Maximize Spectrum Efficiency in Underlay Coexistence With Channel Uncertainty

Shaoran Li, *Student Member, IEEE*, Yan Huang, *Student Member, IEEE*, Chengzhang Li, *Student Member, IEEE*, Brian A. Jalaian, *Member, IEEE*, Y. Thomas Hou, *Fellow, IEEE*, Wenjing Lou, *Fellow, IEEE*, and Stephen Russell

***Abstract*— We consider an underlay coexistence scenario where secondary users (SUs) must keep their interference to the primary users (PUs) under control. However, the channel gains from the PUs to the SUs are uncertain due to a lack of cooperation between the PUs and the SUs. Under this circumstance, it is preferable to allow the interference threshold of each PU to be violated occasionally as long as such violation stays below a probability. In this article, we employ Chance-Constrained Programming (CCP) to exploit this idea of occasional interference threshold violation. We assume the uncertain channel gains are only known by their mean and covariance. These quantities are slow-changing and easy to estimate. Our main contribution is to introduce a novel and powerful mathematical tool called Exact Conic Reformulation (ECR), which reformulates the intractable chance constraints into tractable convex constraints. Further, ECR guarantees an equivalent reformulation from linear chance constraints into deterministic conic constraints without the limitations associated with Bernstein Approximation, on which our research community has been fixated on for years. Through extensive simulations, we show that our proposed solution offers a significant improvement over existing approaches in terms of performance and ability to handle channel correlations (where Bernstein Approximation is no longer applicable).**

***Index Terms*— Channel uncertainty, underlay coexistence, spectrum sharing, chance-constrained programming (CCP).**

I. Introduction

UNDERLAY is one of the most important schemes to achieve coexistence and to improve spectrum efficiency [2]. In underlay coexistence, primary and secondary users (PUs and SUs) are allowed to transmit simultaneously on the same spectrum and the SUs must keep their interference to the PUs under control (i.e., below a threshold). A key benefit of underlay is that no new requirement (neither hardware nor software) is needed for the PUs to achieve coexistence. As a result, underlay is ideal for incremental deployment where a secondary network (new infrastructure) is added on top of a primary network (existing infrastructure) in the same geographical area. Due to this benefit, underlay coexistence has attracted many efforts from the research community (see, e.g., [3]–[6]).

However, such benefits do not come without new technical challenges. To achieve underlay, we must ensure that there is no additional burden on the PUs to cooperate with the SUs. The SUs must take sole responsibility of keeping their interference to the PUs under a threshold, without requiring any cooperation from the PUs. This can be accomplished through transmission power control, based on Channel State Information (CSI) solely on the SUs side. However, in the absence of any cooperation with the PUs, an accurate estimation of CSI by the SUs is impossible. Specifically, the channel gains from the PUs to the SUs should be modeled as random variables (RVs). With such uncertainty in CSI, how to exercise power control by the SUs to protect the PUs is a challenging problem.

On the other hand, in many situations, we notice that occasional violations of the interference threshold are not fatal to the PUs. For example, today's communication systems always use error correction codes (e.g., Turbo code and LDPC code) that is capable of recovering the transmitted bits (to some extent) in the presence of interference [7]. Further, media-rich applications (e.g., audio and video streaming) are quite lenient to occasional transmission errors and packet loss. This is because, for such applications, human biological hearing, and vision systems are quite tolerable to occasional distortions [8] and there are numerous techniques to mitigate or conceal their adverse impacts [8], [9].

In the literature, there are three approaches to address channel uncertainty, namely, *stochastic optimization, worst-case optimization* and *chance-constrained programming (CCP)*. Stochastic optimization assumes the distributions of the RVs are known and rely on the corresponding distribution functions to solve the problem [10]. For example, in [11], the wireless channel was assumed to have log-normal shadowing and Nakagami small-scale fading while in [12], it was assumed to have Rayleigh fading. However, most wireless channels in reality do not exactly follow the assumed distributions. Consequently, a blind application of these simplified distribution models could lead to misleading results (either overly optimistic or pessimistic). Further, the formulated optimization problem could still be very hard to solve depending on the structure of the distributions (e.g., non-convex).

As for worst-case optimization, the RVs are assumed to have some known boundaries (upper or lower bounds) and this approach only focuses on the worst-case that is usually associated with such boundaries [13]. For instance, in [14], the authors relaxed the interference constraints in underlay

Manuscript received December 26, 2019; revised July 11, 2020 and November 17, 2020; accepted December 13, 2020; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor G. Iosifidis. Date of publication January 12, 2021; date of current version April 16, 2021. This work was supported in part by the U.S. Army Research Laboratory under Grant W911NF1820293. The work of Y.T. Hou was supported in part by the NSF under Grant CNS-1642873. The work of W. Lou and Y.T. Hou was supported in part by the Virginia Commonwealth Cyber Initiative (CCI). An abridged version of this article appeared in the Proc. ACM MobiHoc, Catania, Italy, July 2019. *(Corresponding author: Y. Thomas Hou.)*

Shaoran Li, Yan Huang, Chengzhang Li, Y. Thomas Hou, and Wenjing Lou are with Virginia Tech, Blacksburg, VA 24061 USA (e-mail: thou@vt.edu).

Brian A. Jalaian and Stephen Russell are with the CCDC Army Research Laboratory, Adelphi, MD 20783 USA.

Digital Object Identifier 10.1109/TNET.2020.3047760

scenario to linear constraints by defining maximum estimation errors. Although worst-case optimization provides the simplest mathematical formulations, it often gives overly conservative performance due to its focus on upper/lower bounds, which may hardly (or rarely) occur in practice. Further, worst-case optimization will not work when channel models are either unbounded (e.g., Rayleigh fading) or accurate estimation of the bounded set is difficult.

The third approach is called *Chance-Constrained Programming* (CCP). Although CCP was developed in the mathematics community more than six decades ago [15], it has only found applications in wireless communication and networking community in the 2000's [16]. In contrast to stochastic programming and worst-case optimization, CCP can be applied even when only a subset of statistics (e.g., mean and covariance) and properties (e.g., symmetricity and unimodality) is available. Different combinations of these statistics and properties are sufficient to make CCP applicable and as a result, CCP offers great flexibility in terms of problem settings. Another strength of CCP is to allow occasional violations of certain constraints, as long as the violation probability is smaller than a given probability upper bound. This small probability upper bound is termed *risk level.* By choosing different risk levels, we are able to explore a trade-off between achievable performance and tolerance in constraints violations. It is easy to see that CCP effectively circumvents the fundamental limitation associated with worst-case optimization.

However, the fruits of CCP come with a price. A major challenge in CCP is that chance constraints are usually mathematically intractable due to its probabilistic nature and absence of distribution knowledge. A critical step in solving CCP successfully is, therefore, to reformulate (substitute) the chance constraints into deterministic constraints. By doing so, we convert the CCP into a tractable optimization problem and can then design a customized algorithm for the deterministic problem.

In the literature, there are a number of methods to achieve this reformulation (substitution), such as Chebyshev inequalities, Markov inequalities [17], conditional value-at-risk [18], and Bernstein Approximation [19]. Among them, Bernstein Approximation is the state-of-the-art technique to address channel uncertainty in our research community (see, e.g., [20]–[23]). It performs such substitution by treating each RV separately (assuming they are independent and bounded) and deriving deterministic constraints after solving additional optimization problems regarding each RV.

However, there are a number of serious limitations associated with Bernstein Approximation. First, Bernstein Approximation explicitly requires that the RVs must be independent of each other. But this assumption does not always hold as correlations among uncertain RVs (e.g., CSI of sub-channels in our problem) are common and cannot be ignored. Second, the performance of Bernstein Approximation depends heavily on the knowledge of the boundaries of uncertain RVs [19], which is hard to obtain in many cases. Finally, due to its generic nature, Bernstein Approximation does not explore the unique structure of linear CCP, which can characterize a broad class of problems in practice. Consequently, results by Bernstein Approximation tend to be conservative, as to be shown in our simulation results (Section VII).

In this article, we study an underlay coexistence problem using CCP. The goal is to maximize the spectrum efficiency of picocells while keeping SUs' occasional violation of interference threshold within a given risk level. We introduce a novel technique called *Exact Conic Reformulation* (ECR) to reformulate the chance constraints with deterministic constraints. The reformulated deterministic constraints from ECR introduce no relaxation errors (i.e., "exact") and are convex second-order cones (i.e., "conic"). We show that the proposed ECR allows us to handle more practical and general problem settings and to achieve better performance when compared to Bernstein Approximation. The main contributions of this article are summarized as follows:

- To address channel uncertainty in underlay coexistence, we employ CCP under the assumption that only mean and covariance of the uncertain channel gains are available. The advantage of using mean and covariance is that they are rather time-invariant and can be readily estimated.
- To reformulate the intractable chance constraints, we introduce ECR – which offers mathematically "exact" conic reformulation. ECR is able to reformulate the original chance constraints into deterministic constraints while maintaining the same (hence "exact") optimization space w.r.t. decision variables.
- We compare our proposed ECR with the state-of-the-art Bernstein approximation. ECR allows for unbounded and correlated RVs to describe the interference while Bernstein approximation cannot. Further, ECR enables direct reformulation from the original problem without solving additional optimization problems (as required by Bernstein Approximation). To the best of our knowledge, this is the first paper that has successfully addressed the limitations (e.g., correlations and conservativeness) of Bernstein Approximation, on which our research community has become fixated for years. As such, ECR represents a new and more effective technique to solve CCP problems in wireless networks.
- Through extensive simulations, we show that our two solutions (predicated on ECR) outperform Bernstein Approximation in two aspects. First, our solutions achieve higher spectrum efficiency (33% and 30% on average) when channel gains are independent (where Bernstein Approximation is applicable). Second, in the correlated scenario (where Bernstein Approximation is no longer applicable), our proposed approaches can still offer competitive solutions while meeting the target risk level. Therefore, our proposed approaches are able to reap the full benefits of CCP in both general and practical settings thanks to our novel ECR technique.

We organize the remainder of this article as follows. In Section II, we review related work. In Section III, we introduce the system model of our underlay coexistence problem. In Section IV, we offer a CCP formulation to our problem. In Section V, we present the novel ECR technique for the CCP formulation. In Section VI, we present solutions to the equivalent (reformulated) deterministic optimization problem. In Section VII, we present simulation results. Section VIII concludes this article.

II. Related Work

There is a large body of work studying various problems in underlay coexistence [24]–[27]. We categorize these works based on whether they assumed perfect channel knowledge or not (i.e., imperfect channel knowledge, or channel uncertainty).

The majority of works that assumed perfect channel knowledge have looked into problems on spectrum efficiency [3], [28], energy efficiency [29], beamforming design [30], secure transmission [31], to name a few. These works assumed that the channel is static and channel gains are given constants. This allowed simplifying the underlying problem and analysis significantly. However, such an assumption does not hold in reality as channel dynamics are common and must be observed for problems dealing with power control or scheduling on a small time scale.

For those works that considered channel uncertainty (or imperfect channel knowledge) in underlay, they employed mathematical tools such as stochastic programming [11], [12], [32], worst-case optimization [14], [33], [34] and CCP [20]–[23], [35], [36]. Though a learning-based algorithm seems relevant, to the best of our knowledge, there is no existing work that addressed channel uncertainty in underlay. Thus, we will only discuss related work for the above three approaches.

In stochastic programming, exact knowledge of channel gain distributions is needed. Some widely used distributions include Rayleigh fading [12], Rician fading [32], and Nakagami fading [11]. Such an assumption of channel gain distributions could be problematic as they may not match how channels actually behave in reality.

In worst-case optimization, the channel gains are assumed to be inside certain bounded sets, such as norm bounded region [14], [33], polyhedron [14] and maximum Kullback-Leibler divergence [34]. As discussed in Section I, solutions following worst-case optimization tend to be overly conservative, as will be demonstrated in our numerical study in Section VII.

