem, we can get

$$\frac{\log(1 + z_j^* Q_{pk}) - \log(1 + z_{j-1}^* Q_{pk})}{\frac{Q_{pk}}{1+z_j^* Q_{pk}}} \approx (z_j^* - z_{j-1}^*). \quad (38)$$

Substituting (38) into (37), we have

$$F(z_{j+1}^*) - F(z_j^*) \approx f(z_j^*) (z_j^* - z_{j-1}^*) \approx F(z_j^*) - F(z_{j-1}^*). \tag{39}$$

Based on (39), we have

$$F(z_{L+1}^*) - F(z_L^*) \approx \dots \approx F(z_1^*) - F(z_0^*) = \frac{1}{L+1} \tag{40}$$

which gives

$$F(z_j^*) \approx \frac{j}{L+1} \quad j = 1, \dots, L. \tag{41}$$

Therefore, we have

$$z_j^* \approx F^{-1}\left(\frac{j}{L+1}\right), \quad j = 1, \dots, L. \tag{42}$$

Finally, based on (32), we have $F(z) = 1 - NB(N, z + 1)$, which implies that $F^{-1}(j/L + 1)$ can be obtained by solving the following equation for z :

$$1 - NB(N, z + 1) = \frac{j}{L+1}. \tag{43}$$

Using the aforementioned expression, we can obtain the asymptotic expressions for the maximum ergodic capacity of SU as

$$C_L \approx \frac{1}{L+1} \sum_{j=1}^L \log\left(1 + F^{-1}\left(\frac{j}{L+1}\right) Q_{pk}\right). \tag{44}$$

Remark 3: Similar to the aforementioned EPrPR approximation, the EPrPRL approximation is only a heuristic-based approach, and the effectiveness of the aforementioned EPrPRL approximation for large L is illustrated through simulation results in Section VI, *Numerical Results* (see Fig. 7 and the associated explanations in Section VI).

Special cases. When $N = 1$ (only one PU is present), (44) becomes $C_L \approx (1/L + 1) \sum_{j=1}^L \log(1 + (j/L + 1 - j)Q_{pk})$. When $N \rightarrow \infty$ (a large number of PUs are present), applying the asymptotic cdf of z [which is derived in (46), shown below] into (42) and applying a further approximation (see [20] for more details), (44) can be given by the following closed-form approximation: $C_L \approx (1/L + 1) \sum_{j=1}^L \log(1 + ((1/\gamma) - (W(1/\gamma)(L + 1 - j/L + 1)N^{1/\gamma} \log N) / \log N)Q_{pk})$. Here, $\gamma = 0.57721566\dots$ is the Euler–Mascheroni constant, and $W(x)$ is the Lambert W function, which gives the principal solution for w in $x = we^w$. It was illustrated in [20] that these approximations for the special cases are highly efficient.

As in the QPA case, we can also find the asymptotic distribution of Z as $N \rightarrow \infty$. We can simplify the computational burden of solving (34) for the large number of PUs case by applying the asymptotic distribution of Z into (36). We call

this QRA method ALNPs-QRA. Because we know that $Z = g_1/g_m$, by letting $X = g_m$, the cdf of Z is given by

$$F(z) = \int_0^\infty P(g_1 < xz|x) f_X(x) dx = \int_0^\infty (1 - e^{-xz}) f_X(x) dx. \tag{45}$$

It is shown in Appendix C that, as $N \rightarrow \infty$, the asymptotic pdf of X is given by $f_X(x) \sim Ne^{-x}e^{-Ne^{-x}}$. Substituting it into (45), we can approximate $F(z)$ for large N as

$$F(z) \approx 1 - e^{-N} + \int_0^\infty e^{-xz} Ne^{-Ne^{-x}} dx = 1 - e^{-N} + N^{-z} (\Gamma(z + 1, N) - \Gamma(z + 1)) \tag{46}$$

where $\Gamma(a, b)$ is the incomplete gamma function given by $\Gamma(a, b) = \int_b^\infty t^{a-1} e^{-t} dt$. Then, the asymptotic pdf of Z is approximated by

$$f(z) \approx \frac{N}{(z+1)^2} {}_2F_2(z+1, z+1; z+2, z+2; -N) \tag{47}$$

where ${}_2F_2(a_1, a_2; b_1, b_2; x)$ is a generalized hypergeometric function, given by

$${}_2F_2(a_1, a_2; b_1, b_2; x) = \sum_{k=0}^\infty \frac{(a_1)_k (a_2)_k x^k}{(b_1)_k (b_2)_k k!} \tag{48}$$

in which $(\alpha)_n = \Gamma(\alpha + n)/\Gamma(\alpha)$ is called the Pochhammer symbol.

Remark 4: Note that the QPA and QRA methods based on the asymptotic distribution of V and the ratio Z , respectively, i.e., ALNPs-QPA and ALNPs-QRA, result in approximate solutions to the QPA and QRA problems for large N and are strictly suboptimal. The efficacy of ALNPs-QRA is illustrated through numerical results only (see Fig. 9). Similar results are observed for ALNPs-QPA but are excluded to avoid repetition.

As we have already shown, QPA and QRA both have some useful properties in the ILR that are similar, and both approaches can very easily be extended to other symmetric- or asymmetric-fading distributions (corresponding to whether the distributions of g_1 and g_{0i} are identical or different). However, there are also a few differences between these approaches.

- 1) For QPA0, we do not need to know the instantaneous information of g_1 at the SU-Tx or the CR network manager. However, for QRA, both the SU-Tx and the CR network manager are required to have full information of g_1 .
- 2) Compared to QRA, QPA0 requires more complex computations, because it needs to compute expectations with respect to g_1 , which may not always have a closed-form solution for arbitrarily general distributions of g_1 .
- 3) With regard to more general distributions for g_0 , e.g., the Rician distribution, for QPA0, we cannot get a valuable closed-form expression for the quantization thresholds (such as in (20) with Rayleigh distribution), and we have to solve (18) for the thresholds. However, for QRA, regardless of the distribution of g_1 and g_0 , we always have

a closed-form expression for the quantization thresholds (36), which can substantially reduce the complexity of solving the optimization problem, particularly for a large number of feedback bits.

As we will show through simulation studies (Section VI), with the same number of feedback bits, QPA0 outperforms QRA, but as the number of feedback bits increases, the capacity gap between these approaches is reduced, which implies that the performances of these two methods are very close for a large number of feedback bits. Therefore, when g_1 or g_0 has a complicated distribution, we can choose to use QRA instead of QPA0 for a large number of feedback bits (e.g., ≥ 6 b).

IV. QUANTIZED POWER ALLOCATION WITH IMPERFECT g_1 AND g_0 AT THE SECONDARY-USER TRANSMITTER

In Section III, we studied a locally optimum QPA scheme with perfect g_1 and quantized g_0 feedback at the SU-Tx. In this section, we consider the more practical scenario of the availability of partial or imperfect g_1 information at the SU-Tx. We are interested in investigating the effect of partial or imperfect g_1 information at the SU-Tx in designing the QPA scheme, where the following two different forms of partial g_1 information at SU-Tx will be examined: 1) noisy estimated g_1 and 2) quantized g_1 . In Section IV-A, we show that, when a noisy estimate of g_1 is available at the SU-Tx, it may not be possible to guarantee the peak interference constraints with probability one, particularly when the channel estimation error is modeled as a random variable with unbounded support. However, by choosing a reduced (stricter) peak interference threshold, we can keep the IVP under control and design a suboptimal QPA scheme (with respect to the original peak interference threshold). In Section IV-B, we consider the case where quantized information about g_1 and g_0 is available at the SU-Tx and design another suboptimal QPA scheme. Two further suboptimal QPA schemes are proposed based on simple intuitive channel quantization schemes. The performances of all of these suboptimal schemes are investigated and compared through simulation studies in Section VI.

A. QPA With Quantized g_0 and Noisy Estimated g_1

In this section, we assume that a noisy estimate of the instantaneous g_1 is available at the SU-Tx, where the noise in estimation or the estimation error can be unbounded. In particular, we exploit the following well-established model [12], [22] for the complex channel estimate of the h_1 at SU-Tx, i.e., \hat{h}_1 :

$$\hat{h}_1 = \rho_0 h_1 + \sqrt{1 - \rho_0^2} \eta \quad (49)$$

where η is the channel estimation error with the standard complex normal distribution (SCND), i.e., η is distributed according to $\mathcal{CN}(0, 1)$ (which implies that $E[\eta] = 0$, $E[|\eta|^2] = 1$), and η is independent of h_1 . $\rho_0 \in [0, 1]$ is the correlation coefficient between the true channel amplitude gain h_1 and its estimate \hat{h}_1 , given as $\rho_0 = E[h_1 \hat{h}_1] - E[h_1]E[\hat{h}_1] / \sqrt{\text{Var}[h_1]\text{Var}[\hat{h}_1]}$.

Thus, the estimated g_1 , i.e., \hat{g}_1 , is obtained as $\hat{g}_1 = |\hat{h}_1|^2$. As aforementioned, $\sqrt{g_1} = |h_1|$ is Rayleigh distributed, and $E[|h_1|^2] = E[g_1] = 1$; thus, h_1 is also SCND. Based on (49), it is easy to verify that the linear transform \hat{h}_1 is also distributed complex normally, because $E[\hat{h}_1] = 0$, and

$$\begin{aligned} E[\hat{g}_1] &= E \left[\left| \rho_0 h_1 + \sqrt{1 - \rho_0^2} \eta \right|^2 \right] \\ &= \rho_0^2 E[g_1] + (1 - \rho_0^2) E[|\eta|^2] = 1. \end{aligned} \quad (50)$$

Thus, the magnitude $|\hat{h}_1|$ will also have a Rayleigh distribution, and the squared magnitude $|\hat{h}_1|^2$, i.e., \hat{g}_1 will have the unit mean exponential distribution.

