

Channel States Classification in Cognitive Small Cell Networks with Multiple Transmission Powers

Danyang Wang, Zan Li, *Senior Member, IEEE*, Ning Zhang, Huici Wu, and Xuemin (Sherman) Shen, *Fellow, IEEE*

Abstract—Cognitive small cell networks have great potential in improving spectrum efficiency and mitigating inter-cell interference. Comprehensively classifying the channel states in cognitive small cell networks is important for efficiently reusing the spectrum bands that are licensed to macrocell. In this paper, we investigate channel states classification in cognitive small cell networks with multi-level of transmission powers, including occupation detection of spectrum bands and transmission power classification of macrocell base station (MBS). Specifically, two scenarios including priori known signaling features and unknown signaling features are both studied. For the former scenario, we propose an optimal spectrum sensing and power classification algorithm (OSC), based on coherent classification, to achieve accurate sensing performance by fully exploiting the inherent information of the signaling features. Optimal sensing threshold as well as decision regions are derived for detecting and classifying the transmission power of MBS. For the scenario without signaling features, a generic spectrum sensing and power classification (GSC) algorithm is proposed based on non-coherent classification with low implementation complexity. A new performance metric, i.e., classification probability, is introduced to comprehensively evaluate the classification capability of the proposed algorithms. Finally, extensive simulations are provided to verify the proposed algorithms.

Index Terms—Cognitive small cell, sensing based spectrum sharing, multiple transmission powers, channel states classification.

I. INTRODUCTION

Wireless data traffic driven by the mobile devices and evolutionary applications is predicted to increase 1000-fold over the next decade [1]. To effectively accommodate the

unprecedented growth of wireless data traffic, both industry and academia are devoting more efforts in developing next generation (5G) mobile networks. As a promising technology in the 5G, small cells are densely deployed to improve the network capacity and increase the spectrum efficiency by sharing spectrum resources with macrocell [2]. However, severe cross-tier interference (i.e., between small cell and macrocell) and co-tier interference (i.e., among small cells) can arise when the unplanned deployed small cells share the same spectrum resources. To address this problem, cognitive radio technology has been integrated with small cells to automatically monitor the channel condition and intelligently allocate the spectrum resources [3]–[5].

In cognitive small cell networks, the macrocell is deemed as the primary user that has the license of spectrum bands, and the small cell embedded with cognitive capability is deemed as the secondary user that is seeking opportunities to access the licensed spectrum bands of the macrocell [5]. Three spectrum access strategies exist for cognitive small cells: 1) *opportunistic spectrum access* (also known as overlay spectrum access), where the cognitive small cells opportunistically access the spectrum bands that are not occupied by the primary macrocell [6]; 2) *spectrum sharing* (also known as underlay spectrum access), where the cognitive small cells can always access the spectrum bands without causing harmful interference to the macrocell [7]; and 3) *sensing based spectrum sharing* (also known as hybrid overlay-underlay spectrum access), where the cognitive small cells flexibly switch between overlay and underlay models according to the sensing results about the macrocell licensed spectrum bands [8]–[11]. When the macrocell spectrum bands are detected to be unoccupied, the cognitive small cell adopts overlay model to utilize spectrum bands with a higher transmission power. Once the macrocell spectrum bands are detected to be occupied, the cognitive small cell switches to underlay model with a lower transmission power to limit the interference to macrocell.

With the sensing based spectrum sharing strategy, cognitive small cell can not only significantly mitigate the interference to both macrocell and other small cells, but also efficiently improve spectrum utilization for a high network capacity. In [12], a centralized hybrid overlay-underlay selection scheme was investigated to maximize the transmission efficiency of the cognitive small cell network. Moreover, downlink energy consumption of small cells was also analyzed. In [13], a hybrid underlay-overlay cognitive femtocell network was proposed, in which the subchannel allocation problem was formulated as a coalition formation game. In [14], by jointly considering

Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

D. Wang and Z. Li are with the State Key Laboratory of Integrated Service Networks, School of Telecommunications Engineering, Xidian University, Xi'an 710071, China (e-mail: danyangwang@stu.xidian.edu.cn; zanli@xidian.edu.cn).

N. Zhang is with the Department of Computer Science, Texas A&M University at Corpus Christi, Corpus Christi, TX, 78412, USA (e-mail: ning.zhang@tamucc.edu).

H. Wu is with the National Engineering Lab for Mobile Network Technologies, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: dailywu@bupt.edu.cn).

X. Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: xshen@bbr.uwaterloo.ca).

This work was supported in part by the Key Project of National Natural Science Foundation of China under Grant 61631015; by the Fundamental Research Funds for the Central Universities under Grant 7215433803; by the National Natural Science Foundation of China under Grant {61471395,61501354,61501356}; by the Natural Sciences and Engineering Research Council of Canada (NSERC).

Part of this work was presented at IEEE GLOBECOM 2017, Singapore.

the transmitter and receiver energy consumption, a time-frequency power-resource allocation strategy was developed in orthogonal frequency-division multiple-access (OFDMA) networks. In [15], the interference management for heterogeneous networks was investigated, where cognitive radio technique was deemed as an efficient approach in minimizing mutual interference between cells. In these existing works, two common assumptions are adopted: *i*) the macrocell base station (MBS) is either inactive or transmitting with a constant power; and *ii*) the received noise at the cognitive small cell is white Gaussian noise to simplify analysis. However, practical communication systems are usually with multiple transmission powers and spatially correlated noise:

- *Multiple Transmission Powers (MTP)*: In cognitive small cell networks, the MBS can adaptively change the transmission power to accommodate the diverse demands of mobile users [16], [17], since the cognitive small cell can assist to balance the traffic load of macrocell. When the MBS works on multiple transmission powers, i.e., the MTP scenario, the interference temperature of macrocell associated mobile users varies under different primary transmission powers. If the primary macrocell base station works on a low transmission power while the cognitive small cell adopts a constant power with underlay spectrum access model for concurrent transmission, serious interference would be introduced to the macrocell. Hence, the cognitive small cell has to capture the current interference temperature of the macrocell so as to prevent macrocell users from being interfered by the inappropriate access of cognitive small cell. To this end, two main challenges to facilitate the sensing based spectrum sharing in cognitive small cell networks with MTP are 1) detecting the occupation of the macrocell spectrum bands; and 2) classifying the transmission power on the occupied spectrum band.
- *Spatially Correlated Noise*: Due to the oversampling and imperfect filtering in practical communication systems, correlations can arise among the received noise [18]. For example, when the received signal and ambient noise are filtered by a narrowband filter at a receiver, the noise embedded in the output signal is correlated. Moreover, the small cells are densely and randomly deployed, such that the received noise at each cell contains tremendous unpredictable weak signals transmitted from other cells. These weak signals can be highly directive and correlated, which results in a spatially correlated noise [19]–[22]. The spatially correlated noise seriously degrades the detection and classification performance of existing works [23]–[27], e.g., lead to high false alarm probability and low classification probability.

These factors translate into an urgent need in developing channel states classification algorithms which are insensitive to the spatially correlated noise.

In this paper, we investigate the channel states classification in cognitive small cell networks with multi-level of transmission powers, in which we not only detect the occupation of spectrum bands, but also identify the transmission power

over the occupied spectrum band. Specifically, the occupation of spectrum bands is first detected by examining the posterior probability ratio between two different hypotheses. Then, the transmission power on the occupied spectrum band is identified by leveraging minimum Bayes risk criterion. When the signaling features of macrocell transmitted signals are available, we propose an optimal spectrum sensing and power classification (OSC) algorithm. The OSC algorithm is designed a coherent classification algorithm, where the decision metric is conducted as the output of matched filter weighted by channel gain and noise correlation factors. For the case without signaling features at cognitive small cell, we propose a more generic spectrum sensing and power classification (GSC) algorithm, which is designed as a non-coherent classification algorithm. The decision metric of GSC algorithm is conducted as the weighted energy of received signals. We prove that the total error rates of both proposed algorithms are convex function with respect to the sensing threshold, and the optimal thresholds are respectively derived such that the total error rates are minimized. Moreover, closed-form decision regions for classifying the transmission powers are theoretically derived. In a nutshell, the main contributions of this paper are summarized as follows:

- We propose two channel states classification algorithms considering multiple transmission powers. Specifically, when the signaling features of transmitted signal are available, an optimal spectrum sensing and power classification algorithm is proposed, which is a coherent classification algorithm. For the case where the signaling features are unavailable, a non-coherent classification algorithm is proposed, where the decision metric is the summation of received signal energy weighted by channel gain and noise correlation factors.
- The total error rate is proved to be a convex function with respect to the sensing threshold. By minimizing the total error rate, the optimal sensing threshold is obtained for detecting the occupation of the spectrum band. Additionally, closed-form decision regions for classifying transmission powers are theoretically derived.
- A unique phenomenon in MTP scenarios, namely power mask effect, is discussed whenever the priori probability of each transmission power is equal or not. To comprehensively evaluate the classification capability of the proposed algorithms, a new performance metric, i.e., classification probability is introduced. Simulation results demonstrate that both the proposed algorithms can classify the primary transmission power of the MBS and the classification performance improves when the sensing condition becomes better. Moreover, compared to GSC algorithm, OSC algorithm requires less samples to achieve a desired performance such that it can meet the real-time demand of channel states classification in cognitive small cell networks.