In CCP for underlay, some previous works include [20]–[23], [35], [36]. As expected, all these works assumed unknown distributions and allowed occasional violations of the interference threshold. To reformulate the CCP problem, they all employed Bernstein Approximation [19], where safety approximations are derived to replace the chance constraints (with probabilities) by assuming the RVs (e.g., channel gains in underlay) are independent and bounded. However, correlations among channel gains are possible depending on the physical paths. Further, reformulation from Bernstein Approximation introduces relaxation errors, which affect its performance (the issue of conservativeness). In contrast, we show in this article that our proposed ECR can mitigate the limitations of Bernstein Approximation (independent RVs and conservativeness).

III. System Model

Consider a number of picocells residing within a macrocell as shown in Fig. 1. An example of such a scenario is that each picocell is installed as a set-up box inside a residential unit [37], [38]. Users connected directly to the macro base station (BS) are called PUs while users connected to the pico BSs are called SUs. As for spectrum allocation among the macro- and picocells, we follow a well-known scheme called "fractional frequency reuse" in the literature [39]–[41]. Specifically, each picocell is assigned only a fraction of the spectrum allocated to the macrocell and adjacent picocells use different frequency bands to avoid the inter-cell interference between neighbors (as shown in different colors of footprint in Fig. 1).

In an underlay coexistence paradigm [2], the SUs must ensure that the normal operations of the PUs are not interrupted and that the PUs are unaware of the presence of the SUs. We consider a TDD system that the uplink and downlink transmissions of SUs and PUs are performed in the same spectrum. Although the macrocell and picocells can choose uplink and downlink directions arbitrarily, we consider the most difficult problem when transmissions in the macrocell are in downlink and transmissions in the picocells are in uplink. Under this circumstance, there are multiple SUs transmitting to their pico BSs and multiple nearby PUs need to be protected. Here, exercising power control inside each picocell is the key to ensure normal network operations for the PUs (and to increase the spectrum efficiency for the SUs). Once the reader understands the proposed solutions in this article, they can easily extend our proposed solutions to other transmission cases, such as:

- Downlink in both the macrocell and the picocells. Mathematically, this is equivalent to one SU per picocell in our model.
- Uplink in both the macrocell and the picocells. Mathematically, this is equivalent to one PU in our model.
- Uplink in the macrocell and downlink in the picocells, Mathematically, this is equivalent to one PU and one SU in our model.

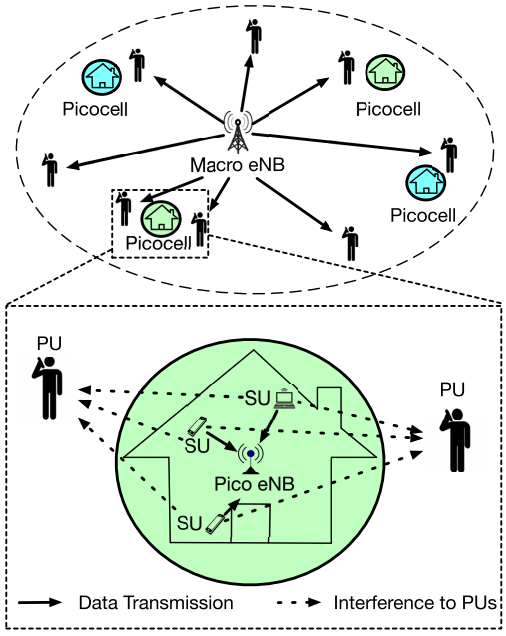


Fig. 1. A network topology for underlay: Multiple picocells inside a macrocell (upper portion) and multiple SUs in a picocell (lower portion).

It is well-known that information on channel gains inside the network is needed for power control. Following cellular terminology, we consider the time domain is divided into transmission time intervals (TTIs) and the frequency domain is divided into sub-channels. Each sub-channel over one TTI is called a resource block (RB). With respect to an RB, channel gains (between the SUs and the pico BS) are location-dependent and vary over time (due to multipath). For different RBs, one would expect a certain degree of correlations among the channel gains due to similar physical paths. Thus, each SU must perform channel sensing before transmission to estimate the channel gains. There are two types of channel gains, namely transmission channel gains (from the SUs to the pico BS) and interference channel gains (from the PUs to the SUs). The transmission channel gains can be estimated by traditional channel sounding procedure such

as pilot-based training (with cooperations between the SUs and their associated pico BSs). The interference channel gains can only be measured based on known signals (e.g., pilots and preambles used in the macrocell) and channel reciprocity property [42]. To differentiate different PUs, a SU can exploit the orthogonality in pilots as well as location techniques by leveraging existing spectrum sensing algorithms [43], [44].

After channel sensing, all the CSI will be reported to the pico BS through a dedicated control channel. Upon receiving the CSI, the pico BS will find the optimal scheduling decisions (in frequency and/or time) and transmission powers for the SUs. The goal is to maximize spectrum efficiency while keeping the aggregate interference from the SUs to each nearby PU below a threshold (see Fig. 1). The optimal solution for scheduling and power control will be sent to the SUs by the pico BSs and then the SUs can execute their uplink transmissions based on this solution. In the case of fractional frequency reuse, we only need to focus on one picocell since neighboring picocells operate independently on non-overlapping frequency bands.

Consider one picocell (the lower portion of Fig. 1) with several nearby PUs. The signal from the macro BS to the PUs are considered as "noise" by the SUs and the pico BS inside the picocell. Such a noise will be included in the respective channel gain calculations between the SUs and the pico BS. To control the aggregate interference to each PU, the CSI from the SUs to each PU is needed. Since there is no feedback from the PUs to the SUs, the SUs can only estimate CSI to the PUs unilaterally based on overhearing known signals from the PUs and channel reciprocity. As a result, the interference channel gains from the SUs to the PUs, the key parameters for controlling the transmission powers of the SUs, can only be characterized as RVs.

In this work, we assume that the mean and covariance of the interference channel gains are available via online estimation. Specifically, whenever the SUs overhear the signals transmitted by a nearby PU during the PU's transmission, the SUs can estimate the corresponding interference channel gains in current TTI based on channel reciprocity. Then the SUs will update the mean and covariance of the interference channel gains. Since these statistics are slow-changing, it is reasonable to assume such statistics are up-to-date at the SUs through continuous tracking of the mean and covariance over time. Thus, it is reasonable and more practical to use the mean and covariance than assuming certain types of distributions for the interference channel gains.

Further, we assume the PUs can tolerate occasional threshold violation as long as the probabilities of such violations are smaller than given constants. For practical purposes, such occasional violations are tolerable, as discussed in Section I. As we shall see in the next section, such tolerance can be formulated as chance constraints under CCP.

IV. Mathematical Modeling and Formulation

In this section, we formulate the optimization problem using CCP. Consider a single picocell in the lower portion of Fig. 1. We are interested in maximizing spectrum efficiency for the SUs in a picocell while keeping their interference threshold violations (to each nearby PU) under small probabilities. The notations used in this article are summarized in Table I.

Denote N as the number of SUs in the picocell, J as the number of nearby PUs, and M as the number of RBs per TTI.

TABLE I
NOTATIONS

Symbol	Definition
c_i^m	Throughput of SU i toward its pico BS on RB m
g_{ij}^m	Interference channel gain from SU i to PU j on RB m
$\mathbf{g}_j$	A column vector: $\left[g_{1j}^1, g_{1j}^2, \cdots, g_{1j}^M, g_{2j}^1, \cdots, g_{Nj}^M\right]^T$
$\bar{\mathbf{g}}_j$	Mean of channel gain vector $\mathbf{g}_j$
$\mathbf{R}_j$	Covariance matrix of channel gain vector $\mathbf{g}_j$
h_i^m	Transmission channel gain from SU i to its pico BS on RB m
I_j	Interference threshold for PU j
J	Number of nearby PUs
$\mathcal{J}$	A set of integers from 1 to J: $\{1, 2, 3, \cdots, J\}$
L_j	Maximum sub-carrier spacing of correlated RBs from PU j
M	Number of RBs for transmission in the picocell
$\mathcal{M}$	A set of integers from 1 to M: $\{1, 2, 3, \cdots, M\}$
N	Number of SUs in the picocell
$\mathcal{N}$	A set of integers from 1 to N: $\{1, 2, 3, \cdots, N\}$
p_i^m	Transmission power from SU i to pico BS on RB m
$\mathbf{p}$	A column vector: $\left[p_1^1, p_1^2, \cdots, p_1^M, p_2^1, \cdots, p_N^M\right]^T$
$P_i^{\max}$	Maximum transmission power of SU i over all RBs
w_i	Weight of SU i
x_i^m	A binary variable indicating whether or not SU i transmits on RB m
ϵ_j	Risk level (probability upper bound) of violating PU j's interference threshold I_j

At the start of each TTI, the pico BS will run a scheduling algorithm to allocate the available RBs to the SUs and decide the SUs' transmission powers for uplink transmissions. We consider single user OFDMA where one RB can be allocated to at most one SU but one SU may be allocated with multiple RBs in each TTI. A popular scheduling objective is to achieve long-term proportional fair (PF) among SUs' throughput [45]. This is equivalent to maximizing a weighted sum of throughput in each TTI, with the weight of each SU being updated at the beginning of each TTI based on their long-term data rates. Thus, the objective function used in this article is the weighted sum (with given weights) of throughput for all SUs in the picocell.

Denote x_i^m as a binary variable to indicate whether or not SU i will transmit to the pico BS on RB m, i.e.,

$$x_i^m = \begin{cases} 1 & \text{if SU } i \text{ will transmit to the pico BS on RB } m, \\ 0 & \text{otherwise.} \end{cases}$$

Under single user OFDMA, each RB (say m) can only be assigned to at most one SU. We have

$$\sum_{i \in \mathcal{N}} x_i^m \leq 1 \qquad (m \in \mathcal{M}). \tag{1}$$

Denote p_i^m as the transmission power from SU i to the pico BS on RB m. Clearly, p_i^m must be zero if x_i^m is zero because the transmission power must follow the scheduling decision. Further, it is common that each SU's device has a maximum power limit over all RBs. Denote $P_i^{\max}$ as the maximum power of SU i. Then we have

$$0 \leq p_i^m \leq x_i^m P_i^{\max} \qquad (i \in \mathcal{N},\ m \in \mathcal{M}), \tag{2}$$

and

$$\sum_{m \in \mathcal{M}} p_i^m \leq P_i^{\max} \qquad (i \in \mathcal{N}), \tag{3}$$

where $\mathcal{N} = \{1, 2, \cdots, N\}$.

Denote g_{ij}^m as the interference channel gain (a RV) from SU i to PU j on RB m and I_j as the interference threshold (constant) for PU j ($j \in \mathcal{J}$ where $\mathcal{J} = \{1, 2, \cdots, J\}$). Under CCP, the aggregate interference from the SUs to PU j is allowed to occasionally violate I_j but the probability of such a violation must be below a target (small) probability. Denote this target probability as ϵ_j, which is also called *risk level*. This behavior, in its complementary form, can be formulated by the following chance constraints:

$$\mathbb{P}\left\{\sum_{i\in\mathcal{N}}\sum_{m\in\mathcal{M}} g_{ij}^m p_i^m \le I_j\right\} \ge 1-\epsilon_j \qquad (j\in\mathcal{J})\,, \tag{4}$$

where $\mathbb{P}\{\cdot\}$ denotes the probability function w.r.t. g_{ij}^m. In (4), p_i^m's are decision variables; I_j and ϵ_j are constants; g_{ij}^m's are RVs with known mean and covariance. Note that g_{ij}^m's are independent of p_i^m's, I_j's, and ϵ_j's. Constraints (4) state that the aggregate interference from the SUs to PU j must be smaller than the threshold I_j with a probability of at least $1-\epsilon_j$. Clearly, proper power control of the SUs is the key to meet these chance constraints. This risk level ϵ_j is a small number that serves as the upper bound for the violation probability. It could vary over a wide range (e.g., 0.01 to 0.5) depending on the application of PU j. A higher ϵ_j means a larger tolerance to violation of interference threshold (and corresponding to a larger optimization space) and hence higher spectrum efficiency.