As stated in Section III-A, for the locally optimum QPA with quantized g_0 only, when $\lambda = 0$ (for a sufficiently high P_{av}), implying that only the PIP constraints are active, the SU-Tx is not required to have any knowledge of g_1 . Thus, with partial g_1 information available at the SU-Tx, the optimum QPA solution for this case is still the same as the aforementioned QPA0. However, when $\lambda > 0$ (the ATP constraint is active), with estimated g_1 at the SU-Tx, the transmit power p_L for the last region \mathcal{R}_L becomes $p_L = (1/\lambda - (1/\hat{g}_1))^+$. The actual PIP becomes $(1/\lambda - (1/\hat{g}_1))^+ g_m$, which may not necessarily be less than or equal to Q_{pk} , although we have $(1/\lambda - (1/g_1))g_m < Q_{pk}$ satisfied. Note that this case is due to the unbounded nature of the channel estimation error (modeled as complex Gaussian). Thus, with estimated g_1 at the SU-Tx, it is not possible to guarantee the actual instantaneous PIP to be $\leq Q_{pk}$ with probability 1. It seems that, to satisfy the PIP constraint with probability one, the SU-Tx has to transmit with zero power in \mathcal{R}_L [9], which renders the whole \mathcal{R}_L in outage. A more appropriate strategy for this case is to allow the actual PIP with estimated g_1 to exceed Q_{pk} with a certain small probability (e.g., $\leq 5\%$) [9], which we call the IVP, given by

$$\text{IVP} = \Pr \left(\left(\frac{1}{\lambda} - \frac{1}{\hat{g}_1} \right)^+ g_m > Q_{pk} | \mathcal{R}_L \right) \Pr(\mathcal{R}_L). \quad (51)$$

To achieve a given percentage IVP, we employ a reduced level of Q_{PK} [12], which is denoted as \bar{Q}_{PK} , to design the locally optimal QPA codebook in Section III-A, with $\lambda > 0$. This case implies that the maximum IVP with a certain nominal Q_{PK} is attained when \bar{Q}_{PK} is chosen to be the maximum allowable peak interference in designing QPA and the proposed QPA becomes strictly suboptimal for the original problem with the peak power constraint Q_{PK} . Next, the optimum QPA (11) with estimated g_1 and \bar{Q}_{PK} becomes

$$\begin{aligned} \max_{\{\bar{v}_2, \dots, \bar{v}_{L-1}\}} \bar{C}_L &= \sum_{j=2}^{L-1} E[\log(1 + \bar{v}_j g_1) | \bar{\mathcal{R}}_j] \Pr(\bar{\mathcal{R}}_j) \\ &\quad + E \left[\log \left(1 + \left(\frac{1}{\lambda} - \frac{1}{\hat{g}_1} \right)^+ g_1 \right) | \bar{\mathcal{R}}_L \right] \Pr(\bar{\mathcal{R}}_L) \\ \text{s.t.} \sum_{j=2}^{L-1} \bar{v}_j \Pr(\bar{\mathcal{R}}_j) &+ E \left[\left(\frac{1}{\lambda} - \frac{1}{\hat{g}_1} \right)^+ | \bar{\mathcal{R}}_L \right] \Pr(\bar{\mathcal{R}}_L) \leq P_{av} \end{aligned} \quad (52)$$

where $\{\bar{v}_2, \dots, \bar{v}_{L-1}\}$, $\{\bar{\mathcal{R}}_1, \dots, \bar{\mathcal{R}}_L\}$ denote the new locally optimum quantization thresholds and regions that are associated with \bar{Q}_{PK} , respectively, and λ' is the nonnegative Lagrange multiplier that is associated with the ATP constraint of (52). Equation (52) can be solved using similar methods as used to solve (11). Note that, in (52), the capacity of last region is given by

$$\begin{aligned} E & \left[\log \left(1 + \left(\frac{1}{\lambda'} - \frac{1}{\hat{g}_1} \right)^+ g_1 \right) \middle| \bar{\mathcal{R}}_L \right] \Pr(\bar{\mathcal{R}}_L) \\ &= \int_{\lambda'}^{\infty} \int_{\lambda'}^{\infty} \log \left(1 + \left(\frac{1}{\lambda'} - \frac{1}{\hat{g}_1} \right) g_1 \right) f(g_1, \hat{g}_1) \\ & \quad \times \left(1 - F \left(\frac{1}{\lambda'} - \frac{1}{g_1} \right) \right) dg_1 d\hat{g}_1 \end{aligned} \quad (53)$$

where $F(x) = 1 - (1 - e^{-(\bar{Q}_{pk}/x)})^N$, and $f(g_1, \hat{g}_1)$ is the joint pdf of g_1 and \hat{g}_1 , which, according to the bivariate exponential distributions in [23, eq. (47.1)], is given as

$$f(g_1, \hat{g}_1) = \frac{1}{1-\rho} I_0 \left(\frac{2\sqrt{\rho g_1 \hat{g}_1}}{1-\rho} \right) e^{-\frac{g_1 + \hat{g}_1}{1-\rho}} \quad (54)$$

where ρ is the correlation coefficient between g_1 and \hat{g}_1 , and $I_0(x) = \sum_{k=0}^{\infty} (x/2k!)^{2k}$ is the well-known modified Bessel function of the first kind with order zero.

With the optimal QPA codebook obtained by \bar{Q}_{PK} , the modified last region $\bar{\mathcal{R}}_L$ becomes $\{g_1 \geq \lambda', (\bar{Q}_{pk}/g_m) \geq (1/\lambda') - (1/g_1)\}$, and thus, IVP can be expressed as

$$\begin{aligned} \text{IVP} &= \Pr \left(\left(\frac{1}{\lambda'} - \frac{1}{\hat{g}_1} \right)^+ > \frac{Q_{pk}}{g_m} \middle| g_1 \geq \lambda' \right. \\ & \quad \left. \frac{\bar{Q}_{pk}}{g_m} \geq \frac{1}{\lambda'} - \frac{1}{g_1} \right) \Pr(\bar{\mathcal{R}}_L) \\ &= \Pr \left(\frac{1}{\lambda'} > \frac{Q_{pk}}{g_m} \hat{g}_1 > \frac{1}{\frac{1}{\lambda'} - \frac{Q_{pk}}{g_m}} \middle| g_1 \geq \lambda' \right. \\ & \quad \left. \frac{\bar{Q}_{pk}}{g_m} \geq \frac{1}{\lambda'} - \frac{1}{g_1} \right) \Pr(\bar{\mathcal{R}}_L). \end{aligned} \quad (55)$$

Because $\bar{Q}_{pk} \leq Q_{pk}$ and $(1/\lambda') > Q_{pk}/g_m$, we have $(1/\lambda') > (\bar{Q}_{pk}/g_m)$. Applying this result to (55), we get

$$\begin{aligned} \text{IVP} &= \Pr \left(\frac{1}{\lambda'} > \frac{Q_{pk}}{g_m}, \hat{g}_1 > \frac{1}{\frac{1}{\lambda'} - \frac{Q_{pk}}{g_m}} \middle| \lambda' \leq g_1 \leq \frac{1}{\frac{1}{\lambda'} - \frac{\bar{Q}_{pk}}{g_m}} \right) \\ & \quad \times \Pr \left(\lambda' \leq g_1 \leq \frac{1}{\frac{1}{\lambda'} - \frac{\bar{Q}_{pk}}{g_m}} \right) \\ &= \int_{\lambda' Q_{pk}}^{\infty} f_{g_m}(g_m) \left(\int_{\lambda'}^{\bar{c}} \int_c^{\infty} f(g_1, \hat{g}_1) d\hat{g}_1 dg_1 \right) dg_m \end{aligned} \quad (56)$$

where $\bar{c} = (1/(1/\lambda') - (\bar{Q}_{pk}/g_m))$, $c = (1/(1/\lambda') - (Q_{pk}/g_m))$, and the pdf of g_m is given by $f_{g_m}(g_m) = N e^{-g_m} (1 - e^{-g_m})^{N-1}$ [1]. Let $\Delta = \int_{\lambda'}^{\bar{c}} \int_c^{\infty} f(g_1, \hat{g}_1) d\hat{g}_1 dg_1$. Applying (54) to Δ , we have

$$\Delta = \int_{\lambda'}^{\bar{c}} \frac{1}{1-\rho} e^{-\frac{g_1}{1-\rho}} \left(\int_c^{\infty} I_0 \left(\frac{2\sqrt{\rho g_1 \hat{g}_1}}{1-\rho} \right) e^{-\frac{\hat{g}_1}{1-\rho}} d\hat{g}_1 \right) dg_1. \quad (57)$$

With a change of variable $x = \sqrt{2\hat{g}_1}/(1-\rho)$, (57) becomes

$$\Delta = \int_{\lambda'}^{\bar{c}} e^{-g_1} Q \left(\sqrt{\frac{2\rho g_1}{1-\rho}}, \sqrt{\frac{2c}{1-\rho}} \right) dg_1 \quad (58)$$

where $Q(a, b) = \int_b^{\infty} x e^{-(x^2+a^2/2)} I_0(ax) dx$ is the first-order Marcum Q-function. With, again, a change of variable, let $y = \sqrt{2g_1}$; then, (58) becomes

$$\begin{aligned} \Delta &= \int_{\sqrt{2\lambda'}}^{\infty} y e^{-\frac{y^2}{2}} Q \left(\sqrt{\frac{\rho}{1-\rho}} y, \sqrt{\frac{2c}{1-\rho}} \right) dy \\ & \quad - \int_{\sqrt{2\bar{c}}}^{\infty} y e^{-\frac{y^2}{2}} Q \left(\sqrt{\frac{\rho}{1-\rho}} y, \sqrt{\frac{2c}{1-\rho}} \right) dy. \end{aligned} \quad (59)$$

Applying (14) of [24] to (59), we obtain

$$\begin{aligned} \Delta &= e^{-\lambda'} Q \left(\sqrt{\frac{2\lambda'\rho}{1-\rho}}, \sqrt{\frac{2c}{1-\rho}} \right) \\ & \quad - e^{-c} Q \left(\sqrt{\frac{2\lambda'\rho}{1-\rho}}, \sqrt{\frac{2c\rho}{1-\rho}} \right) \\ & \quad - e^{-\bar{c}} Q \left(\sqrt{\frac{2\bar{c}\rho}{1-\rho}}, \sqrt{\frac{2c}{1-\rho}} \right) \\ & \quad + e^{-c} Q \left(\sqrt{\frac{2\bar{c}}{1-\rho}}, \sqrt{\frac{2c\rho}{1-\rho}} \right). \end{aligned} \quad (60)$$

Thus, the IVP can be given as

$$\begin{aligned} \text{IVP} &= \int_{\lambda' Q_{pk}}^{\infty} f_{g_m}(g_m) \left\{ e^{-\lambda'} Q \left(\sqrt{\frac{2\lambda'\rho}{1-\rho}}, \sqrt{\frac{2c}{1-\rho}} \right) \right. \\ & \quad - e^{-c} Q \left(\sqrt{\frac{2\lambda'\rho}{1-\rho}}, \sqrt{\frac{2c\rho}{1-\rho}} \right) \\ & \quad - e^{-\bar{c}} Q \left(\sqrt{\frac{2\bar{c}\rho}{1-\rho}}, \sqrt{\frac{2c}{1-\rho}} \right) \\ & \quad \left. + e^{-c} Q \left(\sqrt{\frac{2\bar{c}}{1-\rho}}, \sqrt{\frac{2c\rho}{1-\rho}} \right) \right\} dg_m \end{aligned} \quad (61)$$

which can numerically be calculated.