The remainder of this paper is organized as follows. Section II presents the system model and formulates channel states classification problem in MTP scenarios. Section III proposes the optimal spectrum sensing and power classification

algorithm if the signaling features of macrocell transmitted signal are available. Section IV presents an alternative generic spectrum sensing and power classification algorithm for cases where the signaling features are unavailable. Simulation results are provided in Section V, followed by the conclusion in Section VI.

II. SYSTEM MODEL

Consider a cognitive small cell network where cognitive small cells are overlaid with a primary macrocell to share the same spectrum resources. The spectrum bandwidth of macrocell is divided into several channels. Cognitive small cell base station firstly performs spectrum sensing and power classification to determine the status of each channel, and then access the available channel based on the classification results. The MBS operates on the licensed band with one of the discrete power levels P_i , $i = 1, \dots, L$. Without loss of generality, these power levels are ordered as $P_1 < P_2 < \dots < P_L$. Suppose that once a transmission power is chosen by the MBS, it will be used for a certain period during which the cognitive small cell could perform spectrum sensing and power classification as well as the subsequent data transmission. Cognitive small cell base station is equipped with M antennas to improve the sensing accuracy. To detect the occupation status of the channels, a set of N discrete time vector observations $\mathbf{x}[n]$, $n = 0, \dots, N-1$ are utilized. Hence, the hypothesis testing problem of interest can be expressed as

$$\begin{aligned} \mathcal{H}_0 : \mathbf{x}[n] &= \hat{\mathbf{w}}[n], \\ \mathcal{H}_i : \mathbf{x}[n] &= \sqrt{P_i} \tilde{\mathbf{h}}s[n] + \hat{\mathbf{w}}[n], \quad i = 1, \dots, L \end{aligned} \quad (1)$$

where \mathcal{H}_0 denotes the hypothesis that primary macrocell is inactive while \mathcal{H}_i indicates primary macrocell is operating with power P_i ; $\tilde{\mathbf{h}} = [\tilde{h}_1, \dots, \tilde{h}_M]^T$ represents the block-fading channel from the primary MBS to the cognitive small cell base station; $s[n]$ is the symbol transmitted from the primary macrocell that is normalized to have unit power; and $\hat{\mathbf{w}}[n] = [\hat{w}_1[n], \dots, \hat{w}_M[n]]^T$ denotes the spatially correlated noise. In block-fading channel, the channel coefficient vector $\tilde{\mathbf{h}}$ is considered to be constant for each antenna at the cognitive small cell base station during one sensing period. Moreover, $\tilde{\mathbf{h}}$, $s[n]$ and $\hat{\mathbf{w}}[n]$ are assumed to mutually independent with each other, which is in accordance with practical communication systems [28]–[30].

The priori probability of each state of primary macrocell is defined as $\Pr(\mathcal{H}_i)$, $i = 0, 1, \dots, L$. Also define $\mathcal{H}_{\text{on}} = \bigcup_{i=1}^L \mathcal{H}_i$ as the hypothesis that primary macrocell is active. Then, the priori probability of \mathcal{H}_{on} is $\Pr(\mathcal{H}_{\text{on}}) = \sum_{i=1}^L \Pr(\mathcal{H}_i)$. While the inactive state of primary macrocell, denoted by $\mathcal{H}_{\text{off}} \triangleq \mathcal{H}_0$, has the priori probability $\Pr(\mathcal{H}_{\text{off}}) = \Pr(\mathcal{H}_0)$. We assume that cognitive small cell has the knowledge of the available transmission power levels that primary macrocell can select, since those power levels are normally deterministic and regulated by the wide-used standards [16], [17]. In addition, the corresponding priori probabilities $\Pr(\mathcal{H}_i)$'s are assumed to be known at cognitive small cell.

A. Correlated Noise Model

The noise samples are considered to be correlated across the spatial dimensions and uncorrelated in temporal dimension. To analyze the spatial correlation characteristic of the noise observations, one-sided noise correlation model is adopted [18]. The correlated noise samples across spatial dimension are related to white noise as

$$\hat{\mathbf{W}} = \mathbf{\Theta}^{1/2} \mathbf{W}, \quad (2)$$

where \mathbf{W} is an $M \times N$ matrix with independent and identically distributed (i.i.d.) Gaussian entries with zero mean and variance σ_w^2 , representing the white noise, $\mathbf{\Theta}$ is an $M \times M$ Hermitian matrix with entries corresponding to the correlation among noise samples, and $\mathbf{\Theta}^{1/2}$ denotes the square root of $\mathbf{\Theta}$. A common model to effectively quantify the level of spatial correlation is exponential correlation model [31], [32]. More specifically, the exponential model can be given as

$$\theta_{ij} = \begin{cases} \rho^{j-i}, & i \leq j \\ (\rho^{i-j})^*, & i > j \end{cases} \quad (3)$$

where θ_{ij} is the (i, j) th element of $\mathbf{\Theta}$, $\rho \in \mathbb{C}$ is the correlation coefficient with $|\rho| \leq 1$, and $[\cdot]^*$ indicates the complex conjugate operator. We can easily notice that $\mathbf{\Theta}$ does not affect the noise power since it is normalized, i.e., $(1/M)\text{trace}\{\mathbf{\Theta}\} = 1$.

III. OPTIMAL SPECTRUM SENSING AND POWER CLASSIFICATION ALGORITHM

In this section, the optimal spectrum sensing and power classification algorithm with spatially correlated noise is proposed by leveraging the priori knowledge of signaling features. In this case, the n -th received signal sample has the distribution as follows:

$$\mathbf{x}[n] \sim \begin{cases} \mathcal{CN}(\mathbf{0}, \sigma_w^2 \mathbf{\Theta}), & \mathcal{H}_0; \\ \mathcal{CN}(\sqrt{P_i} \tilde{\mathbf{h}}s[n], \sigma_w^2 \mathbf{\Theta}), & \mathcal{H}_i. \end{cases} \quad (4)$$

Then, the received sample matrix at cognitive small cell base station can be represented as $\mathbf{X} = [\mathbf{x}[1], \dots, \mathbf{x}[N]]$, and the corresponding probability density function (PDF) under hypothesis \mathcal{H}_0 is expressed as

$$\begin{aligned} f(\mathbf{X}|\mathcal{H}_0) &= \prod_{n=1}^N f(\mathbf{x}[n]|\mathcal{H}_0) \\ &= \prod_{n=1}^N \frac{1}{\pi^M \det(\sigma_w^2 \mathbf{\Theta})} \exp \left\{ -\frac{\mathbf{x}^H[n] \mathbf{\Theta}^{-1} \mathbf{x}[n]}{\sigma_w^2} \right\} \\ &= \frac{1}{\pi^{MN} \{\det(\sigma_w^2 \mathbf{\Theta})\}^N} \exp \left\{ -\frac{1}{\sigma_w^2} \sum_{n=1}^N \mathbf{x}^H[n] \mathbf{\Theta}^{-1} \mathbf{x}[n] \right\}. \end{aligned} \quad (5)$$

Define $\mathbf{y}[n] = \mathbf{\Theta}^{-1}\mathbf{x}[n]$ and $\mathbf{Y} = [\mathbf{y}[1], \dots, \mathbf{y}[N]]$, hence there is $\mathbf{Y} = \mathbf{\Theta}^{-1}\mathbf{X}$. Further, (4) can be re-written as

$$\begin{aligned} f(\mathbf{X}|\mathcal{H}_0) &= \frac{1}{\pi^{MN} \{\det(\sigma_w^2 \mathbf{\Theta})\}^N} \exp \left\{ -\frac{1}{\sigma_w^2} \sum_{n=1}^N \mathbf{x}^H[n] \mathbf{y}[n] \right\} \\ &= \frac{1}{\pi^{MN} \{\det(\sigma_w^2 \mathbf{\Theta})\}^N} \exp \left\{ -\frac{1}{\sigma_w^2} \text{tr}(\mathbf{X}^H \mathbf{Y}) \right\} \\ &= \frac{1}{\pi^{MN} \{\det(\sigma_w^2 \mathbf{\Theta})\}^N} \exp \left\{ -\frac{1}{\sigma_w^2} \text{tr}(\mathbf{X}^H \mathbf{\Theta}^{-1} \mathbf{X}) \right\} \\ &\stackrel{(a)}{=} \frac{\exp \left\{ -\frac{\text{tr}(\mathbf{X} \mathbf{X}^H \mathbf{\Theta}^{-1})}{\sigma_w^2} \right\}}{\pi^{MN} \{\det(\sigma_w^2 \mathbf{\Theta})\}^N}, \end{aligned} \quad (6)$$