Per our earlier discussion, in constraints (4), the interference channel gains g_{ij}^m's are modeled as RVs with known mean (a $MN\times 1$ vector) and covariance (a $MN\times MN$ matrix) via online estimation. Since g_{ij}^m's are the interference channel gains, they can be independent or correlated with each other due to similar physical paths. In reality, we can check the covariance matrix of g_{ij}^m's to see whether they are independent or correlated. g_{ij}^m's are independent with each other if and only if the covariance matrix is diagonal.

Assume each RB occupies the same bandwidth, which we normalize to 1 unit. Denote h_i^m as the transmission channel gain from SU i to the pico BS on RB m, including both interferences from the macro BS and thermal noise at the pico BS. Then the capacity of SU i on each RB can be calculated by Shannon theorem. In this work, we use weighted sum of capacity as our objective function to maximize the spectrum efficiency. Denote w_i as the weight of SU i in the current TTI, which is a given constant (but may differ over TTIs). Based on the above notations, our problem can be formulated as follows:

$$\begin{aligned}(\text{P1})\quad \max_{x_i^m, p_i^m}\ & \sum_{i\in\mathcal{N}}\sum_{m\in\mathcal{M}} w_i \log_2(1+h_i^m p_i^m)\\ \text{s.t.}\ & \text{Scheduling constraints } (1),\\ & \text{Transmission power constraints } (2),(3),\\ & \text{Interference power constraints } (4),\\ & x_i^m\in\{0,1\},\quad p_i^m\ge 0.\end{aligned}$$

Clearly, the main challenge in this optimization problem lies in chance constraints (4) due to its probabilistic nature. Although we have the mean and covariance of g_{ij}^m's, we do not have the knowledge of their distributions.[1] For the same mean and covariance, there is an infinite number of possible distributions. Since it is impossible to enumerate all possible distributions for constraints (4), P1 is intractable.

[1]Even if we had the knowledge of distributions, it remains unclear if problem P1 could be solved optimally and efficiently, which heavily depends on the distribution density functions.

V. A Novel Reformulation of Chance Constraint

In general, to solve an intractable CCP problem, we need to perform reformulation of the chance constraints based on available knowledge (e.g., mean and covariance in our problem). In this section, we introduce a novel technique to perform *exact reformulation* of chance constraints (4) as *second-order cones*. By its name *Exact Conic Reformulation* (ECR), we mean that the optimization space of the newly derived deterministic constraints (with tractable conic formulations) is the same as that from the original chance constraints. In other words, for some worst-case distributions, the threshold violation probability is exactly the risk level ϵ_j, while for all other distributions, the threshold violation probability is smaller than ϵ_j [46]. Besides, the derived deterministic constraints belong to second-order cones which are guaranteed to be convex.

In constraints (4), the j-th constraint is the probabilistic interference guarantee for PU j. Since the interference guarantees for the PUs are independent of each other, we can perform ECR for each chance constraint w.r.t. PU j. For ease of exposition, we rewrite j-th constraint of (4) into its vector form

$$\mathbb{P}\left\{\mathbf{g}_j^T\mathbf{p} > I_j\right\} \le \epsilon_j\,, \tag{5}$$

where superscript "T" denotes transposition, $\mathbf{p}$ is an $MN\times 1$ column vector given as

$$\mathbf{p} = \left[p_1^1, p_1^2, \cdots, p_1^M, p_2^1, \cdots, p_N^M\right]^T\,, \tag{6}$$

which represents the MN transmission powers from SUs (over all RBs). $\mathbf{g}_j$ is also an $MN\times 1$ column vector given as

$$\mathbf{g}_j = \left[g_{1j}^1, g_{2j}^2, \cdots, g_{1j}^M, g_{2j}^1, \cdots, g_{Nj}^M\right]^T\,, \tag{7}$$

which represents MN random channel gains from the SUs (over all RBs) to PU j.

Denote $\overline{\mathbf{g}}_j$ (a $MN\times 1$ column vector) and $\mathbf{R}_j$ (a $MN\times MN$ matrix) as the mean and covariance of $\mathbf{g}_j$, i.e., $\mathbf{g}_j \sim (\overline{\mathbf{g}}_j, \mathbf{R}_j)$. Note that $\mathbf{R}_j$ is always symmetric positive semi-definite and is a diagonal matrix when the interference channel gains g_{ij}^m's are independent of each other. Since constraint (5) is satisfied for $\mathbf{g}_j$ under all possible distributions with $\mathbf{g}_j \sim (\overline{\mathbf{g}}_j, \mathbf{R}_j)$, we have

$$\sup_{\mathbf{g}_j\sim(\overline{\mathbf{g}}_j,\mathbf{R}_j)} \mathbb{P}\left\{\mathbf{g}_j^T\mathbf{p} > I_j\right\} \le \epsilon_j\,, \tag{8}$$

where the "sup" is taken over all distributions of $\mathbf{g}_j$ with mean $\overline{\mathbf{g}}_j$ and covariance $\mathbf{R}_j$.

Denote ξ_j as a scalar RV which is defined as $\xi_j = \mathbf{g}_j^T\mathbf{p} - \overline{\mathbf{g}}_j^T\mathbf{p}$. It is easy to see that ξ_j has mean 0 and variance $\mathbf{p}^T\mathbf{R}_j\mathbf{p}$, i.e., $\xi_j \sim (0, \mathbf{p}^T\mathbf{R}_j\mathbf{p})$. For ease of exposition, denote ϕ_j as $\phi_j = I_j - \overline{\mathbf{g}}_j^T\mathbf{p}$. Since we require (5) to hold for all possible g_{ij}^m's, we must have

$$\sup_{\xi_j\sim(0,\mathbf{p}^T\mathbf{R}_j\mathbf{p})} \mathbb{P}\left\{\xi_j > \phi_j\right\} \le \epsilon_j\,. \tag{9}$$

To perform an exact reformulation of chance constraint (9), we need to derive a closed-form expression for the supremum

of violation probability where ξ_j can take any form of distribution with mean 0 and variance $\mathbf{p}^T\mathbf{R}_j\mathbf{p}$. We wish to choose $\mathbf{p}$ so that this closed-form supremum is upped bounded by ϵ_j. Since this closed-form expression has no randomness, we have a deterministic constraint of the decision variables in $\mathbf{p}$.

To derive a closed-form expression for the "sup" in (9), we rely on Cantelli's Inequality [47] (also known as one-sided Chebyshev inequality [48]). Cantelli's Inequality states that for a given random variable ξ, with known mean μ and variance σ^2, we have

$$\begin{cases} 1-\dfrac{\sigma^2}{\sigma^2+\lambda^2} \le \mathbb{P}\left\{\xi-\mu\ge\lambda\right\} \le 1 & \text{if } \lambda<0, \\ 0 \le \mathbb{P}\left\{\xi-\mu\ge\lambda\right\} \le \dfrac{\sigma^2}{\sigma^2+\lambda^2} & \text{if } \lambda\ge 0. \end{cases} \tag{10}$$

The upper bounds in Cantelli's Inequility (10) are tight since we can find worst-case distributions of ξ that achieve these upper bounds. For instance, when $\lambda<0$, an example of worst-case distribution of ξ is given as

$$\mathbb{P}\{\xi=\frac{\lambda}{2}+\mu\}=\frac{4\sigma^2}{4\sigma^2+\lambda^2}\ ,$$
$$\mathbb{P}\{\xi=\frac{-2\sigma^2}{\lambda}+\mu\}=\frac{\lambda^2}{4\sigma^2+\lambda^2}.$$

When $\lambda\ge 0$, an example of worst-case distribution of ξ is given as

$$\mathbb{P}\{\xi=\frac{-\sigma^2}{\lambda}+\mu\}=\frac{\lambda^2}{\sigma^2+\lambda^2}\ ,\ \mathbb{P}\{\xi=\lambda+\mu\}=\frac{\sigma^2}{\sigma^2+\lambda^2}.$$

We encourage readers to verify the violation probability in these distributions.

Based on Cantelli's Inequality, if we set $\xi=\xi_j, \mu=0$, $\lambda=\phi_j$ in (10) and note that

$$\sup_{\xi_j\sim(0,\mathbf{p}^T\mathbf{R}_j\mathbf{p})} \mathbb{P}\left\{\xi_j>\phi_j\right\} = \sup_{\xi_j\sim(0,\mathbf{p}^T\mathbf{R}_j\mathbf{p})} \mathbb{P}\left\{\xi_j\ge\phi_j\right\}, \tag{11}$$

we can derive a closed-from expression of our violation probability in (9), which we state in the following Lemma.

Lemma 1: A closed-form supremum of threshold violation probability in (9) *is given by*

$$\sup_{\xi_j\sim(0,\mathbf{p}^T\mathbf{R}_j\mathbf{p})} \mathbb{P}\left\{\xi_j>\phi_j\right\} = \begin{cases} 1 & \phi_j<0\ , \\ \dfrac{\mathbf{p}^T\mathbf{R}_j\mathbf{p}}{\phi_j^2+\mathbf{p}^T\mathbf{R}_j\mathbf{p}} & \phi_j\ge 0\ . \end{cases} \tag{12}$$

The significance of Lemma 1 is that we can convert an intractable probability in (9) to a deterministic equation in (12) without any relaxation (due to tightness of Cantelli's Inequality). We also know that there exist "worst-case distributions of ξ_j" that achieve (or asymptotically achieve) the supremum of the threshold violation probability $\mathbb{P}\left\{\xi_j>\phi_j\right\}$. For all other distributions, the threshold violation probability will be smaller than the supremum. In fact, we can derive some necessary conditions of the worst-case distributions, as stated in Property 1.

Property 1: **(Worst-case Distributions of ξ_j)** *A worst-case distribution of ξ_j must have the following properties:*
(i) If $\phi_j<0$, then $\mathbb{P}\{\xi_j\le\phi_j\}=0$;
(ii) If $\phi_j\ge 0$, then $\mathbb{P}\{\xi_j>\Phi_j\}=0$ for all $\Phi_j>\phi_j$.

Proof: (i) When $\phi_j<0$, it is a trivial case since otherwise we will have $\mathbb{P}\{\xi_j>\phi_j\}=1-\mathbb{P}\{\xi_j\le\phi_j\}<1$.

(ii) When $\phi_j\ge 0$, our proof is based on contradictions. Suppose (ii) doesn't hold and we have a worst-case distribution ξ_j^* and there exists $\Phi_j^*>\phi_j$ with $\mathbb{P}\{\xi_j^*=\Phi_j^*\}>0$.[2] Then we will construct another distribution ξ_j' with identical mean and variance as ξ_j^* such that $\mathbb{P}\{\xi_j'>\phi_j\}>\mathbb{P}\{\xi_j^*>\phi_j\}>0$ to claim a contradiction.