The capacity loss with estimated g_1 due to using \bar{Q}_{pk} as the PIP to obtain a suboptimal QPA codebook (so that the IVP can be kept below a desired maximum) is calculated as $C_{loss} = C_L - \bar{C}_L$, where C_L is the maximum SU ergodic capacity

obtained from the quantization problem (11) with PIP Q_{pk} and perfect knowledge of g_1 .

B. QPA With Quantized (g_0, g_1) Information

The limited-feedback scheme here is similar to QPA in Section III-A, except that we quantize both V and g_1 , not only V . Let λ_l represent the nonnegative Lagrange multiplier that is associated with the ATP constraint for this case. Again, when P_{av} is large enough to make the ATP constraint inactive, only the PIP constraints are effective ($\lambda_l = 0$), and designing the optimum QPA does not require the SU-Tx to have knowledge of g_1 . Therefore, the optimum QPA for this scenario is the same as QPA with quantized g_0 only, as shown in Section III-A, with $\lambda = 0$.

When $\lambda_l > 0$, due to the difficulty and complexity of the QPA analysis (as shown in Section III-A) for this case, we consider a low-complexity suboptimal QPA, in which, based on the boundary $v = ((1/\lambda_l) - (1/g_1))^+$ (similar to the full-CSI case), two different types of power codebooks are derived. Let $\mathcal{P} = \{p_1, \dots, p_{L_1}, p'_1, \dots, p'_{L_2}\}$ with $L_1 + L_2 = L$ and the corresponding partitioning $\{\mathcal{R}_1, \dots, \mathcal{R}_{L_1}, \mathcal{R}'_1, \dots, \mathcal{R}'_{L_2}\}$ denote an optimal solution for the current suboptimal setting. Let $\mathcal{V} = \{v_1, \dots, v_{L_1}\}$, $\mathbf{q} = \{q_1, \dots, q_{L_2}\}$ denote the quantization thresholds that correspond to this solution on the V -axis (where $0 = v_1 < \dots < v_{L_1} < v_{L_1+1} = 1/\lambda$) and the g_1 -axis (where $\lambda_l < q_1 < \dots < q_{L_2} < q_{L_2+1} = \infty$), respectively, and $p(V, g_1)$ represent the mapping from the instantaneous (V, g_1) to the allocated power level. Then, we have

$$p(V, g_1) = \begin{cases} p_1 = 0, & \text{if } V \geq \left(\frac{1}{\lambda_L} - \frac{1}{g_1}\right), g_1 < q_1, \text{ or} \\ & V < \left(\frac{1}{\lambda_L} - \frac{1}{g_1}\right), v_1 \leq V < v_2 \\ p_j = v_j, & \text{if } V < \left(\frac{1}{\lambda_L} - \frac{1}{g_1}\right), v_j \leq V < v_{j+1} \\ & j \in [2, L_1] \\ p'_k = \left(\frac{1}{\lambda_L} - \frac{1}{q_k}\right), & \text{if } V \geq \left(\frac{1}{\lambda_L} - \frac{1}{g_1}\right), q_k \leq g_1 < q_{k+1} \\ & k \in [1, L_2]. \end{cases} \quad (62)$$

We call this suboptimal QPA scheme with both g_0 and g_1 quantized [when $\lambda_l = 0$, the power allocation is the same as (10), whereas when $\lambda > 0$, the power allocation is given by (62)] vector quantized power allocation (VQPA). Fig. 3 illustrates the structure of the partition regions for VQPA with $\lambda_l > 0$.

When $\lambda_l > 0$, the QPA problem of quantizing both g_0 and g_1 with limited feedback therefore becomes

$$\begin{aligned} & \max_{\{v_1, \dots, v_{L_1}, q_1, \dots, q_{L_2}\}} \sum_{j=1}^{L_1} E \left[\log(1 + v_j g_1) \middle| \mathcal{R}_j \right] \Pr(\mathcal{R}_j) \\ & + \sum_{k=1}^{L_2} E \left[\log \left(1 + \left(\frac{1}{\lambda_L} - \frac{1}{q_k} \right) g_1 \right) \middle| \mathcal{R}'_k \right] \Pr(\mathcal{R}'_k) \\ \text{s.t. } & \sum_{j=1}^{L_1} v_j \Pr(\mathcal{R}_j) + \sum_{k=1}^{L_2} E \left[\frac{1}{\lambda_L} - \frac{1}{q_k} \middle| \mathcal{R}'_k \right] \Pr(\mathcal{R}'_k) \leq P_{av}. \end{aligned} \quad (63)$$

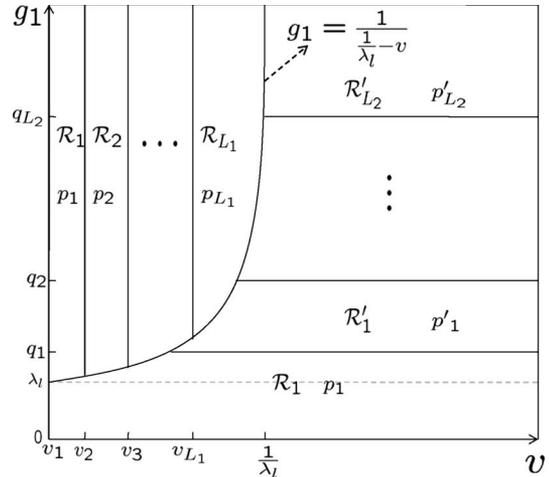


Fig. 3. Quantization region structure of VQPA with $\lambda_l > 0$.

By using the KKT necessary conditions, v_2, \dots, v_{L_1} can be obtained by solving (12) with $\lambda = \lambda_l$ and $j = 2, \dots, L_1$, and q_1, \dots, q_{L_2} can be found by solving the equations in

$$\begin{aligned} & \int_{q_k}^{q_{k+1}} \frac{1}{q_k^2} \left(\frac{g_1}{1 + \left(\frac{1}{\lambda_L} - \frac{1}{q_k} \right) g_1} \lambda \right) f_1(g_1) \left(1 - F \left(\frac{1}{\lambda_L} - \frac{1}{g_1} \right) \right) dg_1 \\ & = (\bar{f}_k(q_k) - \bar{f}_k(q_{k-1})) f_1(q_k) \left(1 - F \left(\frac{1}{\lambda_L} - \frac{1}{q_k} \right) \right) \\ & j = 1, \dots, L_2 \end{aligned} \quad (64)$$

where $\bar{f}_k(q) = \log(1 + (1/\lambda_l - (1/q))q_k) - 1 + (\lambda_l/q)$, $f_1(g_1) = e^{-g_1}$, $F(\cdot)$ is given by (15), and λ_l can be obtained by solving

$$\lambda_L \left(\sum_{j=1}^{L_1} v_j \Pr(\mathcal{R}_j) + \sum_{k=1}^{L_2} E \left[\frac{1}{\lambda_L} - \frac{1}{q_k} \middle| \mathcal{R}'_k \right] \Pr(\mathcal{R}'_k) - P_{av} \right) = 0. \quad (65)$$

Thus, for a fixed λ_l , v_2, \dots, v_{L_1} can be obtained by solving (12), with $\lambda = \lambda_l$ and $j = 2, \dots, L_1$, and given a q_1 , based on (64) we can successively compute q_2, \dots, q_{L_2} numerically. Then, (64) with $k = L_2$, which has only one unknown variable q_1 , can numerically be solved for q_1 . Then, we can update λ_l until convergence using an approximate subgradient-based method (or a bisection-based method, as mentioned in Section III-A), as shown in

$$\begin{aligned} \lambda_L^{n+1} = & \left[\lambda_L^n - \alpha^n \left(P_{av} - \sum_{j=1}^{L_1} v_j \Pr(\mathcal{R}_j) \right. \right. \\ & \left. \left. - \sum_{k=1}^{L_2} E \left[\frac{1}{\lambda_L} - \frac{1}{q_k} \middle| \mathcal{R}'_k \right] \Pr(\mathcal{R}'_k) \right) \right]^+ \end{aligned} \quad (66)$$

where n is the iteration number, and α^n is a positive scalar step size for the n th iteration that satisfies the aforementioned usual conditions. We can thus iteratively repeat the aforementioned two steps until a satisfactory convergence criterion is met.

The aforementioned result is based on a given pair of values for L_1 and L_2 . To find the optimum pair of (L_1, L_2) , we can exhaustively search all possible combinations of L_1 and L_2 so that $L_1 + L_2 = L$ and pick the combination that gives the best SU ergodic capacity.

Alternative suboptimal algorithms: Again, for comparison, we propose the following two alternative intuitive suboptimal algorithms for finding the QPA.