where (a) is derived by utilizing the invariant property of the trace under cyclic permutations [33]. Similarly, the PDF of received sample matrix \mathbf{X} under hypothesis \mathcal{H}_i can be expressed as

$$\begin{aligned} f(\mathbf{X}|\mathcal{H}_i) &= \frac{1}{\pi^{MN} \{\det(\sigma_w^2 \mathbf{\Theta})\}^N} \exp \left\{ -\frac{\text{tr}(\mathbf{X} \mathbf{X}^H \mathbf{\Theta}^{-1})}{\sigma_w^2} \right. \\ &\quad \left. + \frac{\sqrt{P_i} (\tilde{\mathbf{h}}^H \mathbf{\Theta}^{-1} \mathbf{X} \mathbf{s}^H + \mathbf{s} \mathbf{X}^H \mathbf{\Theta}^{-1} \tilde{\mathbf{h}})}{\sigma_w^2} - \frac{P_i \|\mathbf{s}\|^2 \tilde{\mathbf{h}}^H \mathbf{\Theta}^{-1} \tilde{\mathbf{h}}}{\sigma_w^2} \right\}. \end{aligned} \quad (7)$$

where \mathbf{s} is the macrocell transmitted signal vector during one sensing period, denoted by $\mathbf{s} = [s[1], s[2], \dots, s[N]]$.

In MTP scenarios, the primary target at cognitive small cell is to detect the occupation of macrocell channels, while a secondary target is to classify the transmission power on the occupied channel. The algorithm that detects the occupation of macrocell channels is given in Section III-A. Then, in Section III-B, the optimal sensing threshold is derived by minimizing the total error rate. The algorithm and decision thresholds for classifying the transmission power are given in Section III-C.

A. Occupation Detection of Macrocell Channel

To achieve the primary task, we first verify the hypothesis $\mathcal{H}_{\text{on}}/\mathcal{H}_{\text{off}}$. Let us examine the ratio of the posterior probabilities between hypothesis \mathcal{H}_{on} and hypothesis \mathcal{H}_{off}

$$\begin{aligned} \xi(\mathbf{X}) &= \frac{\Pr(\mathcal{H}_{\text{on}}|\mathbf{X})}{\Pr(\mathcal{H}_{\text{off}}|\mathbf{X})} = \frac{\sum_{i=1}^L \Pr(\mathcal{H}_i) f(\mathbf{X}|\mathcal{H}_i)}{\Pr(\mathcal{H}_0) f(\mathbf{X}|\mathcal{H}_0)} \\ &= \sum_{i=1}^L \frac{\Pr(\mathcal{H}_i)}{\Pr(\mathcal{H}_0)} \exp \left\{ \frac{\sqrt{P_i} (\tilde{\mathbf{h}}^H \mathbf{\Theta}^{-1} \mathbf{X} \mathbf{s}^H + \mathbf{s} \mathbf{X}^H \mathbf{\Theta}^{-1} \tilde{\mathbf{h}})}{\sigma_w^2} - \frac{P_i \|\mathbf{s}\|^2 \tilde{\mathbf{h}}^H \mathbf{\Theta}^{-1} \tilde{\mathbf{h}}}{\sigma_w^2} \right\} \end{aligned} \quad (8)$$

Since the signaling features of transmitted signal are known to cognitive small cell base station, it is easily found that $\xi(\mathbf{X})$ is strictly increasing over $\mathbb{T}_{\text{OSC}} \triangleq \tilde{\mathbf{h}}^H \mathbf{\Theta}^{-1} \mathbf{X} \mathbf{s}^H + \mathbf{s} \mathbf{X}^H \mathbf{\Theta}^{-1} \tilde{\mathbf{h}}$, and hence the decision criterion can be made as

$$\mathbb{T}_{\text{OSC}} = \tilde{\mathbf{h}}^H \mathbf{\Theta}^{-1} \mathbf{X} \mathbf{s}^H + \mathbf{s} \mathbf{X}^H \mathbf{\Theta}^{-1} \tilde{\mathbf{h}} \underset{\mathcal{H}_{\text{off}}}{\underset{\mathcal{H}_{\text{on}}}{\geq}} \theta, \quad (9)$$

where θ is the pre-determined sensing threshold. For simplifying analysis, we perform singular value decomposition (SVD) on noise correlation matrix $\mathbf{\Theta}$ given by $\mathbf{\Theta} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H$, where

\mathbf{U} is an $M \times M$ unitary matrix of the eigenvectors of $\mathbf{\Theta}$, and $\mathbf{\Lambda}$ is an $M \times M$ diagonal matrix whose diagonal elements are the corresponding eigenvalues, i.e., $\mathbf{\Lambda}_{ii} = \lambda_i$. Thus, (9) can be expressed as

$$\mathbb{T}_{\text{OSC}} = \mathbf{h}^H \mathbf{\Lambda}^{-1} \mathbf{Y} \mathbf{s}^H + \mathbf{s} \mathbf{Y}^H \mathbf{\Lambda}^{-1} \mathbf{h} \underset{\mathcal{H}_{\text{off}}}{\underset{\mathcal{H}_{\text{on}}}{\geq}} \theta, \quad (10)$$

where $\mathbf{h} = \mathbf{U}^H \tilde{\mathbf{h}}$ and $\mathbf{Y} = \mathbf{U}^H \mathbf{X}$. Note that \mathbf{Y} is a linear transformation of the received signal matrix \mathbf{X} , such that \mathbf{Y} has the same statistical properties as \mathbf{X} . Therefore, under the hypothesis \mathcal{H}_i , the test statistic \mathbb{T}_{OSC} follows Gaussian distribution:

$$\mathcal{H}_i : \mathbb{T}_{\text{OSC}} \sim \mathcal{N}(2\sqrt{P_i} \mu \|\mathbf{s}\|^2, 2\sigma_w^2 \mu \|\mathbf{s}\|^2), \quad (11)$$

where $\|\cdot\|$ denotes the standard vector norm, and $\mu = \mathbf{h}^H \mathbf{\Lambda}^{-1} \mathbf{h} = \sum_{m=1}^M \frac{|h_m|^2}{\lambda_m}$. Similar to the conventional cognitive radio networks, we resort to the false alarm probability \Pr_{fa} and the detection probability \Pr_{d} to evaluate the sensing performance of the proposed OSC algorithm, which can be separately calculated as

$$\begin{aligned} \Pr_{\text{fa}}(\theta) &= \Pr(\mathcal{H}_{\text{on}}|\mathcal{H}_{\text{off}}) = \Pr(\mathbb{T}_{\text{OSC}} > \theta|\mathcal{H}_{\text{off}}) \\ &= Q\left(\frac{\theta}{\sqrt{2\mu\sigma_w} \|\mathbf{s}\|}\right), \\ \Pr_{\text{d}}(\theta) &= \Pr(\mathcal{H}_{\text{on}}|\mathcal{H}_{\text{on}}) = \Pr(\mathbb{T}_{\text{OSC}} > \theta|\mathcal{H}_{\text{on}}) \\ &= \sum_{i=1}^L \frac{\Pr(\mathcal{H}_i)}{\Pr(\mathcal{H}_{\text{on}})} Q\left(\frac{\theta - 2\sqrt{P_i} \|\mathbf{s}\|^2 \mu}{\sqrt{2\mu\sigma_w} \|\mathbf{s}\|}\right), \end{aligned} \quad (12)$$

where $Q(\cdot)$ is the Complementary Cumulative Distribution Function (CCDF), i.e., the right tail probability of the standard normal distribution.