We first derive the upper bound of $\mathbb{P}\{\xi_j^*=\Phi_j^*\}$. Since $\mathbb{P}\{\xi_j^*=\Phi_j^*\}>0$ and $\mathbb{E}\{\xi_j^*\}=0$, we can always find an interval $\mathcal{B}\in(-\infty,0)$ such that

$$\mathbb{E}_{\{\xi_j^*\in\{\mathcal{B},\Phi_j^*\}\}}\xi_j^*=0\ . \tag{13}$$

Then we can calculate the probability and second-order statistics of ξ_j^* in $\{\mathcal{B},\Phi_j^*\}$, denoted as p and q^2 respectively. p and q^2 are calculated by

$$p=\mathbb{P}\{\xi_j^*\in\{\mathcal{B},\Phi_j^*\}\}\ ,\quad q^2=\mathbb{E}_{\{\xi_j^*\in\{\mathcal{B},\Phi_j^*\}\}}(\xi_j^*)^2. \tag{14}$$

Define a new random variable ζ on $\{\mathcal{B},\Phi_j^*\}$ with

$$\mathbb{P}\{\zeta=\phi\}=\frac{\mathbb{P}\{\xi_j^*=\phi\}}{p}\quad(\phi\in\{\mathcal{B},\Phi_j^*\})\ . \tag{15}$$

Clearly, ζ has mean 0 and variance $\frac{q^2}{p}$. Based on Cantelli's Inequality, we have

$$\max\mathbb{P}\{\xi_j^*=\Phi_j^*\}=p\cdot\max\mathbb{P}\{\zeta\ge\Phi_j^*\}=\frac{pq^2}{q^2+p(\Phi_j^*)^2}\ . \tag{16}$$

Then we construct a new random variable ξ_j' by replacing the part of ξ_j^*'s distribution on $\xi_j^*\in\{\mathcal{B},\Phi_j^*\}$ by

$$\mathbb{P}\{\xi_j'=-\frac{q^2}{p\Phi_j'}\}=\frac{p^2(\Phi_j')^2}{q^2+p(\Phi_j')^2}\ ,$$
$$\mathbb{P}\{\xi_j'=\Phi_j'\}=\frac{pq^2}{q^2+p(\Phi_j')^2}\ , \tag{17}$$

where $\phi_j<\Phi_j'<\Phi_j^*$. If can be easily verified that ξ_j' and ξ_j^* have the same mean and variance. Clearly, the violation probability from ξ_j' is higher than ξ_j^* since

$$\mathbb{P}\{\xi_j'>\phi_j\}-\mathbb{P}\{\xi_j'>\Phi_j'\}\ge\frac{pq^2}{q^2+p(\Phi_j')^2}-\frac{pq^2}{q^2+p(\Phi_j^*)^2}$$
$$>0\ . \tag{18}$$

This contradicts the assumption that ξ_j^* is a worst-case distribution. □

Property 1 states that when $\phi_j<0$, any worst-case distribution of ξ_j should task no values in the interval $(-\infty,\phi_j]$; when $\phi_j\ge 0$, any worst-case distribution of ξ_j only takes one value that is sufficiently close to ϕ_j as the possibilities of ξ_j in interval $(\phi_j,+\infty)$ are pushed to ϕ_j. Any distribution that does not satisfy Property 1 is not a worst-case distribution. Consequently, the resulting threshold violation probability will be smaller than the supremum and the gap depends on the distribution of ξ_j.

We note that there may exist many functions for the worst-case distributions even under Property 1 but they all have the same supremum of the threshold violation probability,

[2]Here we use the notation of a discrete RV. The same conclusion holds for a continuous RV through discretization.

as given in Lemma 1. Since we are not assuming any knowledge of the distributions in our problem setting, we cannot ignore the worst-case distributions. In other words, Lemma 1 is the best bound for our problem setting.

With Lemma 1 at hand, we can derive our reformulation by upper bounding the closed-form supremum of threshold violation probability by ϵ_j as in (9). Since $\epsilon_j < 1$, the case with $\phi_j < 0$ is always infeasible. So we only need to consider the case when $\phi_j \geq 0$. That is, chance constraint (9) can be replaced by the following two constraints:

$$\phi_j \geq 0 \,, \tag{19a}$$

$$\frac{\mathbf{p}^T \mathbf{R}_j \mathbf{p}}{\phi_j^2 + \mathbf{p}^T \mathbf{R}_j \mathbf{p}} \leq \epsilon_j \,. \tag{19b}$$

We can rewrite (19b) as

$$\frac{1-\epsilon_j}{\epsilon_j} \cdot \mathbf{p}^T \mathbf{R}_j \mathbf{p} \leq \phi_j^2 \,. \tag{20}$$

Taking the square root of both sides in (20) and considering (19a), we have

$$\sqrt{\frac{1-\epsilon_j}{\epsilon_j}} \cdot \sqrt{\mathbf{p}^T \mathbf{R}_j \mathbf{p}} \leq \phi_j \,. \tag{21}$$

Note that (21) implicitly implies that $\phi_j \geq 0$ and thus (19a) is no longer needed. Substituting $\phi_j = I_j - \mathbf{g}_j^T \mathbf{p}$ into (21), we have the following main result.

Theorem 1: **(ECR)** *With respect to the decision variables in* $\mathbf{p}$*, chance constraints (4) are equivalent to the following second-order cones*

$$\sqrt{\frac{1-\epsilon_j}{\epsilon_j}} \sqrt{\mathbf{p}^T \mathbf{R}_j \mathbf{p}} + \overline{\mathbf{g}}_j^T \mathbf{p} \leq I_j \quad (j \in \mathcal{J}) \,. \tag{22}$$

ECR says that in term of optimization space, the original chance constraints in (4) and the second-order cones in (22) are equivalent. Such equivalence is a direct result of Lemma 1.[3] Constraints (22) are deterministic constraints since the random interference channel gains g_{ij}^m's are eliminated and the mean $\overline{\mathbf{g}}_j$ and covariance matrix $\mathbf{R}_j$ are given constants. Through our derivation, we only rely on mean $\overline{\mathbf{g}}_j$ and covariance $\mathbf{R}_j$ of g_{ij}^m's without requiring any other assumptions.

Comparing ECR to Bernstein Approximation [19]–[23], ECR is more general as Bernstein Approximation explicitly assumes independence and boundaries.[4] Interestingly, constraints (22) appear quite similar to the deterministic constraints obtained by Bernstein Approximation, with the difference being the non-linear terms. This seemingly small difference is, in fact, the main reason for the conservativeness in Bernstein Approximation, as we shall show in simulations in Section VII.

We summarize the benefits of ECR over Bernstein Approximation as follows:

- Bernstein Approximation explicitly assumes that the RVs (interference channel gains g_{ij}^m's) must be independent while ECR can handle the case with correlated RVs. In reality, correlations among g_{ij}^m's are common due to similar physical paths.
- The RVs (g_{ij}^m's) must be bounded to be normalized in Bernstein Approximation while ECR does not need the boundaries and can handle unbounded RVs. In practice, the exact boundaries are also hard to estimate, which affects the overall performance. The authors in [20] proposed a method to handle RVs with unbounded supports by truncation. Unfortunately, such truncation is designed for a specific distribution and cannot be easily extended to other distributions.
- Reformulated constraints in Bernstein Approximation are safety approximations with relaxation errors to guarantee the risk level while ECR offers an exact reformulation with no relaxations (based on Lemma 1).
- In Bernstein Approximation, the parameters to derive the deterministic constraints must be obtained through an optimization problem, which requires additional efforts. In contrast, ECR enables direct reformulation from the original problem based on mean $\overline{\mathbf{g}}_j$ and covariance $\mathbf{R}_j$.

To use ECR for the reformulation of P1, we only need to replace chance constraints (4) by deterministic constraints (22). Then we obtain a deterministic maximization problem as follows:

$$\begin{aligned} \text{(P2)} \quad & \max_{x_i^m, p_i^m} \sum_{i \in \mathcal{N}} \sum_{m \in \mathcal{M}} w_i \log_2(1 + h_i^m p_i^m) \\ & \text{s.t.} \quad \text{Scheduling constraints } (1) \,, \\ & \qquad \text{Transmission power constraints } (2), (3) \,, \\ & \qquad \text{Interference power constraints from ECR } (22) \,, \\ & \qquad x_i^m \in \{0,1\}, p_i^m \geq 0 \,. \end{aligned}$$

P2 is a mixed-integer non-linear program (MINLP). In [49], the authors showed that a special case of P2 is NP-hard where no channel uncertainty is considered. In this special case, the covariance matrix $\mathbf{R}_j$ is set to $\mathbf{R}_j = 0$ and (22) becomes linear constraints. Since P2 is the general case, it must be NP-hard, which takes a long time to compute a solution. Thus, for efficient computation, we devise approximation solutions to P2 in the next section.

VI. Solving the Deterministic Optimization Problem

The main difficulties of P2 lie in the two nonlinear terms in the objective function and (22). In this section, we show how to linearize them and reduce the computation complexity.

A. Logarithmic Terms

To linearize the logarithmic terms in the objective function, we use piecewise linear functions for each logarithmic term. Denote c_i^m as SU i's normalized capacity to the pico BS on RB m (w.r.t. normalized RB bandwidth). Then we have

$$c_i^m = \log_2(1 + h_i^m p_i^m) \quad (i \in \mathcal{N},\ m \in \mathcal{M}) \,. \tag{23}$$

Recall that h_i^m is the transmission channel gain from SU i to the pico BS on RB m. We employ a convex hull to relax each logarithmic term $\log_2(1 + h_i^m p_i^m)$ [50]. Since P2 is a maximization problem, we only need to consider a series of linear constraints to upper bound the convex hull.

For each log term $\log_2(1 + h_i^m p_i^m)$, we break the interval of p_i^m (i.e., $[0, P_i^{\max}]$) into K equal-length sub-intervals, each with length $\frac{P_i^{\max}}{K}$. Then for the $K+1$ end points, we solve the

[3] Incidentally, our results in Theorem 1 has been previously developed independently by J. Pinter in [48] (Proposition 2.1) in his work on deterministic approximations of probability inequalities.

[4] We refer interested readers to [19] for more details of Bernstein Approximation.

tangential lines at each end point and obtain the $K+1$ linear functions as follows:

$$c_i^m \le \frac{1}{\ln 2}\cdot\left[\frac{Kh_i^m}{K+kP_i^{\max}}\cdot p_i^m + \ln\left(1+\frac{kP_i^{\max}}{K}\right) - \frac{kP_i^{\max}}{K+kP_i^{\max}}\right] \quad (k\in\mathcal{K}, i\in\mathcal{N}, m\in\mathcal{M}), \tag{24}$$

where $\mathcal{K}=\{0,1,\cdots,K\}$.

Constraints (24) are linear with c_i^m and p_i^m as variables. Clearly, the larger the K, the tighter the linear relaxation and the relaxation error can be made arbitrarily small (with an increasing number of linear constraints). In reality, we can choose K to let the relaxation errors be as small as desired. Since constraints (24) are linear constraints, a relatively larger K does not add much to computation complexity. Note that most of the parameters in (24) are constants except for h_i^m's and thus the cost of calculating (24) is negligible. After this linearization, the problem is a mixed-integer second-order conic program (MISOCP), which can be handled directly by commercial solvers. However, it may still take a lot of time due to non-linearity. Thus, we design a speed-up procedure by linearizing the second-order cones.

B. Second-Order Cones

Constraints (22) from ECR are second-order cones where the non-linear terms are in the form of $\mathbf{p}^T\mathbf{R}_j\mathbf{p}$. To linearize the non-linear terms $\mathbf{p}^T\mathbf{R}_j\mathbf{p}$, we rely on the L_∞-norm of a vector. This part is heuristic and introduces relaxation errors. Although we are not able to quantify the loss theoretically, we will show the loss numerically in Section VII. We first introduce a constant matrix $\mathbf{V}_j$ as the square root of $\mathbf{R}_j$, i.e., $\mathbf{R}_j=\mathbf{V}_j^T\mathbf{V}_j$. Since $\mathbf{R}_j$ is the covariance of $\mathbf{g}_j$, it is guaranteed to be positive semi-definite and symmetric. Thus, $\mathbf{V}_j$ always exists and can easily be calculated by Cholesky decomposition after obtaining the latest covariance $\mathbf{R}_j$.