- 1) We separately quantize V and g_1 (i.e., separate scalar quantization is used) by minimizing their corresponding distortion $\sum_{n=1}^{L'_1} E[(v - v_n)^2 | R_n] \Pr(R_n)$ and $\sum_{k=1}^{L'_2} E[(g_1 - g_{1k})^2 | R'_k] \Pr(R'_k)$, respectively, with the LA, where v_n, g_{1k} are the reconstruction points for v and g_1 in R_n and R'_k , respectively, and $L'_1 \times L'_2 = L$. We then use the resulting channel quantization regions to find the corresponding optimal power allocation p_{nk}^* for the region where $v \in R_n, g_1 \in R'_k$, which is given by $p_{nk}^* = \min\{\min(v | R_n, R'_k), [p_{nk}]^+\}$, where p_{nk} is the solution that is found by solving $E[g_1/1 + g_1 p_{nk} - \lambda_l | R_n, R'_k] = 0$. We call this method SCQ.
- 2) We jointly quantize V and g_1 by minimizing the distortion $\sum_{m=1}^L E[(v - v_m)^2 + (g_1 - g_{1m})^2 | R_m] \Pr(R_m)$ with the LA and then use the resulting channel quantization regions to find the associated optimal power allocation $p_m^* = \min\{\min(v | R_m), [p_m]^+\}$, where p_m is the solution of solving $E[(g_1/1 + g_1 p_m) - \lambda_l | R_m] = 0$. We call this approach the joint channel quantization (JCQ) method. Note that, when $\lambda_l = 0$, which implies quantizing only V , it is easy to verify that the results in this case of SCQ or JCQ are same as the SSQ scheme in Section III-A, with $\lambda = 0$.

Numerical results illustrate that the VQPA scheme that we propose significantly outperforms SCQ and JCQ. See Section VI for more details.

V. EXTENSION TO THE MULTIPLE-SECONDARY USER CASE

In our system model, when more than one SU transmits to the SU-BS, it becomes a cognitive multiple-access channel (C-MAC) network as shown in [27]. Given K SUs and N PUs, with the assumption of perfect channel knowledge, the throughput (sum-rate) maximization problem under both K individual ATP constraints at each SU and N individual PIP constraints at each PU-Rx has been considered in [27]. It is shown that this convex optimization problem does not have a closed-form solution but can numerically be solved using the interior point method for convex optimization. For the case when only partial or imperfect CSI is available at the SU-Tx's, this optimization problem is still an open problem, is beyond the scope of this paper, and will be investigated in future work. However, for illustration, we show that, with quantized CSI, by considering the AIP constraints instead of the PIP constraints at each PU-Rx, the SU ergodic capacity maximization problem can be solved by using a modified generalized Lloyds-type algorithm (GLA), similar to our previous work [19], where only a single SU was considered.

Let g_{kn} denote the channel power gain for the link from the k th SU-Tx to the n th PU-BS, $k = 1, \dots, K, n = 1, \dots, N$, whereas the channel power gain for the link from the k th SU-Tx to the SU-BS is indicated by $h_k, k = 1, \dots, K$. Let $\mathcal{P} = \{P_1, \dots, P_L\}$ with $P_j = \{p_{1j}, \dots, p_{Kj}\}'$ denote the power codebook and $\{\mathcal{R}_j, j = 1, \dots, L\}$ represent the quantization regions. With limited feedback, the ergodic sum-rate maximization problem subject to both K individual ATP constraints at each SU and N individual AIP constraints at each PU-Rx can be formulated as

$$\begin{aligned} & \max_{\{p_{kj} \geq 0, \mathcal{R}_j \forall k, j\}} \sum_{j=1}^L E \left[\log \left(1 + \sum_{k=1}^K h_k p_{kj} \right) \middle| \mathcal{R}_j \right] \Pr(\mathcal{R}_j) \\ \text{s.t. } & \sum_{j=1}^L E[p_{kj} | \mathcal{R}_j] \Pr(\mathcal{R}_j) \leq P_{av}^k \quad \forall k = 1, \dots, K \\ & \sum_{j=1}^L E \left[\sum_{k=1}^K g_{kn} p_{kj} \middle| \mathcal{R}_j \right] \Pr(\mathcal{R}_j) \leq Q_{av}^n \quad \forall n = 1, \dots, N \end{aligned} \quad (67)$$

where P_{av}^k and Q_{av}^n are the ATP at the k th SU-Tx and the AIP at the n th PU-Rx, respectively. We employ a modified GLA to jointly optimize the power codebook and partition regions (note that the solution that we obtain here is again only locally optimum), stated as follows. Let $\mathbf{h} = \{h_1, h_2, \dots, h_K\}$ and $G = (g_{kn}), k = 1, 2, \dots, K, n = 1, 2, \dots, N$ denote the SU-Tx to SU-BS channel gain vector and the SU-Tx to PU-Rx channel matrix, respectively. Let λ_k, μ_n be the nonnegative Lagrange multipliers that are associated with the k th ATP and n th AIP constraints, respectively. Beginning with a random initial codebook, we can design the associated optimal partitions using the fact that $\mathcal{R}_j = \{(\mathbf{h}, G) : [\log(1 + \sum_{k=1}^K h_k p_{kj}) - \sum_{k=1}^K \lambda_k p_{kj} - \sum_{n=1}^N \mu_n (\sum_{k=1}^K g_{kn} p_{kj})] \geq [\log(1 + \sum_{k=1}^K h_k p_{ki}) - \sum_{k=1}^K \lambda_k p_{ki} - \sum_{n=1}^N \mu_n (\sum_{k=1}^K g_{kn} p_{ki})] \forall i \neq j\}$, where \mathcal{R}_j is the corresponding partition region for power level P_j in the codebook, and ties are arbitrarily broken. Once the optimal partitions are designed, the new optimal power codebook is found by solving for $\arg \max_{\{p_{kj} \geq 0 \forall k, j\}} E[\log(1 + \sum_{k=1}^K h_k p_{kj}) - \sum_{k=1}^K \lambda_k p_{kj} - \sum_{n=1}^N \mu_n (\sum_{k=1}^K g_{kn} p_{kj}) | \mathcal{R}_j] \Pr(\mathcal{R}_j) \quad \forall j = 1, 2, \dots, L$. Given a partition, this optimization problem is convex, and by using the KKT conditions, we can obtain the optimal power as $\max(p_{kj}^*, 0)$, where p_{kj}^* is the solution to the equation $E[h_k/1 + \sum_{k=1}^K h_k p_{kj} - (\lambda_k + \sum_{n=1}^N \mu_n g_{kn}) | \mathcal{R}_j] = 0$. The optimal Lagrange multipliers can be obtained using similar subgradient based methods, as shown in [19]. These two steps can then be repeated until the resulting ergodic capacity converges within a prespecified accuracy. Note, however, that the problem of finding the jointly optimal partitions and power codebook is nonconvex and that we can only guarantee a local optimum. Nevertheless, numerical results illustrate that with only 4 b of feedback, the ergodic capacity performance of the C-MAC optimization problem with limited feedback (67) using the modified GLA can achieve a performance that is very close to the full-CSI case. See Section VI for more details.

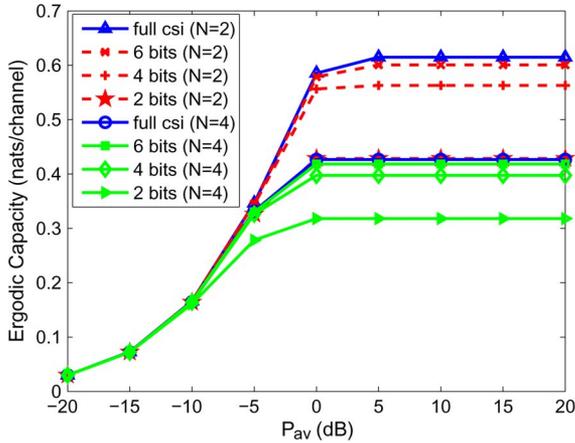


Fig. 4. Ergodic capacity performance of the SU using the QPA-g₀ scheme for different numbers of PUs ($Q_{pk} = 0$ dB).

VI. NUMERICAL RESULTS

In this section, we illustrate the analytical results derived in Sections III and V through numerical simulations. We implement a narrowband spectrum-sharing system with one SU and N PUs, where all the channels involved are assumed to undergo identical Rayleigh fading, i.e., all g_{0i} and g_1 are independent and identically distributed (i.i.d.) and exponentially distributed with unity mean. For the C-MAC case, where K SUs and N PUs share a narrowband, we also assume that all the channels are i.i.d. and $P_{av}^k = P_{av}, \forall k, Q_{av}^n = Q_{av} \forall n$.

Fig. 4 studies the ergodic capacity performance of SUs that share a narrowband spectrum with $N = 2, 4$, respectively, under both the ATP and PIP constraints with the locally optimum QPA - g₀ strategy (i.e., with quantized $V = Q_{pk}/g_m$ and perfect g_1) at $Q_{pk} = 0$ dB and illustrates the effect of increasing the number of feedback bits on the capacity performance. For comparison, we also plot the corresponding capacity performance with full CSI. First, it can easily be observed that the ergodic capacity gradually increases as P_{av} increases until P_{av} reaches a certain threshold, after which the curves become flat (because for a sufficiently large P_{av} , only the PIP constraints are active). The capacity performance also degrades as the number of PUs becomes larger (because the number of PIP constraints increases), as expected.

In Fig. 4, another observation is that, for a fixed value of N , introducing one extra bit of feedback substantially reduces the gap with capacity based on full CSI. To be specific, for $N = 4$, at $P_{av} = 10$ dB, with 2, 4, and 6 b of feedback, the percentage capacity gap between them and full-CSI case are approximately 25.45%, 6.87%, and 1.97%, respectively. In addition, for any N , only 6 b of feedback can result in secondary ergodic capacity very close to the full-CSI case. For example, with $P_{av} = 10$ dB, 6 b of feedback for $N = 2, 4$ only generates around 2.33% and 1.97% percentage capacity losses, respectively, compared with their full-CSI performance, which is clearly an encouraging result.

In addition, we compare the capacity performance of QPA-g₀ with the suboptimal method SSQ at $Q_{pk} = 0$ dB and $N = 4$, as shown in Fig. 5. It is not hard to observe that, with same number of feedback bits, QPA-g₀ can provide dramatic performance

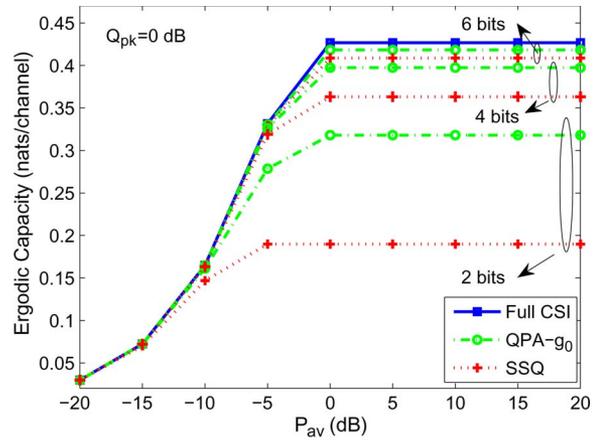


Fig. 5. Capacity performance comparison between the QPA-g₀ scheme and the suboptimal scheme SSQ.