The pre-determined sensing threshold θ is crucial for the cognitive small cell networks to control the system performance. For example, a lower sensing threshold results in a higher false alarm probability $\Pr_{\text{fa}}(\theta)$, which will waste the opportunity to re-use the idle channel. Whereas, a higher sensing threshold leads to a higher miss-detection probability $\Pr_{\text{m}}(\theta)$, which will increase the interference to macrocell. Therefore, there exists a tradeoff between false alarm probability and miss-detection probability. We aim to calculate the optimal sensing threshold such that the total error rate $\Pr_{\text{e}}(\theta)$ of the proposed OSC algorithm is minimized, where the total error rate can be presented as:

$$\begin{aligned} \Pr_{\text{e}}(\theta) &= \Pr_{\text{fa}}(\theta) + \Pr_{\text{m}}(\theta) \\ &= 1 + \Pr_{\text{fa}}(\theta) - \Pr_{\text{d}}(\theta). \end{aligned} \quad (13)$$

B. Optimal Sensing Threshold

Since the sensing performance is related to both false alarm probability and detection probability, we aim to obtain the optimal sensing threshold θ^* that minimizes the total error rate $\Pr_{\text{e}}(\theta)$. Such a problem can be formulated as

$$\begin{aligned} \theta^* &= \arg \min_{\theta} \Pr_{\text{e}}(\theta) \\ &= \arg \min_{\theta} \left\{ 1 + Q\left(\frac{\theta}{\sqrt{2\mu\sigma_w} \|\mathbf{s}\|}\right) - \sum_{i=1}^L \frac{\Pr(\mathcal{H}_i)}{\Pr(\mathcal{H}_{\text{on}})} Q\left(\frac{\theta - 2\sqrt{P_i} \|\mathbf{s}\|^2 \mu}{\sqrt{2\mu\sigma_w} \|\mathbf{s}\|}\right) \right\} \end{aligned} \quad (14)$$

To prove the existence of the optimal sensing threshold for detecting the occupation of macrocell channel, we prove

- 1) $\frac{\partial \text{Pr}_e(\theta)}{\partial \theta} = 0$ has unique solution θ^* and
- 2) $\frac{\partial \text{Pr}_e(\theta)}{\partial \theta} < 0$ when $\theta < \theta^*$ and $\frac{\partial \text{Pr}_e(\theta)}{\partial \theta} > 0$ when $\theta > \theta^*$.

In the following, we provide the theorem and corresponding proof for the existence of optimal sensing threshold θ^* that minimizes the total error rate.

Theorem 1. There exists an optimal sensing threshold for detecting the occupation of macrocell channel such that the total error rate of the proposed OSC algorithm is minimized. Moreover, the optimal sensing threshold θ^* is the solution to the following equation.

$$\sum_{i=1}^L \frac{\Pr(\mathcal{H}_i)}{\Pr(\mathcal{H}_{\text{on}})} \exp\left(\frac{\sqrt{P_i}\theta - P_i \|\mathbf{s}\|^2 \mu}{\sigma_w^2}\right) = 1. \quad (15)$$

Proof: Differentiating $\text{Pr}_e(\theta)$ with respect to θ , we have

$$\begin{aligned} & \frac{\partial \text{Pr}_e(\theta)}{\partial \theta} \\ &= -\frac{1}{\delta\sqrt{2\pi}} e^{-\frac{\theta^2}{2\delta^2}} + \sum_{i=1}^L \frac{\Pr(\mathcal{H}_i)}{\Pr(\mathcal{H}_{\text{on}})} \frac{1}{\delta\sqrt{2\pi}} e^{-\frac{(\theta - 2\sqrt{P_i}\|\mathbf{s}\|^2\mu)^2}{2\delta^2}} \\ &= \frac{1}{\delta\sqrt{2\pi}} e^{-\frac{\theta^2}{2\delta^2}} g(\theta), \end{aligned} \quad (16)$$

where $\delta = \sqrt{2\mu\sigma_w \|\mathbf{s}\|}$, and $g(\theta)$ is defined as

$$g(\theta) = \sum_{i=1}^L \frac{\Pr(\mathcal{H}_i)}{\Pr(\mathcal{H}_{\text{on}})} \exp\left(\frac{\sqrt{P_i}\theta - P_i \|\mathbf{s}\|^2 \mu}{\sigma_w^2}\right) - 1. \quad (17)$$

Since $e^{-\theta^2/2\delta^2} > 0$ always holds when $\theta \in (-\infty, \infty)$, then $\frac{\partial \text{Pr}_e(\theta)}{\partial \theta} = 0$ if and only if $g(\theta) = 0$. With positive P_i , it can be easily found that $g(\theta)$ is monotonically increasing with respect to θ , and

$$\begin{aligned} \lim_{\theta \rightarrow -\infty} g(\theta) &= -1, \\ \lim_{\theta \rightarrow +\infty} g(\theta) &= +\infty. \end{aligned} \quad (18)$$

Hence, there exists a unique sensing threshold θ^* that satisfies $g(\theta^*) = 0$. Moreover, with the unique sensing threshold θ^* , the following inequalities always hold.

- When $\theta < \theta^*$, $g(\theta) < 0$, i.e., $\frac{\partial \text{Pr}_e(\theta)}{\partial \theta} < 0$.
- When $\theta > \theta^*$, $g(\theta) > 0$, i.e., $\frac{\partial \text{Pr}_e(\theta)}{\partial \theta} > 0$.

Note that $\text{Pr}_e(\theta)$ is a monotonically decreasing function with respect to θ when $\theta < \theta^*$, and a monotonically increasing function with respect to θ when $\theta > \theta^*$. Thus, the minimum total error rate is achieved at the unique sensing threshold θ^* .

Therefore, from the above analysis, it is concluded that the optimal sensing threshold θ^* uniquely exists for minimizing the total error rate, which is obtained by solving the equation $\frac{\partial \text{Pr}_e(\theta)}{\partial \theta} = 0$. ■

Remark 1. From (15), the optimal sensing threshold is difficult to be expressed in closed-form, while it can be calculated numerically. Specially, when the macrocell has only one non-zero transmission power, i.e., $L = 1$, the closed-form optimal sensing threshold can be obtained as $\theta^* = \sqrt{P_1} \|\mathbf{s}\|^2 \mu$. If the macrocell has multiple transmission powers, i.e., $L > 1$, the optimal sensing threshold can be calculated numerically.

C. Transmission Power Classification on the Occupied Channel

After detecting the occupation of macrocell channel, the next step is to identify at which power level the macrocell is operating over the occupied channel. When the macrocell has multiple different transmission powers, cognitive small cell might make various erroneous decisions when classifying the transmission powers. To distinguish the non-zero transmission power of macrocell, a multiple hypothesis testing problem is formulated, whereby the Bayes risk is employed. Define C_{ij} as the cost when the cognitive small cell claims that the macrocell is working on power level P_i while the macrocell is actually transmitting with power P_j . Then, the Bayes risk is

$$\begin{aligned} \mathcal{B} &= \sum_{i=1}^L \sum_{j=1}^L C_{ij} \Pr(\mathcal{H}_i | \mathcal{H}_j; \hat{\mathcal{H}}_{\text{on}}) \Pr(\mathcal{H}_j | \hat{\mathcal{H}}_{\text{on}}) \\ &= \sum_{i=1}^L \sum_{j=1}^L \int_{R_i} C_{ij} f(\mathbf{X} | \mathcal{H}_j; \hat{\mathcal{H}}_{\text{on}}) \Pr(\mathcal{H}_j | \hat{\mathcal{H}}_{\text{on}}) d\mathbf{X} \\ &= \sum_{i=1}^L \int_{R_i} \sum_{j=1}^L C_{ij} \Pr(\mathcal{H}_j | \mathbf{X}; \hat{\mathcal{H}}_{\text{on}}) f(\mathbf{X} | \hat{\mathcal{H}}_{\text{on}}) d\mathbf{X} \\ &= \sum_{i=1}^L \int_{R_i} C_i(\mathbf{X}) f(\mathbf{X} | \hat{\mathcal{H}}_{\text{on}}) d\mathbf{X} \end{aligned} \quad (19)$$

where $R_i = \{\mathbf{X}; \text{decide } \mathcal{H}_i\}$ is the partition of the observation space, $C_i(\mathbf{X}) = \sum_{j=1}^L C_{ij} \Pr(\mathcal{H}_j | \mathbf{X}; \hat{\mathcal{H}}_{\text{on}})$ denotes the average cost of deciding \mathcal{H}_i if \mathbf{X} is observed, and $\hat{\mathcal{H}}_{\text{on}}$ represents that $\mathbb{T}_{\text{OSC}} > \theta^*$, which has the equivalent region as $\mathbf{X} \in \mathcal{X}$. In order to minimize the Bayes risk \mathcal{B} , we minimize $C_i(\mathbf{X})$, i.e., the decision rule is given as

$$i^* = \arg \min_{i \in \{1, 2, \dots, L\}} C_i(\mathbf{X}). \quad (20)$$

For the particular cost assignment

$$C_{ij} = \begin{cases} 0, & i = j; \\ 1, & i \neq j, \end{cases} \quad (21)$$

the average cost can be expressed as

$$C_i(\mathbf{X}) = \sum_{j=1}^L \Pr(\mathcal{H}_j | \mathbf{X}; \hat{\mathcal{H}}_{\text{on}}) - \Pr(\mathcal{H}_i | \mathbf{X}; \hat{\mathcal{H}}_{\text{on}}) \quad (22)$$

Note that the first term is independent with i , and $C_i(\mathbf{X})$ is minimized by maximizing $\Pr(\mathcal{H}_i | \mathbf{X}; \hat{\mathcal{H}}_{\text{on}})$. Thus, the decision rule turns into MAP criterion, i.e.,

$$i^* = \arg \max_{i \in \{1, 2, \dots, L\}} \Pr(\mathcal{H}_i | \mathbf{X}; \hat{\mathcal{H}}_{\text{on}}), \quad \mathbf{X} \in \mathcal{X}. \quad (23)$$