With $\mathbf{V}_j$, we can rewrite constraints (22) as

$$\sqrt{\frac{1-\epsilon_j}{\epsilon_j}}||\mathbf{V}_j\mathbf{p}||_2 + \overline{\mathbf{g}}_j^T\mathbf{p} \le I_j \quad (j\in\mathcal{J}), \tag{25}$$

where $||\cdot||_2$ is the L_2-norm of a column vector. We propose to relax this L_2-norm based on the $L_\infty-$norm, which is given as

$$\sqrt{\frac{1-\epsilon_j}{\epsilon_j}}\sqrt{r_j}||\mathbf{V}_j\mathbf{p}||_\infty + \overline{\mathbf{g}}_j^T\mathbf{p} \le I_j \quad (j\in\mathcal{J}), \tag{26}$$

where $||\mathbf{V}_j\mathbf{p}||_\infty$ is the $L_\infty-$norm of $\mathbf{V}_j\mathbf{p}$ and r_j is a design parameter that satisfies $||\mathbf{V}_j\mathbf{p}||_2 \le \sqrt{r_j}||\mathbf{V}_j\mathbf{p}||_\infty$. Note that constraints (26) are linear with decision variables in $\mathbf{p}$ since a L_∞-norm can be easily converted to linear constraints with auxiliary variables. Further, $\mathbf{V}_j$ and r_j are only updated when the slow-changing $\mathbf{R}_j$ is updated and the computation of $\mathbf{V}_j$ and r_j can be done within a TTI since the dimension of $\mathbf{R}_j$ ($MN\times MN$) are usually less than 1,000 in a picocell scenario. Thus, the cost of computing (26) is negligible.

Clearly, we want to choose a small r_j to reduce relaxation errors. Thus, we set r_j as an upper bound on the number of non-zero elements in the column vector $\mathbf{V}_j\mathbf{p}$ for any feasible $\mathbf{p}$. Such an upper bound can be solved based on the sparse structures of $\mathbf{V}_j$ and $\mathbf{p}$ in our problem, as we describe below.

Recall $\mathbf{p}$ (i.e., $\left[p_1^1,\cdots,p_1^M,\cdots,p_N^1,\cdots,p_N^M\right]^T$) has at most M non-zero elements because each RB can only be allocated to at most one SU. Further, $\mathbf{V}_j$, as the square root of the covariance matrix $\mathbf{R}_j$, is a $MN\times MN$ matrix. Since the channel gains from different SUs to PU j are usually independent, $\mathbf{V}_j$ is a block diagonal matrix with the i-th block (denoted as V_{ij}) corresponding to SU i ($i=1,\cdots,N$), i.e.,

$$\mathbf{V}_j = \begin{bmatrix} \mathrm{V}_{1j} & & & \\ & \mathrm{V}_{2j} & & \\ & & \cdots & \\ & & & \mathrm{V}_{Nj} \end{bmatrix}. \tag{27}$$

Moreover, for SU i, the correlation between two RBs decreases as their carrier spacing increases. Define L_j as the maximum subcarrier spacing that has correlation, meaning that an RB is correlated with at most $2L_j$ neighboring RBs. That is, RB m is at most correlated with RB $M-L_j, M-L_j+1,\cdots,M-1,M+1,\cdots,M+L_j$. In reality, L_j can be determined upon solving $\mathbf{V}_j$. Clearly, $L_j=0$ and $L_j=M-1$ represent the scenarios of independent RBs and fully correlated RBs respectively. Then, each block V_{ij} is a band matrix [51] in the following form:

$$\mathrm{V}_{ij} = \begin{bmatrix} v_{ij}^{11} & \cdots & v_{ij}^{1(L_j+1)} & & & & \\ v_{ij}^{21} & v_{ij}^{22} & \cdots & v_{ij}^{2(L_j+2)} & & & \\ & \cdots & \cdots & \cdots & \cdots & \cdots & \\ & & & & v_{ij}^{M(M-L_j)} & \cdots & v_{ij}^{MM} \end{bmatrix}$$

Based on the sparse properties of $\mathbf{V}_j$ and $\mathbf{p}$, we consider the following two cases to calculate r_j.

Case 1. When $M\ge 2L_j+1$, denote r_j^m as the number of non-zero elements in $\mathbf{V}_j\mathbf{p}$ if p_i^m is the only positive element of $\mathbf{p}$. r_j^m is calculated (based on m) by

$$r_j^m = \begin{cases} m+L_j & \text{if } m\le L_j, \\ (2L_j+1) & \text{if } L_j+1\le m\le M-L_j, \\ M-m+L_j+1 & \text{if } m\ge M-L_j+1. \end{cases} \tag{28}$$

The three cases in (28) corresponds to the RBs on the lower sub-channels ($1\sim L_j$), middle sub-channels ($L_j+1\sim M-L_j$) and upper sub-channels ($M-L_j+1\sim M$) of the spectrum allocated to the pico cell.

Since only one SU can transmit on RB m, i.e., only one positive p_i^m in set $\{P_1^m,P_2^m,\cdots,P_N^m\}$, we can set r_j to

$$r_j = \sum_{m\in\mathcal{M}} r_j^m = M(2L_j+1) - L_j(L_j+1). \tag{29}$$

Case 2. When $M<2L_j+1$, r_j can be set to $r_j=\min\{MN, M(2L_j+1)-L_j(L_j+1)\}$ since MN is always an upper bound of r_j.

Combining both cases, we set r_j to

$$r_j = \min\{MN, M(2L_j+1)-L_j(L_j+1)\} \quad (j\in\mathcal{J}). \tag{30}$$

By replacing logarithmic terms in the objective function and (22) with constraints (24) and (26) respectively, we have the

following relaxed optimization problem:

$$\text{(P3)} \quad \max_{x_i^m, p_i^m} \sum_{i\in\mathcal{N}} \sum_{m\in\mathcal{M}} w_i c_i^m$$
$$\text{s.t.} \quad \text{Scheduling constraints } (1),$$
$$\text{Transmission power constraints } (2), (3),$$
$$\text{Linearized interference power constraints } (26),$$
$$\text{Linearized throughput calculations } (24),$$
$$x_i^m \in \{0,1\}, \quad p_i^m \geq 0.$$

P3 belongs to mixed-integer linear program (MILP), which can also be solved directly by commercial solvers with a smaller amount of time than P2. Note that the computation cost to derive P3 is negligible since the parameters in (24) and (26) are all constants. The actual running time of P3 depends on the number of binary variables. Luckily, the number of binary variables in P3 is MN, which is usually small in underlay settings. To further reduce the computation time of MILP (e.g., to millisecond time scale), one can employ a recent work in real-time optimization by using GPU platforms [52] to solve a large MILP. But such an effort is beyond the scope of this article.

VII. Simulation Results

In this section, we conduct simulations to evaluate our proposed solutions predicated on ECR. We focus on the achievable spectrum efficiency (our objective value) and the actual threshold violation probability. We first show performance under independent channels where Bernstein Approximation is applicable. Then we evaluate our proposed solutions in more general settings (e.g., other fading channels, correlated channels, and large-scale networks). Our numerical study also considers different risk levels, interference thresholds, and the number of RBs.

A. Settings

For all topologies in the simulation study, we set the distance between the macro BS and the pico BS to 400 meters. The radius of a picocell is set to 40 meters, which is on the same scale from 3GPP standard [53]. The maximum transmission power $P_i^{\max}$ of each SU across all RBs is set to 20 dBm. The weight w_i of each SU, the interference threshold I_j of each PU, and the risk level ϵ_j of each PU will be given when they are used in different settings.

The wireless channel gains (h_i^m's for transmission and g_{ij}^m's for interference) are calculated based on the ITU path loss model and fast fading. We consider two types of path loss models as follows [53]:

(i) The path loss from the macro BS follows the ITU outdoor path loss model, i.e., $PL_{\text{BS}}(\text{dB}) = 128.1 + 37.6 \times \log_{10}(d_{\text{BS}})$, where d_{BS} is the distance from the macro BS to the pico BS (in kilometers).

(ii) The path loss from SU i to the pico BS and PU j follows the ITU indoor path loss model, i.e., $PL_i(\text{dB}) = 38 + 30 \times \log_{10}(d_i)$, $PL_{ij}(\text{dB}) = 38 + 30 \times \log_{10}(d_{ij})$, where d_i and d_{ij} are the distances from SU i to the pico BS and PU j (in meters), respectively.

Denote f_i^m and f_{ij}^m as the fast fading (in power domain) from SU i to the pico BS and PU j on RB m, respectively. In this work, we consider the most common fast fadings such as Rayleigh, Rician, and Nakagami fading [54]. Denote P_{BS} and σ^2 as the transmission power from the macro BS and the thermal noise, respectively. Based on these notations, we calculate h_i^m and g_{ij}^m as follows:

$$h_i^m = \frac{PL_i f_i^m}{PL_{BS} P_{BS} + \sigma^2}, \quad g_i^m = PL_{ij} f_{ij}^m .$$

In our simulation study, we set P_{BS} =46 dBm and $\sigma^2 = 1 \times 10^{-7}$ mW. Note that we do not assume any knowledge of distributions in our system model and proposed solutions. But in simulation, we have to generate the random parameters based the above distributions. This is only for numerical study in simulations and does not infer such knowledge is available in our proposed solutions. In other words, the proposed solutions are "blindfolded" when we are generating these distributions in the simulation study.

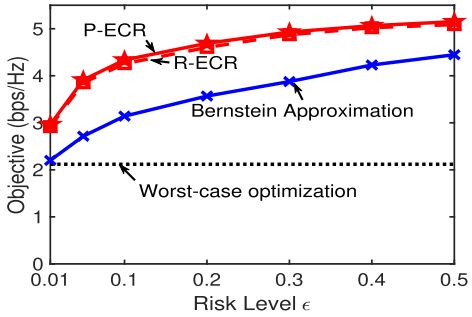
B. Independent Channels With Rayleigh Fading

In this section, the channels are generated independently with Rayleigh fading as fast fading. This scenario is widely considered in previous works and we will compare our proposed solutions with Bernstein Approximation. We consider two scenarios for network topology: (i) the SUs are randomly distributed in the picocell; (ii) the SUs are in close proximity to the PU (a stressful scenario in terms of interference).

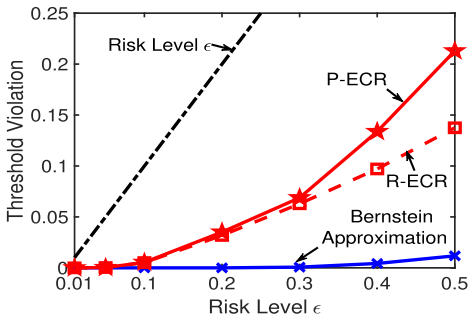
We tested four algorithms where the linearization of logarithmic terms (Section VI.A) is always used since it only introduces less than 1% performance loss (with $K = 50$). Two proposed algorithms are called "P-ECR" and "R-ECR" representing the one after linearization of logarithmic terms (i.e., MISOCP) and P3 respectively. For comparison, we include results from Bernstein Approximation and worst-case optimization in the same figure whenever applicable. Worse-case optimization uses the upper bounds of interference channel gains to remove uncertainty and consequently the chance constraints (4) become deterministic linear constraints. As for Bernstein Approximation, since our channel model is unbounded, we employ the same truncation method proposed in [20]. Then the reformulated MINLP is solved by CPLEX without any further relaxation. For a fair comparison, Bernstein Approximation is also "blindfolded" when we are generating these distributions in the simulation study, same with our proposed solutions.

For each topology, we perform 200 simulation runs and the results are the average values for P1 based on the obtained feasible solutions. For each simulation, we generate 10,000 samples of the channel gain from each SU to each PU. The first and second-order statistics from the 10,000 samples are used in our solutions. These 10,000 samples are also used to calculate the actual threshold violation probability after the solutions (including scheduling decisions and transmission powers) in each simulation are obtained. All optimization problems are solved on 16-core Intel Xeon E5-2687w. We use CPLEX version 12.8.0 for a mixed-integer solution with a relative gap (between integer solution and continuous solution) of less than 1%.