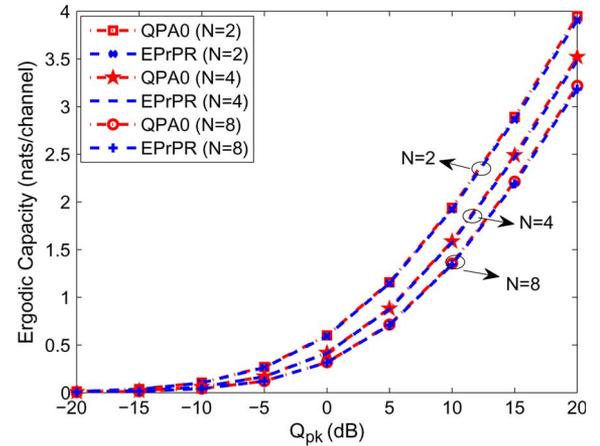


Fig. 6. Comparison of SU ergodic capacity performance using the EPrPR approximation and the corresponding locally optimal QPA0 case for different numbers of PUs with 6 b of feedback.

advance over the SSQ case, particularly at the case of a small number of feedback bits. With a larger number of feedback bits, e.g., 6 b, the gap between these two methods becomes smaller, because when the number of feedback bits is large, the actual structure of the quantization regions becomes less important.

Fig. 6 illustrates the performance of the high-resolution quantization approximation of QPA0 (i.e., EPrPR), which, with 6 b of feedback, compares the SU ergodic capacity performance between the EPrPR approximation and its corresponding locally optimal QPA0 for different numbers of PUs ($N = 2, 4, 8$), respectively. In Fig. 6, one interesting observation is that, for a given N , the capacity performances using the asymptotic EPrPR approximation and the optimal scheme (QPA0) are almost indistinguishable. To be specific, with 6 b of feedback at $Q_{pk} = 10$ dB, for $N = 2, 4, 8$, the percentage capacity loss due to using the EPrPR approximation instead of using the optimal scheme is only around 0.93%, 0.9%, and 0.91%, respectively. This result implies that the EPrPR approximation performs very close to the optimum and confirms that EPrPR is an efficient suboptimal scheme for a large number of quantization levels L . A similar observation can also be made for QRA in

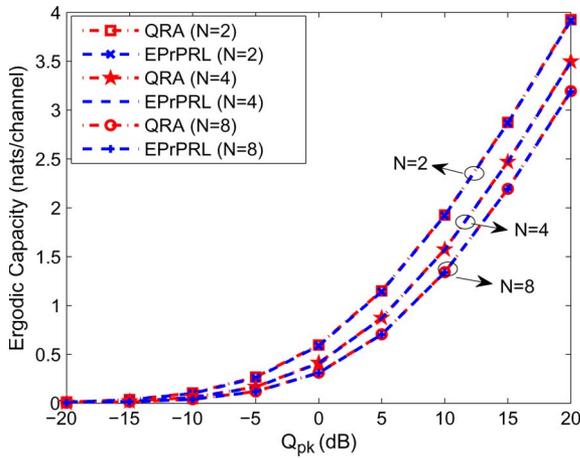


Fig. 7. Comparison of SU ergodic capacity performance using the EPrPRL approximation and the corresponding locally optimal QRA case for different numbers of PUs with 6 b of feedback.

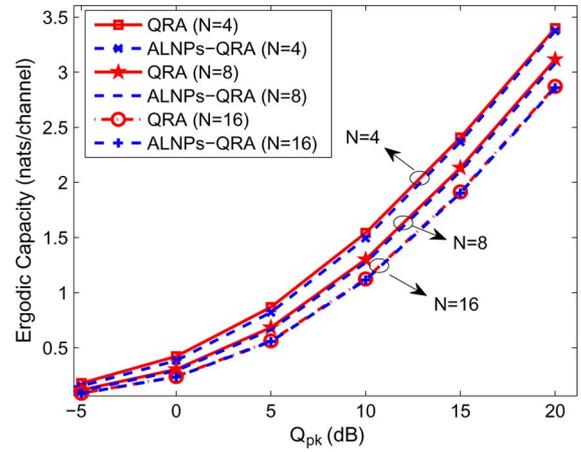


Fig. 9. Comparison of SU ergodic capacity performance between ALNPs-QRA with 4 b of feedback and the corresponding optimal 4-b QRA case for different numbers of PUs.

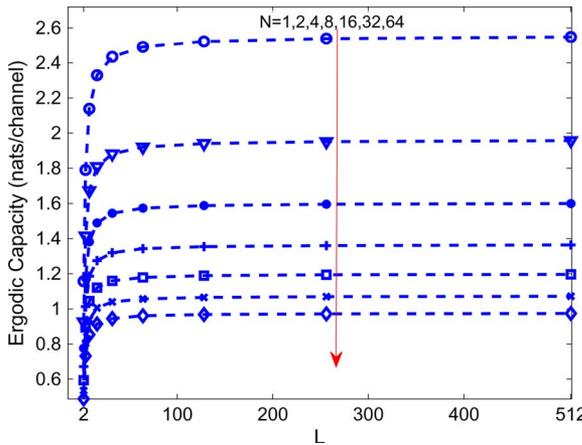


Fig. 8. Asymptotic capacity behavior versus the number of quantization levels L of QPA0 obtained using EPrPR.

Fig. 7. In addition, Fig. 8 depicts the asymptotic SU capacity behavior of QPA0 obtained from (26) versus the number of quantization levels L for different numbers of PUs ($N = 1, 2, 4, 8, 16, 32, 64$) with $Q_{pk} = 10$ dB. Based on Fig. 8, note that, for a given N , the capacity increases as the number of quantization levels L increases; however, as L increases beyond a certain number ($L \geq 2^6$), the capacity curves start to saturate, which further confirms that only a small number of feedback bits (6 b) is required to approach the perfect CSI performance. A similar behavior is also shown for the QRA scheme. We do not include a figure to avoid repetition.

Next, we illustrate the performance of the ALNPs approximation method for QPA and QRA. Here, we only plot the results for QRA. A similar result is also shown for QPA. Fig. 9 investigates the ergodic capacity performance of the SU with quantized feedback (4 b) using the asymptotic analysis method (ALNPs-QRA) for different numbers of PUs ($N = 4, 8, 16$) and compares the results with the corresponding optimal QRA (4-b) case. Interestingly, it can be observed that, with the same number of feedback bits, increasing the number of PUs substantially shrinks the capacity performance gap between

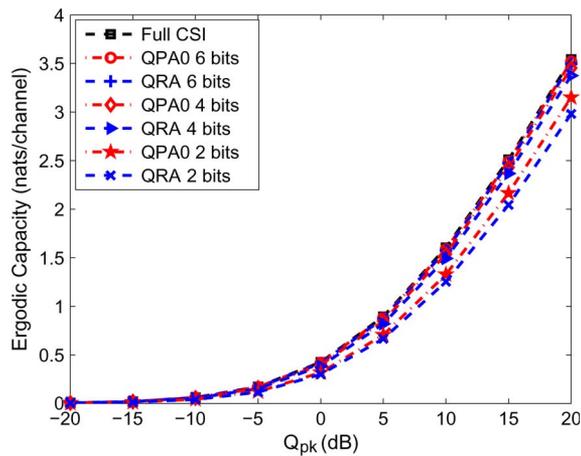


Fig. 10. Comparison of SU ergodic capacity performance using the QPA0 and QRA schemes with $N = 4$.

ALNPs-QRA and the optimal scheme. When $N = 16$, the capacity performance of ALNPs-QRA and the optimal scheme are almost the same. For example, with 4 b of feedback at $Q_{pk} = 10$ dB, for $N = 4, 8, 16$, the percentage capacity gap between ALNPs and the optimal scheme is around 3.21%, 2.21%, and 0.81%, respectively. These results confirm that the ALNPs approximation is an efficient alternative for a large number of PUs.

Fig. 10 compares the ergodic capacity performance of two alternative quantization methods (QPA0 and QRA) for the high- P_{av} case with $N = 4$ for different numbers of feedback bits ($B = 2, 4, 6$). Based on Fig. 10, it can be observed that, with same number of feedback bits, QPA outperforms QRA. However, as the number of feedback bits increases, the capacity gap between the two methods decreases, and as we can see, with 6 b of feedback, the performance of the QRA scheme is very close to the QPA case. For example, at $Q_{pk} = 10$ dB, with 2, 4, and 6 b, the percentage capacity gap between the two quantization schemes is around 5.56%, 2.86%, and 0.84%, respectively.

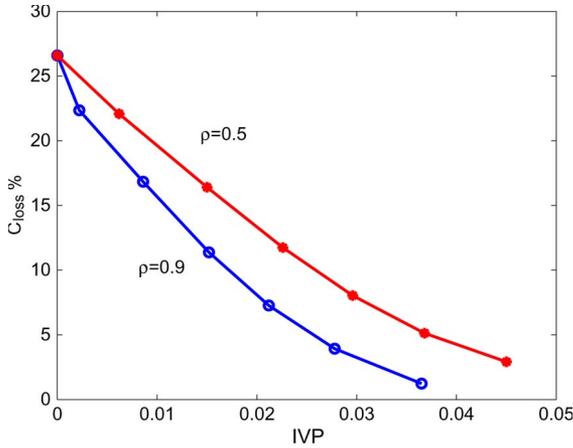


Fig. 11. SU ergodic capacity loss of QPA schemes with quantized g_0 and noisy estimated g_1 versus the IVP for various values of the correlation coefficient ρ ($N = 4, Q_{pk} = 0\text{ dB}, P_{av} = -5\text{ dB}$).