Using Bayes rule, the posterior probability of hypothesis \mathcal{H}_i can be written as

$$\begin{aligned} \Pr(\mathcal{H}_i | \mathbf{X}; \hat{\mathcal{H}}_{\text{on}}) &= \frac{f(\mathbf{X} | \mathcal{H}_i; \hat{\mathcal{H}}_{\text{on}}) \Pr(\mathcal{H}_i; \hat{\mathcal{H}}_{\text{on}})}{f(\mathbf{X} | \hat{\mathcal{H}}_{\text{on}})} \\ &= \frac{f(\mathbf{X} | \mathcal{H}_i) \Pr(\mathcal{H}_i)}{f(\mathbf{X} | \hat{\mathcal{H}}_{\text{on}}) \Pr(\hat{\mathcal{H}}_{\text{on}})}. \end{aligned} \quad (24)$$

Then, the MAP criterion can be expressed as follows:

$$i^* = \arg \max_{i \in \{1, 2, \dots, L\}} f(\mathbf{X}|\mathcal{H}_i)\Pr(\mathcal{H}_i), \mathbf{X} \in \mathcal{X}. \quad (25)$$

For a hypothesis pair $(\mathcal{H}_i, \mathcal{H}_j), \forall i, j \geq 1, \mathcal{H}_i$ is decided rather than \mathcal{H}_j if

$$f(\mathbf{X}|\mathcal{H}_i)\Pr(\mathcal{H}_i) > f(\mathbf{X}|\mathcal{H}_j)\Pr(\mathcal{H}_j), \mathbf{X} \in \mathcal{X}. \quad (26)$$

With (7), we can expand (26) as

$$(\sqrt{P_i} - \sqrt{P_j})\mathbb{T}_{\text{OSC}} > \sigma_w^2 \ln \left[\frac{\Pr(\mathcal{H}_j)}{\Pr(\mathcal{H}_i)} \right] + (P_i - P_j) \|\mathbf{s}\|^2 \mu. \quad (27)$$

Define

$$\Psi(i, j) \triangleq \frac{\sigma_w^2}{\sqrt{P_i} - \sqrt{P_j}} \ln \left[\frac{\Pr(\mathcal{H}_j)}{\Pr(\mathcal{H}_i)} \right] + (\sqrt{P_i} + \sqrt{P_j}) \|\mathbf{s}\|^2 \mu. \quad (28)$$

To obtain the decision rule for each hypothesis pair with respect to the test statistics \mathbb{T}_{OSC} , we have the following inequalities.

- When $i > j$, i.e., $P_i > P_j$, there is $\mathbb{T}_{\text{OSC}} > \Psi(i, j), \forall i > j$.
- When $i < j$, i.e., $P_i < P_j$, there is $\mathbb{T}_{\text{OSC}} < \Psi(i, j), \forall i < j$.

For $1 < i < L$, the lower bound of $\mathcal{R}(\mathcal{H}_i)$ for hypothesis \mathcal{H}_i is $\max_{1 \leq j < i} \Psi(i, j)$ and the upper bound of $\mathcal{R}(\mathcal{H}_i)$ should be $\min_{i < j \leq L} \Psi(i, j)$. Additionally, the MAP detection for recognizing the non-zero transmission power of MBS is defined on the domain $\mathbf{X} \in \mathcal{X}$. Namely, all the decision regions of non-zero transmission powers should stay in $(\theta^*, +\infty)$. In summary, the decision regions of hypotheses \mathcal{H}_i 's, $i \in \{1, 2, \dots, L\}$ are given as (29) shown on the top of next page.

For the hypothesis \mathcal{H}_0 , the decision region is $(-\infty, \theta^*)$. Define $\theta_i, i \in \{1, 2, \dots, L\}$ as the threshold between $\mathcal{R}(\mathcal{H}_{i-1})$ and $\mathcal{R}(\mathcal{H}_i)$. Note that θ_1 is equal to θ^* . Moreover, define $\theta_0 \triangleq -\infty$ and $\theta_{L+1} \triangleq +\infty$ for consistence and completeness.

Note that for a certain decision region $\mathcal{R}(\mathcal{H}_{i_0})$, its lower bound may be greater than the upper bound, i.e.,

$$\max\{\theta_{\text{on/off}}, \max_{1 \leq j < i_0} \Psi(i_0, j)\} > \min_{i_0 < j \leq L} \Psi(i_0, j). \quad (30)$$

In this case, the decision region $\mathcal{R}(\mathcal{H}_{i_0})$ is empty and the corresponding transmission power P_{i_0} can never be detected. We name this phenomenon as power mask effect. Many reasons can lead to the power mask effect phenomenon, and we conclude main reasons as:

- 1) If the licensed channel is idle in most of the time, i.e., the priori probability of $\Pr(\mathcal{H}_0)$ is very large, P_0 may likely mask its adjacent transmission powers;
- 2) If the transmission power P_{i_0} is seldom used by the macrocell, i.e., the priori probability of $\Pr(\mathcal{H}_{i_0})$ is very small, then the transmission power P_{i_0} may be easily ignored by the cognitive small cell;
- 3) If P_{i_0} is very close to the adjacent transmission power P_{i_0-1} and P_{i_0+1} , then P_{i_0} is very likely to be masked by P_{i_0-1} or P_{i_0+1} when the instantaneous noise influence is large.

Remark 2. The leftmost transmission power P_0 and the rightmost transmission power P_L cannot be masked and are always detectable.

Fig. 1 shows an example to illustrate the power mask effect with unequal priori probabilities of transmission powers. Four non-zero transmission powers (i.e., P_1, P_2, P_3, P_4) are considered, and the corresponding priori probabilities $\Pr(\mathcal{H}_i), i = 1, 2, 3, 4$ are set as 0.26, 0.14, 0.03, 0.17, respectively. The priori probability $\Pr(\mathcal{H}_0)$ is set to 0.4. In Fig. 1, we demonstrate the decision thresholds and corresponding decision regions for each hypothesis. It can be seen that transmission power P_3 is masked due to the power mask effect. This is because that the corresponding priori probability of $\Pr(\mathcal{H}_3)$ is very small, i.e., the transmission power P_3 is rarely used. When the transmission power P_3 is masked, we set the threshold $\theta_4 = \theta_3$, and thus the decision region $\mathcal{R}(\mathcal{H}_3)$ is empty.

Lemma 1. Consider a special but practical case where $\Pr(\mathcal{H}_i) = \Pr(\mathcal{H}_j)$ in MTP scenarios. $\Psi(i, j)$ is independent to the priori probability of each transmission power, and is a monotonically increasing function with respect to P_j for any given $P_i, \forall i, j \in \{1, 2, \dots, L\}$.

Proof: When $\Pr(\mathcal{H}_i) = \Pr(\mathcal{H}_j)$ is considered, we have

$$\Psi(i, j) = (\sqrt{P_i} + \sqrt{P_j}) \|\mathbf{s}\|^2 \mu. \quad (31)$$

Since $\|\mathbf{s}\|^2$ and μ are positive constants, $\Psi(i, j)$ is monotonically increasing with respect to P_j for any given $P_i, \forall i, j \in \{1, 2, \dots, L\}$. ■

According to Lemma 1, when $\Pr(\mathcal{H}_i) = \Pr(\mathcal{H}_j)$, there is

$$\max_{1 \leq j < i} \Psi(i, j) = \Psi(i, i-1) < \Psi(i, i+1) = \min_{i < j \leq L} \Psi(i, j). \quad (32)$$

Hence, the non-zero transmission powers cannot mask each other, and the power mask effect may only happen when P_0 masks the non-zero transmission powers. In addition, for $i \geq 1$, there exists $\Psi(i, i+1) = \Psi(i+1, i)$. If the transmission power P_{i_0} has not been masked by the power P_0 , the lower bound of \mathcal{H}_{i_0} , i.e., $\Psi(i_0, i_0-1)$ equals to the upper bound of \mathcal{H}_{i_0-1} , i.e., $\Psi(i_0-1, i_0)$, and the upper bound of \mathcal{H}_{i_0} , i.e., $\Psi(i_0, i_0+1)$ equals to the lower bound of \mathcal{H}_{i_0+1} , i.e., $\Psi(i_0+1, i_0)$. Thus, there are no gaps between the decision regions of any two contiguous hypotheses.

Fig. 2 illustrates the power mask effect when each transmission power has equal priori probability. Similarly, four non-zero powers (i.e., P_1, P_2, P_3, P_4) are considered, while the corresponding priori probabilities $\Pr(\mathcal{H}_i), i = 1, 2, 3, 4$ are all set equally to 0.1. Hence, the priori probability $\Pr(\mathcal{H}_0)$ is 0.6. It can be noted that transmission power P_1 is masked, because the priori probability $\Pr(\mathcal{H}_0)$ is very large, i.e., the licensed channel is idle in most of the time. When P_1 is masked, we set the threshold $\theta_2 = \theta_1$, thus the decision region $\mathcal{R}(\mathcal{H}_1)$ is empty.