For the cases tested in this subsection and similar scales later ($N = 6$, $M = 12$, and $J = 1, 3$), the running times of "Bernstein Approximation" and P-ECR are on the order of seconds while R-ECR and "worst-case optimization" runs on the order of sub-seconds. Specifically, we found that R-ECR runs $3\sim 6$ times faster than R-ECR. Thus, we will omit the running times and focus on performance except for the large scale network considered in Section VII.E.



(a) Objective value



(b) Actual threshold violation probability

Fig. 2. Performance as a function of risk level ϵ with one PU.

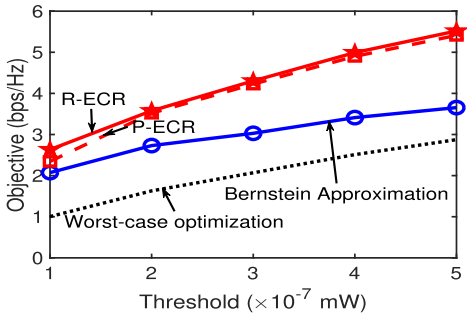


Fig. 3. Objective as a function of interference threshold I with one PU.

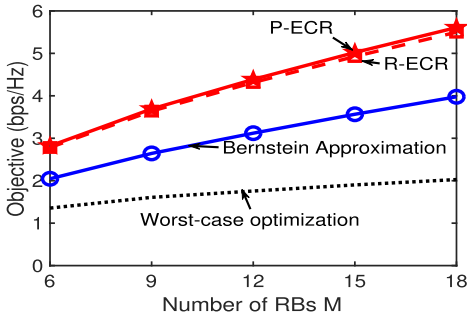


Fig. 4. Objective as a function of number of RBs M with one PU.

1) Randomly Distributed SUs: In this subsection, we consider the settings with six SUs ($N = 6$) and one or three PUs ($J = 1$ or 3). Assuming the pico BS is at the origin and the coordinates of SUs are generated based on uniform distribution. For this numerical study, the coordinates (all in meters) of SUs are $(-10.67, -19.01)$, $(32.47, -11.77)$, $(-0.23, 8.90)$, $(12.04, -5.03)$, $(28.89, -10.00)$ and $(-22.66, 5.34)$. The corresponding weights of SUs are $0.22, 0.09, 0.09, 0.23, 0.19$, and 0.18. The spectrum allocated to the picocell is divided into 12 RBs ($M = 12$) and remains unchanged in the simulation study. The coordinates and interference threshold I_j for each PU are given per discussion below.

(i) One PU. Consider the case with one PU located at $(0, 50)$ and the interference threshold I is set to 3×10^{-7} mW. The achievable spectrum efficiency and actual threshold violation probability are shown in Fig. 2.

In Fig. 2(a), we find that the objective value of our solutions monotonically increases with the risk level. The spectrum efficiency improvement is due to a larger optimization space since the PU can tolerate more interference threshold violations. Further, the objective values of P-ECR and R-ECR are very close. In particular, P-ECR and R-ECR achieve 2.97 bps/Hz and 2.92 bps/Hz with a risk level of $\epsilon = 0.01$; 5.15 bps/Hz and 5.089 bps/Hz with $\epsilon = 0.5$.

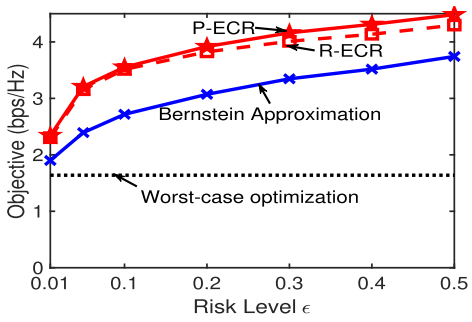
In Fig. 2(a), the objective value of worst-case optimization stays the same since it does not involve any risk. Clearly, the performance of worst-case optimization is overly pessimistic (due to zero tolerance of interference threshold violation). Further, as shown in Fig. 2(a), with a small risk level $\epsilon = 0.01$ (or 1%), the performance of Bernstein Approximation drops to 2.20 bps/Hz (almost the same with the one from worst-case optimization) while P-ECR and R-ECR achieve 2.97 bps/Hz and 2.92 bps/Hz respectively. In terms of relative improvement ratio, our solutions can achieve 15% to 44% improvement over that by Bernstein Approximation. The 44% improvement is achieved when $\epsilon = 0.05$ by P-ECR (a common risk level that one may use in practice).

Fig. 2(b) shows the actual threshold violation probability from four algorithms. This actual threshold violation probability is obtained by calculating the percentage of observed instances where the interference threshold is violated. Interestingly, P-ECR has a higher violation probability than R-ECR but their objective values are rather close. This is due to the concavity of the log functions in our objective function. As shown in Fig. 2(b), the actual threshold violation probability from Bernstein Approximation stays below 0.02 even though the risk level is 0.5, which is unnecessarily conservative and thus loses substantial benefits of CCP. On the other hand, our solutions violate the interference threshold with a probability lower than the risk level but much larger than that from Bernstein Approximation. Thus, our solutions can better reap the benefits of CCP while still keeping the threshold violation probability below the risk level. The gap between the actual threshold violation probability and risk level ϵ (the supremum of threshold violation probability) is attributed to our channel model (ITU path loss and Rayleigh fading), which is not the worst-case distribution (since it does not satisfies the properties in Property 1). As discussed in Section V, the threshold violation probability does not achieve the supremum (the risk level). Consequently, a gap exists between the actual threshold violation probability and the risk level ϵ.

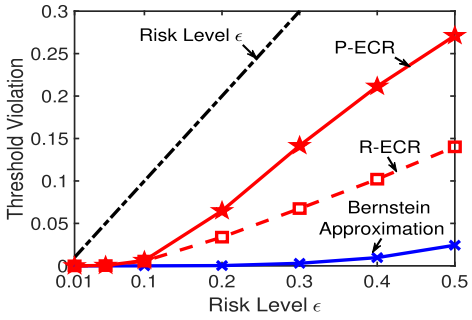
The results of different interference threshold I and number of RBs M are shown in Fig. 3 and Fig. 4 respectively. As expected, the achievable objective values (spectrum efficiency) under all three solutions increase with higher interference threshold I or the number of RBs M. The objective values of P-ECR and R-ECR are very close, meaning the relaxation errors from linearization is small. Our solutions provide a higher spectrum efficiency with relative 13%$\sim$50% improvements compared to Bernstein Approximation.

(ii) Multiple PUs In this study, we consider three PUs ($J = 3$) that are located at $(50, 0)$, $(-45, 5)$ and $(-25, -35)$, respectively.

Our proposed solutions are able to handle the general case when each PU has a different interference threshold I_j and

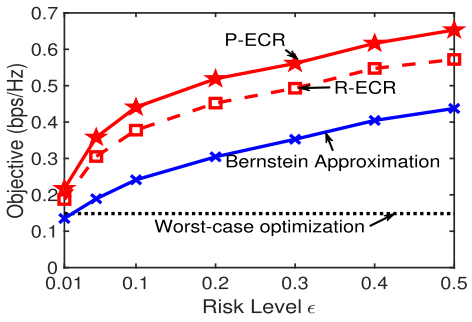


(a) Objective value

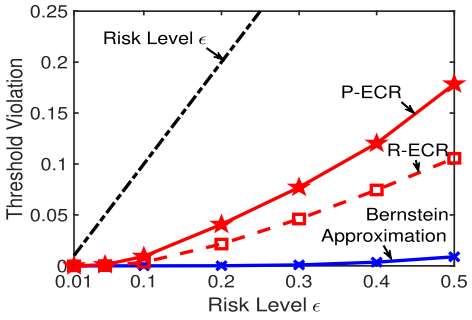


(b) Actual threshold violation probability

Fig. 5. Performance as a function of risk level ϵ with three PUs.



(a) Objective value



(b) Actual threshold violation probability

Fig. 6. Performance as a function of risk level ϵ when SUs are close to the PU.

risk level ϵ_j. Without loss of generality, we use the same risk level for the PUs but different interference threshold I_j. The interference thresholds for the three PUs are 3×10^{-7} mW, 4×10^{-7} mW, and 2×10^{-7} mW respectively. The results are summarized in Fig. 5.

As shown in Fig. 5(a), our solutions outperforms Bernstein Approximation with up to 37% improvement (by P-ECR when $\epsilon = 0.05$). R-ECR only loses 1∼4% spectrum efficiency when compared to P-ECR. We also note that the objective values are less than that of the single PU scenario in Fig. 2(a). This is intuitive as more interference constraints are imposed on the SUs due to multiple PUs. The actual threshold violation probability is given in Fig. 5(b). We use the maximum threshold violation probability among the three PUs in Fig. 5(b). It shows that the violation probabilities from our proposed solutions is smaller than the risk level but higher than that from Bernstein Approximation, which is similar to that in Fig. 2(b).

2) SUs in Close Proximity to the PU: In this subsection, we consider a more stressful network topology with one PU where all the SUs in the picocell are close to the PU. The PU is located at $(50, 0)$ with an interference threshold $I = 3 \times 10^{-7}$ mW. The coordinates of six SUs are $(20.79, -21.49)$, $(17.86, -24.49)$, $(32.16, 14.79)$, $(30.65, 5.08)$, $(24.90, -15.41)$, and $(34.21, -5.52)$, respectively. Their corresponding weights are 0.19, 0.13, 0.26, 0.12, 0.20, and 0.10, respectively. Under this circumstance, the channel gains from the SUs to the PU are larger, and more stringent power control on the SUs should be exercised.

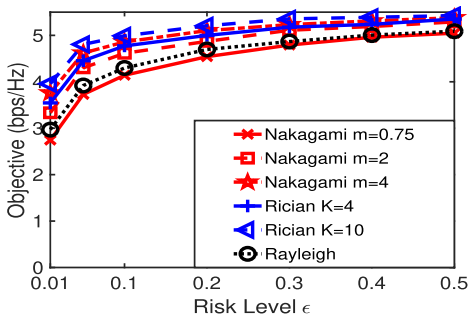
Fig. 6(a) shows the achievable spectrum efficiency under this stressful scenario. Since the SUs are closer to the PU, they must further lower their transmission power to control their interference to the PUs. Consequently, the spectrum efficiency will be lower than that when SUs are randomly distributed. P-ECR has the highest objective value followed by R-ECR. Specifically, R-ECR loses $11\% \sim 14\%$ spectrum efficiency when compared to P-ECR. Thus, when the interference channel gains become larger (due to closer proximity), the relaxation error from the linearization will increase. Further, Bernstein Approximation is inferior to our proposed ECR-based solutions (P-ECR and R-ECR).

Fig. 6(b) shows the actual threshold violation probability. We see the threshold violation probabilities from our solutions are lower than the risk level but higher than that from Bernstein Approximation.

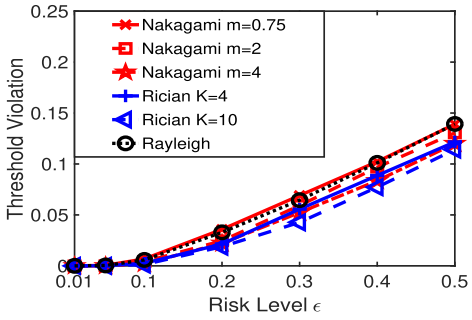
C. Independent Channels With Rician and Nakagami Fading

In this section, we evaluate our proposed solutions under different channel models. We only show the results of R-ECR but the same conclusions also hold for P-ECR. We want to show our proposed solution can guarantee the risk level with only a knowledge of mean and covariance (without distribution). Note that g_{ij}^m's are always independent regardless of their distributions. The topology we use is the same as in Section VII-B.1. We consider the three most popular fast fading models: Rayleigh, Rician, and Nakagami fading. Since these channel models are unbounded and no truncation method has been designed for Bernstein Approximation, we will only present the results from R-ECR.