Next, we illustrate the SU ergodic capacity performance with the additional effect of imperfect g_1 in designing QPA schemes. Fig. 11 shows (with $Q_{pk} = 0\text{ dB}$, $P_{av} = -5\text{ dB}$, and $N = 4$), the resulting percentage SU ergodic capacity loss of using QPA schemes with estimated g_1 and \bar{Q}_{pk} against the IVP for different values of the correlation coefficient ρ . Here, the range of \bar{Q}_{pk} is from -2.871 dB to 0 dB , corresponding to $IVP = 0$ to the maximum value of the IVP. As illustrated in Fig. 11, for any ρ , the capacity loss dramatically rises as the IVP decreases, because to obtain a lower IVP, we need to further decrease \bar{Q}_{pk} , which leads to further capacity loss. Increasing ρ lowers the capacity loss due to having better estimates of g_1 . When $\bar{Q}_{pk} = Q_{pk} = 0\text{ dB}$, we obtain the maximum value of the IVP, which is 0.0450 and 0.0365 for $\rho = 0.5$ and $\rho = 0.9$, respectively, and the least value of the capacity loss is 2.91% and 1.23% for $\rho = 0.5$ and $\rho = 0.9$, respectively. Interestingly, regardless of ρ , zero IVP is observed when \bar{Q}_{pk} decreases up to -2.871 dB , which achieves the maximum capacity loss, roughly 26.59% for all ρ . This case is because, when \bar{Q}_{pk} is sufficiently small to make the ATP constraint inactive and only the PIP constraints are active, the optimum QPA does not depend on g_1 , and hence, even with estimated g_1 , $IVP = 0$.

Fig. 12 depicts the SU ergodic capacity performance of VQPA (i.e., with quantized V and g_1) with 2, 4, and 6 b of feedback, respectively, and compares the results with the corresponding performance of the QPA- g_0 scheme (quantized V and perfect g_1). Based on Fig. 12(a), we can easily observe that, with the same number of feedback bits, these two performances almost overlap with each other (recall that, when $\lambda_l = 0$, they are identical, and a difference only exists when $\lambda_l > 0$), and with 6 b of feedback, VQPA is also very close to the full-CSI case. For clearer visualization, in Fig. 12(b), we zoom into the detail of the area of A in Fig. 12(a), which shows that, with the same number of feedback bits, the performance of QPA- g_0 is only slightly better than VQPA, and with an increasing number of feedback bits, the capacity loss due to imperfect g_1 information is reduced. These results confirm that VQPA is a very efficient scheme. Furthermore, as shown in Fig. 13, we also compare the ergodic capacity performance

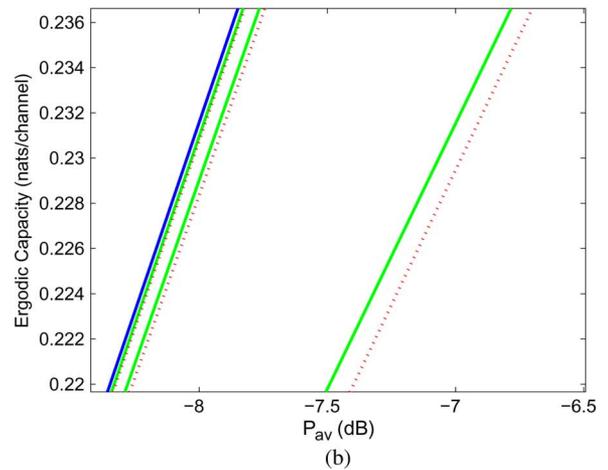
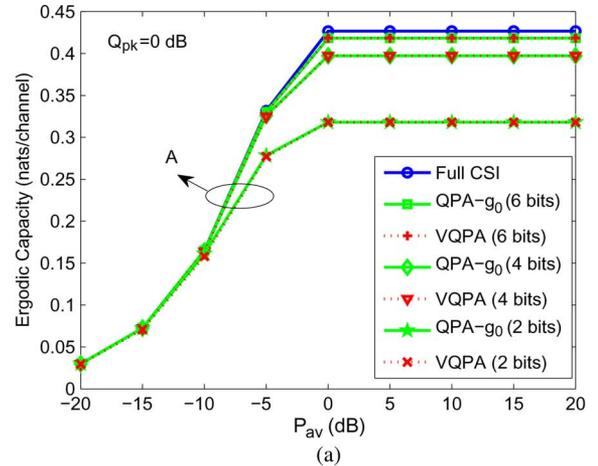


Fig. 12. (a) SU ergodic capacity performance comparison between QPA- g_0 and VQPA ($N = 4$). (b) Zooming in the area of A in (a).

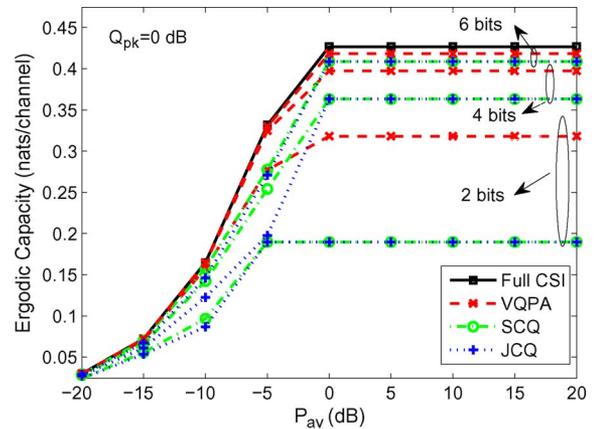


Fig. 13. Capacity performance comparison between the VQPA scheme and two other suboptimal schemes ($N = 4$).

of VQPA with the two other proposed possible suboptimal methods (SCQ and JCQ), with $Q_{pk} = 0\text{ dB}$ and $N = 4$. For the SCQ case, various combinations of L'_1, L'_2 such that $L'_1 \times L'_2 = L$ are investigated, and the combination with the best performance is reported for every value of P_{av} . In Fig. 13, we can easily observe that, with the same number of feedback bits, the performance of both JCQ and SCQ are much worse than

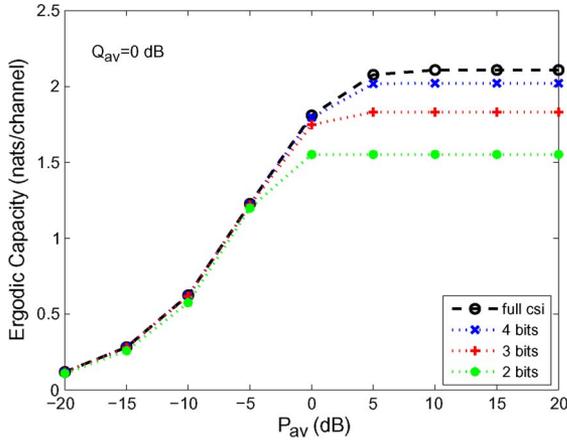


Fig. 14. Ergodic capacity performance for the fading C-MAC under limited feedback with $K = 4$, $N = 1$ ($Q_{av} = 0$ dB).

VQPA, particularly at the case of a small number of feedback bits, which further confirms the efficiency of VQPA.

Fig. 14 shows the ergodic capacity performance of the C-MAC optimization problem (67) with limited feedback solved by the modified GLA, with $K = 4$, $N = 1$, and $Q_{av} = 0$ dB. The important observation in this figure is that, again, only a few bits of feedback can eliminate most of the gap with the full-CSI performance. For example, at $P_{av} = 10$ dB, with 4 b of feedback, there is only around 4.13% capacity loss compared with the full-CSI case.

VII. CONCLUSION

In this paper, we have investigated the problem of ergodic capacity maximization of a SU that shares the same frequency band with a number of PUs in a narrowband spectrum-sharing CR framework under an ATP constraint at the SU-Tx and individual peak interference constraints at each primary receiver. The following three different quantized power codebook schemes are designed for the throughput maximization problem, which correspond to three different forms of channel information of g_1 and g_0 at the SU-Tx: 1) perfect g_1 and quantized information on g_0 ; 2) noisy estimated g_1 and quantized information on g_0 ; and 3) quantized information on both g_1 and g_0 . Note that these quantized power codebook schemes are either locally optimal due to nonconvexity or strictly suboptimal when only partial information on g_1 is available. A discussion on the extension to multiple SUs that share a C-MAC is provided. Numerical results present the efficiency of our quantized feedback schemes. A general observation for these schemes is that, with only 4–6 b of feedback, the SU ergodic capacity with quantized channel information closely approximates the SU ergodic capacity with full CSI at the SU-Tx.

APPENDIX A PROOF OF LEMMA 1

Let λ denote the nonnegative Lagrange multiplier that is associated with the ATP constraint of (8). When $\lambda = 0$

(the ATP is inactive), (8) becomes

$$\begin{aligned} & \max_{\{p_1, \dots, p_L\}} \sum_{j=1}^L E \left[\log(1 + p_j g_1) \middle| \mathcal{R}_j \right] \Pr(\mathcal{R}_j) \\ & \text{s.t. } 0 \leq p_j \leq \min\{V | \mathcal{R}_j\} \quad j = 1, \dots, L. \end{aligned} \quad (68)$$

In this case, it is easy to verify that all the constraints in (68) are satisfied with equality. Let $\{v_1, \dots, v_L\}$ denote the optimum quantization thresholds on the V -axis and $v_{L+1} = \infty$ ($0 = v_1 < \dots < v_L < \infty$). We can obtain

$$p(V, g_1) = p_j = v_j \quad \text{if } v_j \leq v < v_{j+1} \quad j = 1, \dots, L \quad (69)$$

which is independent of the g_1 information.

When $\lambda > 0$, the dual problem of (8) is given as

$$\min_{\lambda \geq 0} g(\lambda) + \lambda P_{av} \quad (70)$$

where the Lagrange dual function $g(\lambda)$ is defined as

$$\begin{aligned} & \max_{\{p_1, \dots, p_L\}} \sum_{j=1}^L E \left[\log(1 + p_j g_1) - \lambda p_j \middle| \mathcal{R}_j \right] \Pr(\mathcal{R}_j) \\ & \text{s.t. } 0 \leq p_j \leq \min\{V | \mathcal{R}_j\} \quad j = 1, \dots, L \end{aligned} \quad (71)$$

which can be decomposed into L parallel subproblems, i.e., for each region \mathcal{R}_j , $j = 1, \dots, L$, solve

$$\begin{aligned} & \max_{p_j} E \left[\log(1 + p_j g_1) - \lambda p_j \middle| \mathcal{R}_j \right] \Pr(\mathcal{R}_j) \\ & \text{s.t. } 0 \leq p_j \leq \min\{V | \mathcal{R}_j\}. \end{aligned} \quad (72)$$

First, based on (72), we can obtain the following results.