Different from the conventional binary hypothesis detection, purely evaluating the false alarm probability and detection probability is insufficient, since the cognitive small cell might make various error decisions in MTP scenarios. Therefore, we define $\Pr(\mathcal{H}_i|\mathcal{H}_j)$ as the probability that cognitive small cell

$$\mathcal{R}(\mathcal{H}_i) = \begin{cases} \mathbb{T}_{\text{OSC}} \in (\theta^*, \min_{1 < j \leq L} \Psi(1, j)), & i = 1; \\ \mathbb{T}_{\text{OSC}} \in (\max\{\theta^*, \max_{1 \leq j < i} \Psi(i, j)\}, \min_{i < j \leq L} \Psi(i, j)), & 1 < i < L; \\ \mathbb{T}_{\text{OSC}} \in (\max\{\theta^*, \max_{1 \leq j < L} \Psi(L, j)\}, +\infty), & i = L. \end{cases} \quad (29)$$

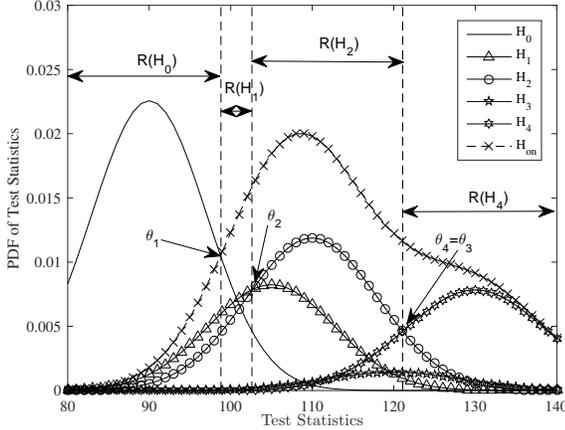


Fig. 1. Illustration for the power mask effect with unequal priori probabilities.

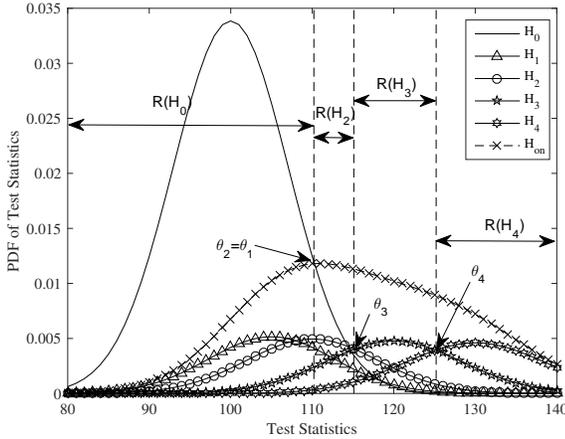


Fig. 2. Illustration for the power mask effect with equal priori probabilities.

claims the MBS is operating at transmission power P_i while the MBS actually operates at P_j . Then, there is

$$\Pr(\mathcal{H}_i|\mathcal{H}_j) = Q\left(\frac{\theta_i - 2\sqrt{P_j} \|\mathbf{s}\|^2 \mu}{\sqrt{2\mu\sigma_w} \|\mathbf{s}\|}\right) - Q\left(\frac{\theta_{i+1} - 2\sqrt{P_j} \|\mathbf{s}\|^2 \mu}{\sqrt{2\mu\sigma_w} \|\mathbf{s}\|}\right). \quad (33)$$

Additionally, for MTP scenarios, we introduce a new performance metric, i.e., classification probability to describe the classification capability of the proposed OSC algorithm, which is defined as

$$\Pr_c = \frac{1}{\Pr(\mathcal{H}_{\text{on}})} \sum_{i=1}^L \Pr(\mathcal{H}_i) \Pr(\mathcal{H}_i|\mathcal{H}_i). \quad (34)$$

Remark 3. The false alarm probability \Pr_{fa} and detection probability \Pr_d can also be calculated using the probability $\Pr(\mathcal{H}_i|\mathcal{H}_j)$:

$$\Pr_{\text{fa}} = \Pr(\mathcal{H}_{\text{on}}|\mathcal{H}_{\text{off}}) = \sum_{i=1}^L \Pr(\mathcal{H}_i|\mathcal{H}_0), \quad (35)$$

$$\Pr_d = \Pr(\mathcal{H}_{\text{on}}|\mathcal{H}_{\text{on}}) = \frac{1}{\Pr(\mathcal{H}_{\text{on}})} \sum_{i=1}^L \sum_{j=1}^L \Pr(\mathcal{H}_j|\mathcal{H}_i) \Pr(\mathcal{H}_i). \quad (36)$$

IV. GENERIC SPECTRUM SENSING AND POWER CLASSIFICATION ALGORITHM

When the cognitive small cell has no information regarding the signaling features of the transmitted signal, we propose an alternative generic spectrum sensing and power classification algorithm. According to the practical communication systems, the transmitted signal can be reasonably supposed as a complex Gaussian process [34]–[36]. Moreover, this assumption indicates the worst case in spectrum sensing, where the derived detection and classification probabilities describe the low bound of the performance. In this case, the n -th received signal sample has the distribution under each hypothesis as:

$$\mathbf{x}[n] \sim \begin{cases} \mathcal{CN}(\mathbf{0}, \sigma_w^2 \boldsymbol{\Theta}), & \mathcal{H}_0; \\ \mathcal{CN}(\mathbf{0}, P_i \tilde{\mathbf{h}} \tilde{\mathbf{h}}^H + \sigma_w^2 \boldsymbol{\Theta}), & \mathcal{H}_i. \end{cases} \quad (37)$$

Thereafter, under hypothesis \mathcal{H}_0 , the PDF of the observation matrix \mathbf{X} can be written as:

$$p(\mathbf{X}|\mathcal{H}_0) = \frac{\exp\{-\text{tr}(\mathbf{X}\mathbf{X}^H(\sigma_w^2 \boldsymbol{\Theta})^{-1})\}}{\pi^{MN} \{\det(\sigma_w^2 \boldsymbol{\Theta})\}^N}, \quad (38)$$

where $\text{tr}(\cdot)$ denotes the trace of matrix. Similarly, under \mathcal{H}_i the PDF of the observation matrix \mathbf{X} is as follows:

$$p(\mathbf{X}|\mathcal{H}_i) = \frac{\exp\{-\text{tr}(\mathbf{X}\mathbf{X}^H(P_i \tilde{\mathbf{h}} \tilde{\mathbf{h}}^H + \sigma_w^2 \boldsymbol{\Theta})^{-1})\}}{\pi^{MN} \{\det(P_i \tilde{\mathbf{h}} \tilde{\mathbf{h}}^H + \sigma_w^2 \boldsymbol{\Theta})\}^N}. \quad (39)$$

A. Occupation Detection of Macrocell Channel

In this subsection, we first verify the hypothesis $\mathcal{H}_{\text{on}}/\mathcal{H}_{\text{off}}$. The ratio of the posterior probabilities between two hypotheses can be written as

$$\begin{aligned} \zeta(\mathbf{X}) &= \frac{\Pr(\mathcal{H}_{\text{on}}|\mathbf{X})}{\Pr(\mathcal{H}_{\text{off}}|\mathbf{X})} = \frac{\sum_{i=1}^L \Pr(\mathcal{H}_i) p(\mathbf{X}|\mathcal{H}_i)}{\Pr(\mathcal{H}_0) p(\mathbf{X}|\mathcal{H}_0)} \\ &= \sum_{i=1}^L \frac{\Pr(\mathcal{H}_i)}{\Pr(\mathcal{H}_0)} \frac{\det(\sigma_w^2 \boldsymbol{\Theta})^N}{\det(P_i \tilde{\mathbf{h}} \tilde{\mathbf{h}}^H + \sigma_w^2 \boldsymbol{\Theta})^N} \\ &\quad \times \exp\left\{ \frac{\text{tr}(\mathbf{X}\mathbf{X}^H(\sigma_w^2 \boldsymbol{\Theta})^{-1})}{-\text{tr}(\mathbf{X}\mathbf{X}^H(P_i \tilde{\mathbf{h}} \tilde{\mathbf{h}}^H + \sigma_w^2 \boldsymbol{\Theta})^{-1})} \right\}. \end{aligned} \quad (40)$$

Using the following Sherman-Morrison formula [37]

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1})^{-1}\mathbf{DA}^{-1}, \quad (41)$$

we obtain

$$(\Theta + \frac{P_i}{\sigma_w^2} \tilde{\mathbf{h}}\tilde{\mathbf{h}}^H)^{-1} = \Theta^{-1} - \frac{\Theta^{-1}\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H\Theta^{-1}}{\sigma_w^2/P_i + \tilde{\mathbf{h}}^H\Theta^{-1}\tilde{\mathbf{h}}}. \quad (42)$$