Fig. 7 shows the results of our simulations. To better understand the results, we first discuss the relationship among these wireless channel models. Note that our solution only require the mean and covariance of the RVs regardless of the underlying distributions. Nakagami fading is a general model for the wireless channel that includes both Rayleigh and Rician fading. It uses a parameter called "Nakagami-m" to represent the seriousness of fading, where a small "Nakagami-m" means a faster (more serious) fading situation [54]. For Rayleigh fading, it is equivalent to Nakagami fading with Nakagami-$m = 1$. On the other hand, Rician fading uses a parameter

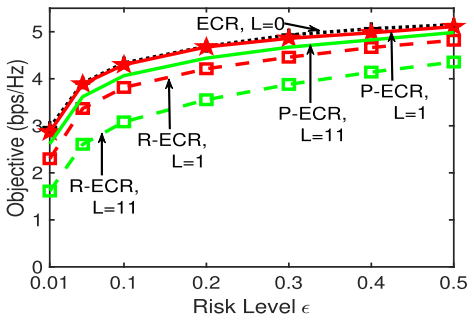


(a) Objective value

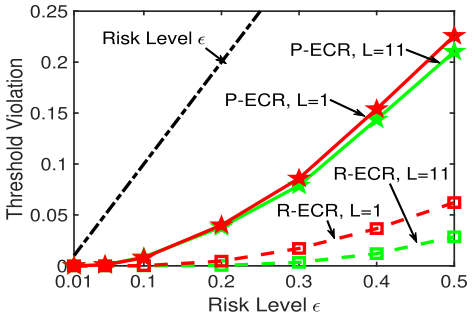


(b) Actual threshold violation probability

Fig. 7. Performance of R-ECR as a function of risk level ϵ under different channel models.



(a) Objective value



(b) Actual threshold violation probability

Fig. 8. Performance as a function of risk level ϵ in correlated channels.

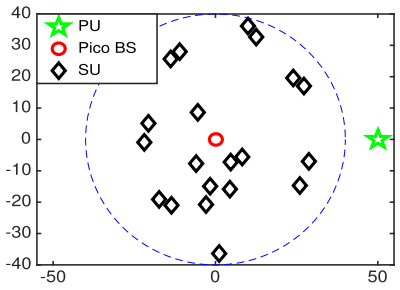


Fig. 9. Network topology of a large-scale network with 20 SUs and 1 PU.

called "Rician-K" which has a mathematical relationship with "Nakagami-m" through $m = (K+1)^2/(2K+1)$.

We employ six different settings of fast fading: Nakagami-$m = 0.75, 2, 4$, Rician $K = 4, 10$ and Rayleigh fading, as shown in Fig. 7. These six channel models have the following equivalent Nakagami-m parameters: 0.75, 2, 4, 2.78, 5.76, 1, respectively. Fig. 7 shows that a smaller Nakagami-m leads to a lower spectrum efficiency. Given the same mean, a small Nakagami-m generates a higher covariance (more serious fading) and consequently, we have a smaller optimization space that leads to worse performance.

Fig 7(b) shows that our proposed solution can successfully maintain the risk level (with the actual threshold violation probability smaller than the risk level). The gap between risk level ϵ and the actual threshold violation probability still exists because none of these channel models belongs to the worst-case distributions.

D. Correlated Channels With Rayleigh Fading

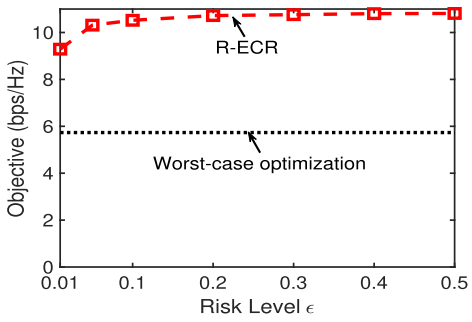
In this section, we consider the general case where RBs have correlations. We want to validate that our proposed solutions are applicable to correlated channels. Since Bernstein Approximation explicitly requires independent channels and is no longer applicable to correlated channels, we will only show the results from our proposed solutions P-ECR and R-ECR. Specifically, we use the same topology in Section VII-B.1 but introduce different levels of correlations among RBs. The channels are based on the path-loss model and Rayleigh fading.

We consider two levels of correlations in this subsection of numerical study. For low correlation case, only the adjacent RBs are correlated with a relative correlation of 0.5 (i.e., $L = 1$, $r = 34$). For high correlation case, the correlation coefficient between any two RBs (adjacent or not) is set to 0.5 (i.e., $L = 11$, $r = 72$). L is maximum sub-carrier spacing of correlation and r_j is a design parameter defined in (29). We set the interference threshold $I = 3 \times 10^{-7}$ mW.

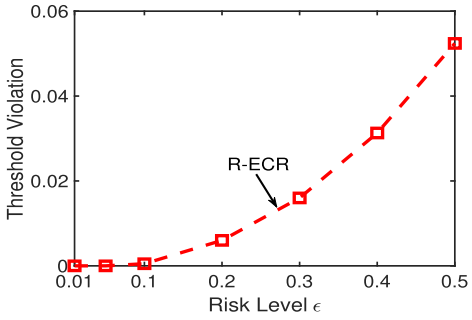
Fig. 8 shows the results. In Fig. 8(a), we find that under the same risk level, as correlation increases, the achievable spectrum efficiency decreases for both P-ECR and R-ECR due to a smaller optimization space. This can be seen from constraints (22) that $\mathbf{p}^T\mathbf{R}_j\mathbf{p}$ is higher for a given $\mathbf{p}$. Further, in correlated channel, R-ECR loses 2%$\sim$30% spectrum efficiency compared to P-ECR. This is because the relaxation errors from L_2-norm to $L_\infty-$norm is larger as correlation increases. The actual threshold violation probability is shown in Fig. 8(b). We see that the actual threshold violation probabilities are all smaller than the risk level ϵ_j (our required performance guarantee) but the gaps from R-ECR are larger than from P-ECR due to the relaxation errors from norm linearization.

E. Large-Scale Network

To check the scalability of our proposed solutions, we consider a large-scale network with 20 SUs, 50 RBs and 1 PU (i.e., $N = 20$, $M = 50$ and $J = 1$). The topology we use is shown in Fig. 9 and the channels are based on Rayleigh fading (independent). We found that R-ECR runs on the order of tens of seconds while the solver cannot close the gap even



(a) Objective value



(b) Actual threshold violation probability

Fig. 10. Performance as a function of risk level ϵ in a large-scale network.

after several days in Bernstein Approximation and P-ECR. Thus, we will only show the results of R-ECR (in Fig. 10).

As shown in Fig. 10(a), the achievable spectrum efficiency is larger under higher risk level ϵ, which is consistent with the one from the small-scale network. Further, the objective value is 10.5 bps/Hz when $\epsilon = 0.05$, which is much higher than that of small-scale networks. This is because we assume the bandwidth of each RB is normalized to 1 and thus the bandwidth allocated to the picocell is proportional to the number of RBs (M). In other words, the picocell in this large-scale setting is actually using a wider spectrum. Moreover, with a larger number of SUs, it is more likely that some of them are not close to the PUs. As a result, the interference channel gains from these SUs to the PUs are relatively smaller and there are more opportunities for these SUs to transmit without harmful interference to the PUs.

However, the improvement of spectrum efficiency is not proportional to the total bandwidth. As we increase the total bandwidth by more than four times (from 12 to 50), the achievable spectrum efficiency only increases 2$\sim$3 times. The main reason is that we use the threshold for aggregate interference across all RBs to protect the PUs. With more RBs, the average allowable interference on each RB is decreasing. Thus, the ratio of objective improvement is less than the ratio of spectrum bandwidth.

Fig. 10(b) shows that the actual threshold violation probability from our solution is indeed lower than the risk level. We see that under the same risk levels, the actual threshold violation probability becomes smaller comparing to that in small-scale networks. This is because that the relaxation errors from linearization is higher in large-scale networks.

VIII. Conclusions

In this article, we studied underlay coexistence with channel uncertainty where the knowledge of interference channel gains is limited to mean and covariance. We used the notation of risk level to exploit SUs' occasional interference threshold violation to the PUs and formulated a chance-constrained program (CCP). To address the intractable CCP formulation, we introduced a powerful reformulation tool called Exact Conic Reformulation (ECR) that offers an exact reformulation from intractable chance constraints into convex second-order cones. ECR overcomes the limitations of the state-of-the-art Bernstein Approximation (i.e., certain assumptions for random variables and conservative performance). Through extensive performance evaluation, we show that: (i) for independent channels, our proposed approaches outperform Bernstein Approximation by 33% and 30% on average in terms of spectrum efficiency; (ii) for correlated channels, our proposed approaches offer competitive solutions while Bernstein Approximation is no longer applicable. We believe that ECR is not limited to the problem studied in this article. It represents a novel and powerful technique to fully reap the benefits of CCP when addressing uncertainty in wireless networking problems.

Acknowledgments

The authors thank the anonymous reviewers for their comments and suggestions for improving the quality of this article. Specifically, the authors thank one reviewer for pointing out Cantelli's Inequality and a prior result in reference [48].

References

[1] S. Li, Y. Huang, C. Li, B. A. Jalaian, Y. T. Hou, and W. Lou, "Coping uncertainty in coexistence via exploitation of interference threshold violation," in *Proc. ACM MobiHoc*, Catania, Italy, Jul. 2019, pp. 71–80.

[2] 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.

[3] L. Le and E. Hossain, "Resource allocation for spectrum underlay in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5306–5315, Dec. 2008.

[4] S.-J. Kim and G. B. Giannakis, "Optimal resource allocation for MIMO ad hoc cognitive radio networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3117–3131, May 2011.

[5] K. Doppler, M. Rinne, C. Wijting, C. Ribeiro, and K. Hugl, "Device-to-device communication as an underlay to LTE-advanced networks," *IEEE Commun. Mag.*, vol. 47, no. 12, pp. 42–49, Dec. 2009.

[6] *Scenarios and Requirements for Small Cell Enhancements for E-UTRA and E-UTRAN*, document TR 36.932, Version 15.0.0, 3GPP, Jul. 2018. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2590

[7] D. J. Costello and G. D. Forney, "Channel coding: The road to channel capacity," *Proc. IEEE*, vol. 95, no. 6, pp. 1150–1177, Jun. 2007.

[8] T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proc. IEEE*, vol. 88, no. 4, pp. 451–515, Apr. 2000.

[9] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: A review," *Proc. IEEE*, vol. 86, no. 5, pp. 974–997, May 1998.

[10] J. R. Birge and F. Louveaux, *Introduction to Stochastic Programming*. Springer, 2011, ch. 2.

[11] E. Dall'Anese, S.-J. Kim, G. B. Giannakis, and S. Pupolin, "Power control for cognitive radio networks under channel uncertainty," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3541–3551, Oct. 2011.

[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.

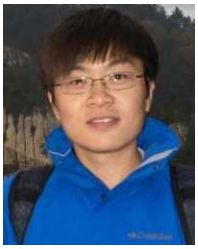
[13] A. Ben-Tal and A. Nemirovski, "Robust solutions of uncertain linear programs," *Oper. Res. Lett.*, vol. 25, no. 1, pp. 1–13, Aug. 1999.

[14] S. Parsaeefard and A. R. Sharafat, "Robust worst-case interference control in underlay cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 61, no. 8, pp. 3731–3745, Oct. 2012.

[15] A. Charnes and W. W. Cooper, "Chance-constrained programming," *Manage. Sci.*, vol. 6, no. 1, pp. 73–79, Oct. 1959.