Case 1. If $\min\{V | \mathcal{R}_j\} < ((1/\lambda) - (1/g_1))$, based on the power constraint in (72), we must have $p_j = \min\{V | \mathcal{R}_j\}$; otherwise, we can always increase p_j up to $\min\{V | \mathcal{R}_j\}$ to achieve better capacity performance for (72);

Case 2. If $\min\{V | \mathcal{R}_j\} \geq ((1/\lambda) - (1/g_1))^+$, we must have $p_j = ((1/\lambda) - (1/g_1))^+$, because the capacity of (72) in this case is maximized at $p_j = ((1/\lambda) - (1/g_1))^+$.

Now, let $\{v_1, \dots, v_{L-1}\}$ denote the optimum quantization thresholds on the V -axis, where $0 = v_1 < \dots < v_{L-1} < 1/\lambda$. Let $v_L = 1/\lambda$. Based on cases 1 and 2, we can easily get the following results.

- 1) If $\mathcal{R}_1 = \{(V, g_1) | g_1 \leq \lambda\} \cup \{(V, g_1) | V < ((1/\lambda) - (1/g_1)), v_1 \leq V < v_2\}$, then we must have $p_1 = 0$.
- 2) If $\mathcal{R}_j = \{(V, g_1) | V < ((1/\lambda) - (1/g_1)), v_j \leq V < v_{j+1}\}$, then we must have $p_j = v_j$, $j = 2, \dots, L-1$.
- 3) If $\mathcal{R}_L = \{(V, g_1) | V \geq ((1/\lambda) - (1/g_1)) > 0\}$, we must have $p_L = ((1/\lambda) - (1/g_1))$.

Next, we will show that, with the power levels in 1), 2), and 3), the aforementioned partition regions are optimal. Let $\mathcal{P} = \{p_1, \dots, p_L\}$ and the corresponding channel partitioning $\mathcal{R}_1, \dots, \mathcal{R}_L$ denote the optimal solution to the optimization problem (8) such that $p(V, g_1) = p_j$ if $(V, g_1) \in \mathcal{R}_j$.

Let $\mathcal{R}_1^* = \{(V, g_1) | g_1 \leq \lambda\} \cup \{(V, g_1) | V < ((1/\lambda) - (1/g_1))\}$, $v_1 \leq V < v_2\}$ and assume that the set $\mathcal{R}_1^* \setminus \mathcal{R}_1$ is a nonempty set, where \setminus is the set subtraction operation (i.e., if $(V, g_1) \in \mathcal{R}_1^* \setminus \mathcal{R}_1$, then $(V, g_1) \in \mathcal{R}_1^*$, but $(V, g_1) \notin \mathcal{R}_1$). Then, we have $\mathcal{R}_1^* \setminus \mathcal{R}_1 \subseteq (\cup_{k=2}^L \mathcal{R}_k)$, which gives $\forall (v, g_1) \in (\mathcal{R}_1^* \setminus \mathcal{R}_1)$, $p(V, g_1) > 0$. However, this condition violates the power constraint in $\mathcal{R}_1^* \setminus \mathcal{R}_1$, i.e., $0 \leq p(V, g_1) \leq \min\{V | \mathcal{R}_1^* \setminus \mathcal{R}_1\}$, implying $p(V, g_1) = 0$, which is in contradiction with the optimality of the solution $\mathcal{P}, \mathcal{R}_j \forall j$. Therefore, $\mathcal{R}_1^* \setminus \mathcal{R}_1 = \emptyset$, i.e., $\mathcal{R}_1^* \subseteq \mathcal{R}_1$.

$\forall j = 2, \dots, L-1$, let $\mathcal{R}_j^* = \{(V, g_1) | V < ((1/\lambda) - (1/g_1))\}$, $v_j \leq V < v_{j+1}\}$ and assume that the set $\mathcal{R}_j^* \setminus \mathcal{R}_j$ has nonzero probability. Then, the set $\mathcal{R}_j^* \setminus \mathcal{R}_j$ can be partitioned into the following two subsets: 1) $S_j^- = (\mathcal{R}_j^* \setminus \mathcal{R}_j) \cap (\cup_{k=1}^{j-1} \mathcal{R}_k)$ and 2) $S_j^+ = (\mathcal{R}_j^* \setminus \mathcal{R}_j) \cap (\cup_{k=j+1}^L \mathcal{R}_k)$. The set $S_j^- = \emptyset$; otherwise, we can reassign the set S_j^- into region \mathcal{R}_j without violating the power constraints in (71), whereas the total capacity of (71) is increased, because $\forall (V, g_1) \in (\cup_{k=1}^{j-1} \mathcal{R}_k)$, $p(V, g_1) < p_j < ((1/\lambda) - (1/g_1))$, and with the reassignment the capacity of set S_j^- , it achieves better performance, which contradicts the optimality of the solution $\mathcal{P}, \mathcal{R}_j \forall j$. We must also have $S_j^+ = \emptyset$; otherwise, the power constraints in S_j^+ , i.e., $0 \leq p(v, g_1) \leq v_j$, will be violated, because $\forall (v, g_1) \in S_j^+$, $p(v, g_1) > v_j$, which is a contradiction to optimality. Therefore, we must have $\mathcal{R}_j^* \setminus \mathcal{R}_j = \emptyset$, which implies $\mathcal{R}_j^* \subseteq \mathcal{R}_j$.

Let $\mathcal{R}_L^* = \{(V, g_1) | V \geq ((1/\lambda) - (1/g_1)) > 0\}$ and assume that the set $(\mathcal{R}_L^* \setminus \mathcal{R}_L) \neq \emptyset$. Then, we have $\mathcal{R}_L^* \setminus \mathcal{R}_L \subseteq (\cup_{k=1}^{L-1} \mathcal{R}_k)$. Again, we can repartition the set $\mathcal{R}_L^* \setminus \mathcal{R}_L$ into region \mathcal{R}_L , which still satisfies the power constraints in (71). However, this new partition increases the total capacity of (71), because $\forall (V, g_1) \in (\cup_{k=1}^{L-1} \mathcal{R}_k)$, $p(V, g_1) < ((1/\lambda) - (1/g_1))$, and after the repartitioning, the capacity of set $\mathcal{R}_L^* \setminus \mathcal{R}_L$ achieves its maximum value, which contradicts optimality. Therefore, we must have $\mathcal{R}_L^* \setminus \mathcal{R}_L = \emptyset$, i.e., $\mathcal{R}_L^* \subseteq \mathcal{R}_L$.

In summary, we have shown that $\forall j = 1, \dots, L$, $\mathcal{R}_j^* \subseteq \mathcal{R}_j$. Because $\cup_{j=1}^L \mathcal{R}_j^*$ equals the whole space covered by (V, g_1) or $\cup_{j=1}^L \mathcal{R}_j$ and $\mathcal{R}_j^* \subseteq \mathcal{R}_j$, $\forall j$, we can obtain $\mathcal{R}_j^* = \mathcal{R}_j$, $\forall j = 1, \dots, L$.

APPENDIX B DERIVATION OF THE CDF AND THE PDF OF $V = Q_{pk} / \max_i g_{0i}$

Let $g_0 = \max_i g_{0i}$, $i = 1, \dots, N$. Then, the pdf of g_0 is given by [1]

$$f(g_0) = N e^{-g_0} (1 - e^{-g_0})^{N-1}. \quad (73)$$

The cdf of $V = Q_{pk} / \max_i g_{0i}$ can be obtained as

$$\begin{aligned} F(v) &= \Pr \left(\frac{Q_{pk}}{\max_i g_{0i}} < v \right) = \Pr \left(g_0 > \frac{Q_{pk}}{v} \right) \\ &= \int_{\frac{Q_{pk}}{v}}^{\infty} f(g_0) dg_0 = 1 - \left(1 - e^{-\frac{Q_{pk}}{v}} \right)^N. \end{aligned} \quad (74)$$

After differentiation, the pdf of V is given as

$$f(v) = \frac{N Q_{pk}}{v^2} e^{-\frac{Q_{pk}}{v}} \left(1 - e^{-\frac{Q_{pk}}{v}} \right)^{N-1}. \quad (75)$$

Derivation of the Asymptotic pdf of $\max_i g_{0i}$, $i = 1, \dots, N$, as $N \rightarrow \infty$: Given that $g_{01}, g_{02}, \dots, g_{0N}$ are i.i.d. random variables and are exponentially distributed with unity mean, let the cdf $F(x) = \Pr(g_{0i} < x) = 1 - e^{-x}$ and pdf $f(x) = e^{-x}$. Let $X = \max(g_{01}, g_{02}, \dots, g_{0N})$. We want to derive the asymptotic pdf of X as $N \rightarrow \infty$. First, note that

$$\Pr(X < x) = F^N(x). \quad (76)$$

Because $f(x) > 0$ and is differentiable for all x in $(x_1, F^{-1}(1))$ for some x_1 , and

$$\lim_{x \rightarrow F^{-1}(1)} \frac{d}{dx} \left[\frac{1 - F(x)}{f(x)} \right] = \lim_{x \rightarrow \infty} \frac{d}{dx} [1] = 0 \quad (77)$$

according to [25], there exist constants $a_N > 0$ and b_N such that

$$F^N(a_N x + b_N) \rightarrow e^{-e^{-x}}, \text{ as } N \rightarrow \infty \quad (78)$$

where \rightarrow denotes the limit as $N \rightarrow \infty$. We can choose

$$b_N = F^{-1} \left(1 - \frac{1}{N} \right) = \log N \quad a_N = [N f(b_N)]^{-1} = 1. \quad (79)$$

Therefore, as $N \rightarrow \infty$, we have

$$F^N(x + \log N) \rightarrow e^{-e^{-x}} \quad F^N(x) \sim e^{-e^{-(x - \log N)}} \quad (80)$$

$$f_X(x) = \frac{\partial F^N(x)}{\partial x} \sim N e^{-x} e^{-N e^{-x}}. \quad (81)$$

Proof of Lemma 2: The proof is similar to [26], given a set of optimum quantization thresholds $\mathbf{z} = \{z_2, \dots, z_L\}$, we assume that the corresponding optimum rate codebook $\mathbf{r} = \{r_1, \dots, r_L\}$ satisfies $z_j^* \notin [z_j, z_{j+1})$ (i.e., $z_j^* \geq z_{j+1}$ or $z_j^* < z_j$). We construct a new codebook $\mathbf{r}' = \{r_1, \dots, r_{j-1}, r'_j, r_{j+1}, \dots, r_L\}$, where $r'_j = R(z_j)$ with corresponding $z_j'^* = z_j$. If $z_j^* \geq z_{j+1}$, we have $C_L(\mathbf{r}') - C_L(\mathbf{r}) = r'_j \Pr(\mathcal{R}_j) > 0$, which contradicts the optimality of the rate codebook \mathbf{r} . If $z_j^* < z_j$, we have $C_L(\mathbf{r}') - C_L(\mathbf{r}) = (r'_j - r_j) \Pr(\mathcal{R}_j) > 0$, which is also a contradiction with the assumed optimality.