Additionally, with the aid of $\Theta = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, $\mathbf{Y} = \mathbf{U}^H\mathbf{X}$, and $\tilde{\mathbf{h}} = \mathbf{U}^H\tilde{\mathbf{h}}$, the ratio of the posterior probabilities (40) can be simplified as

$$\zeta(\mathbf{Y}) = \sum_{i=1}^L \frac{\Pr(\mathcal{H}_i)}{\Pr(\mathcal{H}_0)} \frac{\det(\sigma_w^2\mathbf{\Lambda})^N}{\det(P_i\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H + \sigma_w^2\mathbf{\Lambda})^N} \times \exp \left\{ \frac{\text{tr}(\mathbf{Y}\mathbf{Y}^H\mathbf{\Lambda}^{-1}\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H\mathbf{\Lambda}^{-1})}{\sigma_w^2(\sigma_w^2/P_i + \tilde{\mathbf{h}}^H\mathbf{\Lambda}^{-1}\tilde{\mathbf{h}})} \right\}. \quad (43)$$

Note that $\zeta(\mathbf{Y})$ is strictly increasing over $\mathbb{T}_{\text{GSC}} \triangleq \text{tr}(\mathbf{Y}\mathbf{Y}^H\mathbf{\Lambda}^{-1}\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H\mathbf{\Lambda}^{-1})$, and therefore the decision criterion for detecting the occupation of macrocell channel can be given as

$$\mathbb{T}_{\text{GSC}} = \text{tr}(\mathbf{Y}\mathbf{Y}^H\mathbf{\Lambda}^{-1}\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H\mathbf{\Lambda}^{-1}) \underset{\mathcal{H}_{\text{off}}}{\underset{\mathcal{H}_{\text{on}}}{\gtrless}} \eta, \quad (44)$$

where η is the pre-determined sensing threshold. To obtain the sensing threshold, we first calculate the distribution of the test statistics. Note that the noise covariance matrix Θ is a Hermitian matrix, and hence $\mathbf{\Lambda}$ is real and diagonal, i.e., there is $(\mathbf{\Lambda}^{-1})^H = \mathbf{\Lambda}^{-1}$. Therefore, the test statistics can be written as

$$\begin{aligned} \mathbb{T}_{\text{GSC}} &= \text{tr}(\mathbf{Y}\mathbf{Y}^H\mathbf{\Lambda}^{-1}\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H\mathbf{\Lambda}^{-1}) \\ &= \text{tr}(\tilde{\mathbf{h}}^H\mathbf{\Lambda}^{-1}\mathbf{Y}\mathbf{Y}^H(\mathbf{\Lambda}^{-1})^H\tilde{\mathbf{h}}) \\ &= \|\mathbf{g}^H\mathbf{Y}\|^2, \end{aligned} \quad (45)$$

where $\mathbf{g}^H = \tilde{\mathbf{h}}^H\mathbf{\Lambda}^{-1}$. The distribution of the test statistics is

$$\frac{\mathbb{T}_{\text{GSC}}}{\frac{P_i\mu^2 + \mu\sigma_w^2}{2}} = \frac{\|\mathbf{g}^H\mathbf{Y}\|^2}{\frac{P_i\mu^2 + \mu\sigma_w^2}{2}} \sim \chi_{2N}^2, \quad (46)$$

where χ_{2N}^2 is the centralized chi-squared distribution with $2N$ degrees of freedom. As a result, the false alarm probability and detection probability can be calculated as

$$\Pr_{\text{fa}}(\eta) = \Pr(\mathcal{H}_{\text{on}}|\mathcal{H}_{\text{off}}) = \Pr(\mathbb{T}_{\text{GSC}} > \eta|\mathcal{H}_{\text{off}}) = \frac{\Gamma(N, \frac{\eta}{\mu\sigma_w^2})}{\Gamma(N)}, \quad (47)$$

$$\begin{aligned} \Pr_{\text{d}}(\eta) &= \Pr(\mathcal{H}_{\text{on}}|\mathcal{H}_{\text{on}}) = \sum_{i=1}^L \frac{\Pr(\mathcal{H}_i)}{\Pr(\mathcal{H}_{\text{on}})} \Pr(\mathbb{T}_{\text{GSC}} > \eta|\mathcal{H}_i) \\ &= \sum_{i=1}^L \frac{\Pr(\mathcal{H}_i)}{\Pr(\mathcal{H}_{\text{on}})} \frac{\Gamma(N, \frac{\eta}{P_i\mu^2 + \mu\sigma_w^2})}{\Gamma(N)}, \end{aligned} \quad (48)$$

where $\Gamma(\cdot, \cdot)$ is the upper incomplete Gamma function with the definition $\Gamma(s, x) = \int_x^\infty t^{s-1}e^{-t}dt$ and $\Gamma(\cdot)$ denotes the Gamma function.

B. Optimal Sensing Threshold

Similar to Section III-B, we aim to obtain the optimal sensing threshold η^* to minimize the total error rate $\Pr_e(\eta)$, where the total error rate is defined as

$$\begin{aligned} \Pr_e(\eta) &= 1 + \Pr_{\text{fa}}(\eta) - \Pr_{\text{d}}(\eta) \\ &= \frac{\Gamma(N, \frac{\eta}{\mu\sigma_w^2})}{\Gamma(N)} + \sum_{i=1}^L \frac{\Pr(\mathcal{H}_i)}{\Pr(\mathcal{H}_{\text{on}})} \frac{\gamma(N, \frac{\eta}{P_i\mu^2 + \mu\sigma_w^2})}{\Gamma(N)}, \end{aligned} \quad (49)$$

where $\gamma(\cdot, \cdot)$ is the lower incomplete Gamma function with the definition $\gamma(s, x) = \int_0^x t^{s-1}e^{-t}dt$.

To prove the existence of the optimal sensing threshold for detecting the occupation of macrocell channel, we prove

- 1) $\frac{\partial \Pr_e(\eta)}{\partial \eta} = 0$ has unique solution η^* and
- 2) $\frac{\partial \Pr_e(\eta)}{\partial \eta} < 0$ when $\eta < \eta^*$ and $\frac{\partial \Pr_e(\eta)}{\partial \eta} > 0$ when $\eta > \eta^*$.

In the following, we provide the theorem and the corresponding proof for the existence of the optimal sensing threshold η that minimizes the total error rate $\Pr_e(\eta)$.

Theorem 2. There exists an optimal sensing threshold for detecting the occupation of macrocell channel such that the total error rate of the proposed GSC algorithm is minimized. Moreover, the optimal sensing threshold η is the solution to the following equation.

$$\sum_{i=1}^L \frac{\Pr(\mathcal{H}_i)}{\Pr(\mathcal{H}_{\text{on}})} \left(\frac{\sigma_w^2}{P_i\mu + \sigma_w^2} \right)^N \exp \left\{ \frac{P_i\eta}{\sigma_w^2(P_i\mu + \sigma_w^2)} \right\} = 1. \quad (50)$$

Proof: Differentiating $\Pr_e(\eta)$ in (49) with respect to η , we have

$$\begin{aligned} \frac{\partial \Pr_e(\eta)}{\partial \eta} &= -\frac{\eta^{N-1}e^{-\frac{\eta}{\mu\sigma_w^2}}}{\Gamma(N)(\mu\sigma_w^2)^N} + \frac{1}{\Gamma(N)} \sum_{i=1}^L \frac{\Pr(\mathcal{H}_i)}{\Pr(\mathcal{H}_{\text{on}})} \frac{\eta^{N-1}e^{-\frac{\eta}{P_i\mu^2 + \mu\sigma_w^2}}}{(P_i\mu^2 + \mu\sigma_w^2)^N} \\ &= \frac{\eta^{N-1}e^{-\frac{\eta}{\mu\sigma_w^2}}}{\Gamma(N)(\mu\sigma_w^2)^N} d(\eta), \end{aligned} \quad (51)$$

where $d(\eta)$ is defined as

$$d(\eta) = \sum_{i=1}^L \frac{\Pr(\mathcal{H}_i)}{\Pr(\mathcal{H}_{\text{on}})} \left(\frac{\sigma_w^2}{P_i\mu + \sigma_w^2} \right)^N \exp \left\{ \frac{P_i\eta}{\sigma_w^2(P_i\mu + \sigma_w^2)} \right\} - 1. \quad (52)$$

Note that the test statistics $\mathbb{T}_{\text{GSC}} = \|\mathbf{g}^H\mathbf{Y}\|^2$ is always nonnegative, i.e., $\mathbb{T}_{\text{GSC}} \in [0, \infty)$, hence the sensing threshold η must be positive, i.e., $\eta \in (0, \infty)$. In this case, $\frac{\partial \Pr_e(\eta)}{\partial \eta} = 0$ holds if and only if $d(\eta) = 0$ happens. With the definition of the priori probability of each transmission power, it holds that $\sum_{i=1}^L \frac{\Pr(\mathcal{H}_i)}{\Pr(\mathcal{H}_{\text{on}})} = 1$. As a result, we have

$$\sum_{i=1}^L \frac{\Pr(\mathcal{H}_i)}{\Pr(\mathcal{H}_{\text{on}})} \left(\frac{\sigma_w^2}{P_i\mu + \sigma_w^2} \right)^N < 1, \quad (53)$$

since $\sigma_w^2/(P_i\mu + \sigma_w^2) < 1$ is always true. We can easily verify that $\lim_{\eta \rightarrow 0} d(\eta) < 0$ when the sampling numbers N is

sufficiently large, and $d(\eta)$ is monotonically increasing with respect to η such that there exists a unique sensing threshold η^* satisfies $d(\eta^*) = 0$. In addition, with the unique sensing threshold η^* , the following inequalities hold.