[16] Y. Rong, S. A. Vorobyov, and A. B. Gershman, "Robust linear receivers for multiaccess space-time block-coded MIMO systems: A probabilistically constrained approach," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1560–1570, Aug. 2006.
[17] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. 2. Hoboken, NJ, USA: Wiley, 2008, ch. 1.
[18] R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," *J. Risk*, vol. 2, pp. 21–42, Feb. 2000.
[19] A. Ben-Tal and A. Nemirovski, "Selected topics in robust convex optimization," *Math. Program.*, vol. 112, no. 1, pp. 125–158, Jul. 2007.
[20] N. Y. Soltani, S.-J. Kim, and G. B. Giannakis, "Chance-constrained optimization of OFDMA cognitive radio uplinks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 3, pp. 1098–1107, Mar. 2013.
[21] W. W.-L. Li, Y. Jun Zhang, A. M.-C. So, and M. Z. Win, "Slow adaptive OFDMA systems through chance constrained programming," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3858–3869, Jul. 2010.
[22] S. Wang, W. Shi, and C. Wang, "Energy-efficient resource management in OFDM-based cognitive radio networks under channel uncertainty," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3092–3102, Sep. 2015.
[23] T. A. Le, Q.-T. Vien, H. X. Nguyen, D. W. K. Ng, and R. Schober, "Robust chance-constrained optimization for power-efficient and secure SWIPT systems," *IEEE Trans. Green Commun. Netw.*, vol. 1, no. 3, pp. 333–346, Sep. 2017.
[24] M. El Tanab and W. Hamouda, "Resource allocation for underlay cognitive radio networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 1249–1276, 2nd Quart., 2017.
[25] S. K. Sharma, T. E. Bogale, S. Chatzinotas, B. Ottersten, L. B. Le, and X. Wang, "Cognitive radio techniques under practical imperfections: A survey," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 1858–1884, Jul. 2015.
[26] M. Peng, C. Wang, J. Li, H. Xiang, and V. Lau, "Recent advances in underlay heterogeneous networks: Interference control, resource allocation, and self-organization," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 700–729, 2nd Quart., 2015.
[27] D. L. Wasden, H. Moradi, and B. Farhang-Boroujeny, "Design and implementation of an underlay control channel for cognitive radios," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 10, pp. 1875–1889, Nov. 2012.
[28] D. Feng, L. Lu, Y. Yuan-Wu, G. Y. Li, G. Feng, and S. Li, "Device-to-device communications underlaying cellular networks," *IEEE Trans. Commun.*, vol. 61, no. 8, pp. 3541–3551, Aug. 2013.
[29] F. Wang, C. Xu, L. Song, and Z. Han, "Energy-efficient resource allocation for device-to-device underlay communication," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2082–2092, Apr. 2015.
[30] C. Pan, J. Wang, W. Zhang, B. Du, and M. Chen, "Power minimization in multi-band multi-antenna cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 5056–5069, Sep. 2014.
[31] Y. Cao *et al.*, "Optimization or alignment: Secure primary transmission assisted by secondary networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 4, pp. 905–917, Apr. 2018.
[32] B. Li, Z. Fei, Z. Chu, F. Zhou, K.-K. Wong, and P. Xiao, "Robust chance-constrained secure transmission for cognitive satellite–terrestrial networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4208–4219, May 2018.
[33] M. H. Al-Ali and K. C. Ho, "Transmit precoding in underlay MIMO cognitive radio with unavailable or imperfect knowledge of primary interference channel," *IEEE Trans. Wireless Commun.*, vol. 15, no. 8, pp. 5143–5155, Aug. 2016.
[34] S. Gong, P. Wang, and L. Duan, "Distributed power control with robust protection for PUs in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3247–3258, Jun. 2015.
[35] S.-J. Kim, N. Y. Soltani, and G. B. Giannakis, "Resource allocation for OFDMA cognitive radios under channel uncertainty," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3578–3587, Jul. 2013.
[36] M. Hasan, E. Hossain, and D. I. Kim, "Resource allocation under channel uncertainties for relay-aided device-to-device communication underlaying LTE-A cellular networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 2322–2338, Apr. 2014.
[37] A. Damnjanovic *et al.*, "A survey on 3GPP heterogeneous networks," *IEEE Wireless Commun.*, vol. 18, no. 3, pp. 10–21, Jun. 2011.
[38] V. Chandrasekhar, J. Andrews, and A. Gatherer, "Femtocell networks: A survey," *IEEE Commun. Mag.*, vol. 46, no. 9, pp. 59–67, Sep. 2008.
[39] F. Jin, R. Zhang, and L. Hanzo, "Fractional frequency reuse aided twin-layer femtocell networks: Analysis, design and optimization," *IEEE Trans. Commun.*, vol. 61, no. 5, pp. 2074–2085, May 2013.
[40] H.-B. Chang and I. Rubin, "Optimal downlink and uplink fractional frequency reuse in cellular wireless networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 2295–2308, Apr. 2016.
[41] A. R. Elsherif, W.-P. Chen, A. Ito, and Z. Ding, "Adaptive resource allocation for interference management in small cell networks," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2107–2125, Jun. 2015.
[42] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Commun. Surveys Tuts.*, vol. 11, no. 1, pp. 116–130, 1st Quart., 2009.
[43] A. Furtado, L. Irio, R. Oliveira, L. Bernardo, and R. Dinis, "Spectrum sensing performance in cognitive radio networks with multiple primary users," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1564–1574, Mar. 2016.
[44] L. Wei and O. Tirkkonen, "Spectrum sensing in the presence of multiple primary users," *IEEE Trans. Commun.*, vol. 60, no. 5, pp. 1268–1277, May 2012.
[45] S.-B. Lee, I. Pefkianakis, A. Meyerson, S. Xu, and S. Lu, "Proportional fair frequency-domain packet scheduling for 3GPP LTE uplink," in *Proc. 28th Conf. Comput. Commun. (INFOCOM)*, Rio de Janeiro, Brazil, Apr. 2009, pp. 2611–2615.
[46] G. C. Calafiore and L. E. Ghaoui, "On distributionally robust chance-constrained linear programs," *J. Optim. Theory Appl.*, vol. 130, no. 1, pp. 1–22, Dec. 2006.
[47] B. K. Ghosh, "Probability inequalities related to Markov's theorem," *Amer. Statistician*, vol. 56, no. 3, pp. 186–190, Aug. 2002.
[48] J. Pinter, "Deterministic approximations of probability inequalities," *Zeitschrift für Oper.-Res.*, vol. 33, no. 4, pp. 219–239, 1989.
[49] S. M. Almalfouh and G. L. Stüber, "Interference-aware radio resource allocation in OFDMA-based cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 4, pp. 1699–1713, May 2011.
[50] Y. T. Hou, Y. Shi, and H. D. Sherali, *Applied Optimization Methods for Wireless Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2014, ch. 5.
[51] A. Gorokhov and J.-P. Linnartz, "Robust OFDM receivers for dispersive time-varying channels: Equalization and channel acquisition," *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 572–583, Apr. 2004.
[52] Y. Huang, S. Li, Y. T. Hou, and W. Lou, "GPF: A GPU-based design to achieve $\sim$100 μs scheduling for 5G NR," in *Proc. ACM MobiCom*, New Delhi, India, Oct./Nov. 2018, pp. 207–222.
[53] *Radio Frequency (RF) Requirements for LTE Pico Node B*, document TR 36.931, Version 15.0.0, 3GPP, Jun. 2018. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2589
[54] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005, ch. 3.

Shaoran Li (Student Member, IEEE) received the B.S. degree from Southeast University, Nanjing, China, in 2014, and the M.S. degree from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2017. He is currently pursuing the Ph.D. degree with Virginia Tech, Blacksburg, VA, USA. His research interest includes algorithm design and implementation for wireless networks.

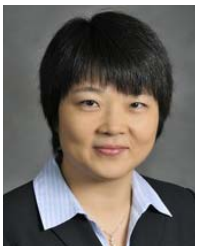


Yan Huang (Student Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2012 and 2015, respectively. He is currently pursuing the Ph.D. degree with Virginia Tech, Blacksburg, VA, USA. His research interests include GPU-based real-time optimizations for wireless networks and machine learning for communications.

Chengzhang Li (Student Member, IEEE) received the B.S. degree in electronics engineering from Tsinghua University, Beijing, China, in 2017. He is currently pursuing the Ph.D. degree with The Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, USA. His current research interests include modeling, analysis, and algorithm design for wireless networks, with a focus on age of information (AoI) and latency research.

Brian A. Jalaian (Member, IEEE) received the M.S. degree in electrical engineering (network and communication systems) and the M.S. degree in industrial system engineering (operation research) from Virginia Tech in 2013 and 2014, respectively, and the Ph.D. degree from The Bradley Department of Electrical and Computer Engineering, Virginia Tech, in 2016. He is currently a Research Scientist and the Research Lead at ARL and an Adjunct Research Assistant Professor with Virginia Tech. His research interests include optimization, machine learning, and network science.

Y. Thomas Hou (Fellow, IEEE) is currently the Bradley Distinguished Professor of electrical and computer engineering with Virginia Tech, Blacksburg, VA, USA, which he joined in 2002. He has published more than 300 articles in IEEE/ACM journals and conferences. He holds five U.S. patents. He has authored/coauthored two graduate textbooks: *Applied Optimization Methods for Wireless Networks* (Cambridge University Press, 2014) and *Cognitive Radio Communications and Networks: Principles and Practice* (Academic Press/Elsevier, 2009). His current research focuses on developing innovative solutions to complex science and engineering problems arising from wireless and mobile networks. His research interest includes wireless security. He was named as an IEEE Fellow for contributions to modeling and optimization of wireless networks. His articles were recognized by eight best paper awards from IEEE and ACM. He served as the Steering Committee Chair for IEEE INFOCOM conference and was a member of the IEEE Communications Society Board of Governors. He was also a Distinguished Lecturer of the IEEE Communications Society. He was/is on the editorial boards of a number of IEEE and *ACM Transactions* and journals.



Wenjing Lou (Fellow, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Florida. She is currently the W. C. English Endowed Professor of computer science with Virginia Tech. Her research interest includes cybersecurity field. Her current research interests include wireless networks, privacy protection in machine learning systems, and security and privacy problems in the Internet-of-Things (IoT) systems. She is a highly cited researcher by the Web of Science Group. She received the Virginia Tech Alumni Award for Research Excellence in 2018, which is the highest university level faculty research award. She received the INFOCOM Test-of-Time Paper Award in 2020. She is also the TPC Chair of IEEE INFOCOM 2019 and ACM WiSec 2020. She was the Steering Committee Chair of IEEE CNS conference from 2013 to 2020. She is also a Steering Committee Member of IEEE INFOCOM and IEEE TRANSACTIONS ON MOBILE COMPUTING. She served as the Program Director for the U.S. National Science Foundation (NSF) from 2014 to 2017.

Stephen Russell received the B.Sc. degree in computer science and the M.S. and Ph.D. degrees in information systems from University of Maryland. He is currently the Information Sciences Division Chief at the Army Research Laboratory. In addition to his research work, he has taught at the Department of Information Systems and Technology Management, George Washington University, and instructed courses under the Federal CIO Certificate Program, management, and information systems subject areas. He has published more than 100 articles in his primary research areas of decision support systems, machine learning, systems architectures, and intelligent systems. He has several years of information technology experience in the telecommunications, healthcare, and manufacturing industries. He has also been a serial entrepreneur, owning companies that have specialized in software engineering, information resource management services, and telecommunications equipment manufacturing. His current research interest includes the Internet of Things and its applicability to the U.S. Army and the U.S. Military. He leads the ARL research program on Internet of Battlefield Things (IOBT), which is focused on multiple challenges of incorporating the IoT ideas and capabilities within the battlefield environment.