Proof of Lemma 3: Given an optimum rate codebook $\mathbf{r} = \{r_1, \dots, r_L\}$, we assume that the optimum quantization thresholds $\mathbf{z} = \{z_2, \dots, z_L\}$ satisfies $z_j \neq z_j^*$. Then, based on Lemma 2, we have $z_j < z_j^* < z_{j+1}$. Now, we construct a set of new quantization thresholds $\mathbf{z}' = \{z_2, \dots, z_{j-1}, z_j^*, z_{j+1}, \dots, z_L\}$, and we can show that $C_L(\mathbf{z}') - C_L(\mathbf{z}) = [r_{j-1}(F(z_j^*) - F(z_{j-1}^*)) + r_j(F(z_{j+1}) - F(z_j^*))] - [r_{j-1}(F(z_j) - F(z_{j-1}^*)) + r_j(F(z_{j+1}) - F(z_j^*))] = r_{j-1}(F(z_j^*) - F(z_j)) > 0$, which contradicts the optimality of the quantization thresholds \mathbf{z} .

REFERENCES

- [1] A. Ghasemi and E. S. Sousa, "Fundamental limits of spectrum-sharing in fading environments," *IEEE Trans. Wireless Commun.*, vol. 6, no. 2, pp. 649–658, Feb. 2007.
- [2] J. Mitola, III, "Cognitive radio for flexible mobile multimedia communications," in *Proc. IEEE Int. Workshop MoMuC*, San Diego, CA, Nov. 1999, pp. 3–10.
- [3] A. Goldsmith, S. A. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information-theoretic perspective," *Proc. IEEE*, vol. 97, no. 5, pp. 894–914, May 2009.
- [4] X. Kang, Y. Liang, A. Nallanathan, H. K. Garg, and R. Zhang, "Optimal power allocation for fading channels in cognitive radio networks: Ergodic capacity and outage capacity," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 940–950, Feb. 2009.
- [5] R. Zhang, Y.-C. Liang, and S. Cui, "Dynamic resource allocation in cognitive radio networks," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 102–114, Mar. 2010.
- [6] M. Gastpar, "On capacity under received-signal constraints," in *Proc. 42nd Annual Allerton Conf. Commun., Control Comput.*, Monticello, IL, Sep. 29–Oct. 1, 2004, pp. 1322–1331.
- [7] H. A. Suraweera, J. Gao, P. J. Smith, M. Shafi, and M. Faulkner, "Channel capacity limits of cognitive radio in asymmetric-fading environments," in *Proc. IEEE ICC*, Beijing, China, May 2008, pp. 4048–4053.
- [8] L. Musavian and S. Aissa, "Capacity and power allocation for spectrum-sharing communications in fading channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 148–156, Jan. 2009.
- [9] L. Musavian and S. Aissa, "Fundamental capacity limits of cognitive radio in fading environments with imperfect channel information," *IEEE Trans. Commun.*, vol. 57, no. 11, pp. 3472–3480, Nov. 2009.
- [10] K. Huang and R. Zhang, "Cooperative feedback for multiantenna cognitive radio networks," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 747–758, Feb. 2011.
- [11] L. Zhang, Y.-C. Liang, Y. Xin, and H. V. Poor, "Robust cognitive beamforming with partial channel-state information," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4143–4153, Aug. 2009.
- [12] H. A. Suraweera, P. J. Smith, and M. Shafi, "Capacity limits and performance analysis of cognitive radio with imperfect channel knowledge," *IEEE Trans. Veh. Technol.*, vol. 59, no. 4, pp. 1811–1822, May 2010.
- [13] A. G. Marques, G. B. Giannakis, L. M. Lopez-Ramos, and J. Ramos, "Stochastic resource allocation for cognitive radio networks based on imperfect state information," in *Proc. IEEE ICASSP*, Prague, Czech Republic, May 2011, pp. 3196–3199.
- [14] M. Abdallah, A. Salem, M.-S. Alouini, and K. A. Qaraqe, "Adaptive rate transmission for spectrum sharing system with quantized channel-state information," in *Proc. 45th Annu. CISS*, Baltimore, MD, Mar. 2011, pp. 1–5.
- [15] A. G. Marques, X. Wang, and G. B. Giannakis, "Dynamic resource management for cognitive radios using limited-rate feedback," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3651–3666, Sep. 2009.
- [16] R. Zhang, "On peak versus average interference power constraints for protecting primary users in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 2112–2120, Apr. 2009.
- [17] R. Chen, J. G. Andrews, R. W. Heath, and A. Ghosh, "Uplink power control in multicell spatial multiplexing wireless systems," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2700–2711, Jul. 2007.
- [18] B. L. Ng, J. S. Evans, S. Hanly, and D. Aktas, "Distributed downlink beamforming with cooperative base stations," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5491–5499, Dec. 2008.
- [19] Y. Y. He and S. Dey, "Power allocation in spectrum sharing cognitive radio networks with quantized channel information," *IEEE Trans. Commun.*, vol. 59, no. 6, pp. 1644–1656, Jun. 2011.
- [20] Y. Y. He and S. Dey, "Throughput maximization in cognitive radio with limited feedback in the interference-limited regime," in *Proc. EW*, Vienna, Austria, Apr. 2011, pp. 1–8.
- [21] A. Khoshnevis and A. Sabharwal, "Performance of quantized power control in multiple-antenna systems," in *Proc. IEEE ICC*, Paris, France, Jun. 2004, pp. 803–807.
- [22] K. S. Ahn and R. W. Heath, Jr., "Performance analysis of maximum-ratio combining with imperfect channel estimation in the presence of cochannel interferences," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1080–1085, Mar. 2009.
- [23] S. Kotz, N. Balakrishnan, and N. L. Johnson, *Continuous Multivariate Distributions—Volume 1: Models and Applications*, 2nd ed. New York: Wiley, 2000.
- [24] A. H. Nuttall, "Some integrals involving the Q_M function (Corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 1, pp. 95–96, Jan. 1975.
- [25] H. A. David and H. N. Nagaraja, *Order Statistics*, 3rd ed. New York: Wiley Interscience, 2003.
- [26] L. Lin, R. D. Yates, and P. Spasojevic, "Adaptive transmission with discrete code rates and power levels," *IEEE Trans. Commun.*, vol. 51, no. 12, pp. 2115–2125, Dec. 2003.
- [27] R. Zhang, S. Cui, and Y. Liang, "On ergodic sum capacity of fading cognitive multiple access and broadcast channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5161–5178, Nov. 2009.
- [28] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel-side information," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.
- [29] M. A. Khojastepour and B. Aazhang, "The capacity of average and peak power constrained fading channels with channel-side information," in *Proc. IEEE WCNC*, Mar. 2004, vol. 1, pp. 77–82.
- [30] M. Khoshnevisan and J. N. Laneman, "Power allocation in wireless systems subject to long-term and short-term power constraints," in *Proc. IEEE ICC*, Kyoto, Japan, Jun. 2011, pp. 1–5.
- [31] Fed. Commun. Comm.: Rules and Regulations. [Online]. Available: <http://www.fcc.gov/oet/info/rules/>
- [32] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1310–1322, Jul. 2006.
- [33] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific, 1999.
- [34] R. Cendrillon, W. Yu, M. Noonen, J. Verlinden, and T. Bostoen, "Optimal multiuser spectrum balancing for digital subscriber lines," *IEEE Trans. Commun.*, vol. 54, no. 5, pp. 922–933, May 2006.



YuanYuan He (S'08) received the B.Eng. degree in communication engineering from Yanshan University, Qinhuangdao, China, in 2005, the M.Eng. degree in electronic engineering from Harbin Institute of Technology, Harbin, China, in 2007, and the Ph.D. degree from the University of Melbourne, Melbourne, Vic., Australia, in 2011.

She is currently a Postdoctoral Research Fellow with the Department of Electrical and Electronic Engineering, University of Melbourne. Her research interests include wireless communications, particularly resource allocation, cognitive radio, and limited feedback.



Subhrakanti Dey (SM'06) was born in India in 1968. He received the B.Tech. and M.Tech. degrees from the Indian Institute of Technology, Kharagpur, India, in 1991 and 1993, respectively, and the Ph.D. degree from the Australian National University, Canberra, A.C.T., Australia, in 1996.

From September 1995 to September 1997 and from September 1998 to February 2000, he was a Postdoctoral Research Fellow with the Department of Systems Engineering, Australian National University. From September 1997 to September 1998, he was a Postdoctoral Research Associate with the Institute for Systems Research, University of Maryland, College Park. Since February 2000, he has been with the Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Vic., Australia, where he is currently a Full Professor. He currently serves on the editorial board of *Elsevier Systems and Control Letters*. His research interests include networked control systems, wireless communications and networks, signal processing for sensor networks, and stochastic and adaptive estimation and control.

Dr. Dey was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE TRANSACTIONS ON AUTOMATIC CONTROL.