- When $\eta < \eta^*$, $d(\eta) < 0$, i.e., $\frac{\partial \text{Pr}_e(\eta)}{\partial \eta} < 0$.
- When $\eta > \eta^*$, $d(\eta) > 0$, i.e., $\frac{\partial \text{Pr}_e(\eta)}{\partial \eta} > 0$.

Hence, $\text{Pr}_e(\eta)$ is a monotonically decreasing function with respect to η when $\eta < \eta^*$, and a monotonically increasing function when $\eta > \eta^*$.

Therefore, to sum up the above analysis, it is concluded that the unique optimal sensing threshold for minimizing the total error rate $\text{Pr}(\eta)$ exists, and it can be obtained by solving the equation $\frac{\partial \text{Pr}_e(\eta)}{\partial \eta} = 0$. ■

C. Transmission Power Classification on the Occupied Channel

After the macrocell channel is detected to be occupied, power classification is then performed to identify at which power level the MBS is operating. Recalling the MAP criterion derived in section III-C, the optimal decision is “ i ” if

$$i^* = \arg \max_{i \in \{1, 2, \dots, L\}} p(\mathbf{X}|\mathcal{H}_i)\text{Pr}(\mathcal{H}_i), \mathbf{X} \in \mathcal{X} \quad (54)$$

For a hypothesis pair $(\mathcal{H}_i, \mathcal{H}_j)$, $\forall i, j \geq 1$, \mathcal{H}_i is decided rather than \mathcal{H}_j if

$$p(\mathbf{X}|\mathcal{H}_i)\text{Pr}(\mathcal{H}_i) > p(\mathbf{X}|\mathcal{H}_j)\text{Pr}(\mathcal{H}_j), i \neq j. \quad (55)$$

With (39) and (42), the inequality (55) can be re-written as

$$\frac{(P_i - P_j)\mathbb{T}_{\text{GSC}}}{(\sigma_w^2 + P_i\mu)(\sigma_w^2 + P_j\mu)} > \ln \left[\frac{\text{Pr}(\mathcal{H}_j)}{\text{Pr}(\mathcal{H}_i)} \right] + N \ln \left[\frac{\det(P_i \mathbf{h}\mathbf{h}^H + \sigma_w^2 \mathbf{\Lambda})}{\det(P_j \mathbf{h}\mathbf{h}^H + \sigma_w^2 \mathbf{\Lambda})} \right], i \neq j. \quad (56)$$

As $P_1 < P_2 < \dots < P_L$, all $L - 1$ inequalities in (56) can be simplified to

$$\max_{1 \leq j < i} \Xi(i, j) < \mathbb{T}_{\text{GSC}} < \min_{i < j \leq L} \Xi(i, j), \quad (57)$$

where

$$\Xi(i, j) = \frac{(\sigma_w^2 + P_i\mu)(\sigma_w^2 + P_j\mu)}{P_i - P_j} \times \left\{ \ln \left[\frac{\text{Pr}(\mathcal{H}_j)}{\text{Pr}(\mathcal{H}_i)} \right] + N \ln \left[\frac{\det(P_i \mathbf{h}\mathbf{h}^H + \sigma_w^2 \mathbf{\Lambda})}{\det(P_j \mathbf{h}\mathbf{h}^H + \sigma_w^2 \mathbf{\Lambda})} \right] \right\} \quad (58)$$

Hence, when $1 < i < L$, the lower bound of decision region $\mathcal{D}(\mathcal{H}_i)$ for hypothesis \mathcal{H}_i , is $\max_{1 \leq j < i} \Xi(i, j)$ and the upper bound of $\mathcal{D}(\mathcal{H}_i)$ should be $\min_{i < j \leq L} \Xi(i, j)$. Moreover, all the decision regions of non-zero transmission power should stay in $(\eta^*, +\infty)$. Hence, the decision regions of hypothesis \mathcal{H}_i 's are given as (59) shown on the top of this page.

For the hypothesis \mathcal{H}_0 , the decision region is $[0, \eta^*)$. Define $\eta_i, i \in \{1, 2, \dots, L\}$ as the threshold between $\mathcal{D}(\mathcal{H}_{i-1})$ and $\mathcal{D}(\mathcal{H}_i)$. Note that η_0 is equal to 0 and η_L is equal to η^* . Additionally, define $\theta_{L+1} \triangleq +\infty$ for consistence and completeness.

Remark 4. Similarly, the power mask effect still exists in the GSC algorithm, since it is mainly caused by the rare usage of some transmission powers. We provide the following Lemma to show that the non-zero transmission powers cannot mask each other when there is $\text{Pr}(\mathcal{H}_i) = \text{Pr}(\mathcal{H}_j)$.

Lemma 2. When the special case that $\text{Pr}(\mathcal{H}_i) = \text{Pr}(\mathcal{H}_j)$ is considered in MTP scenarios, $\Xi(i, j)$ is independent with the priori probability of each transmission power and is a monotonically increasing function over P_j for any given $P_i, \forall i, j \in \{1, 2, \dots, L\}$.

Proof: Please refer to Appendix A. ■

According to Lemma 2, there is $\Xi(i, i - 1) < \Xi(i, i + 1)$ when $\text{Pr}(\mathcal{H}_i) = \text{Pr}(\mathcal{H}_j), i, j > 1$. Hence, the non-zero transmission powers cannot mask each other, and the power mask effect may only happen when P_0 masks the non-zero transmission powers.

The probability $\text{Pr}(\mathcal{H}_i|\mathcal{H}_j)$ that the cognitive small cell claims the MBS is operating at transmission power P_i while the MBS actually operates at P_j , can be calculated as

$$\begin{aligned} \text{Pr}(\mathcal{H}_i|\mathcal{H}_j) &= \text{Pr}(\eta_{i+1} > \mathbb{T}_{\text{GSC}} > \eta_i|\mathcal{H}_j) \\ &= \frac{\gamma(N, \frac{\eta_{i+1}}{P_j\mu^2 + \mu\sigma_w^2})}{\Gamma(N)} - \frac{\gamma(N, \frac{\eta_i}{P_j\mu^2 + \mu\sigma_w^2})}{\Gamma(N)}. \end{aligned} \quad (60)$$

Moreover, the classification probability, false alarm probability, and detection probability can be obtained using (34), (35), and (36) by replacing $\text{Pr}(\mathcal{H}_i|\mathcal{H}_j)$ with (60).

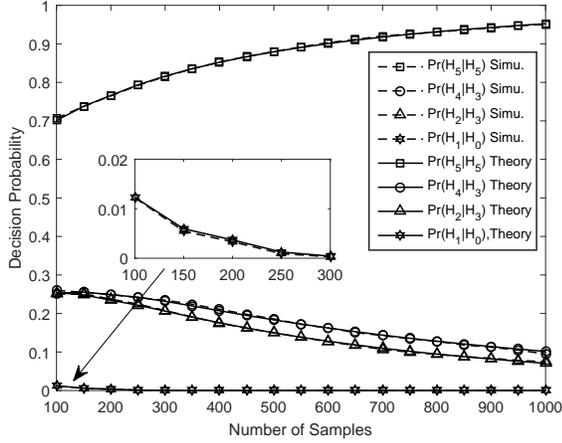
V. SIMULATION RESULTS

In this section, simulation results are provided to evaluate the performance of the proposed algorithms. We consider that the MBS has five transmission powers, and the corresponding probabilities are set to $\text{Pr}(\mathcal{H}_0) = 0.4$ and $\text{Pr}(\mathcal{H}_i) = 0.12, i = 1, \dots, 5$, respectively. The transmission powers satisfy $P_1 : P_2 : P_3 : P_4 : P_5 = 3 : 5 : 7 : 9 : 12$. The average signal-to-noise ratio (SNR) is defined as $1/5 \sum_{i=1}^5 P_i \mathbb{E}[\|\mathbf{h}\|^2] / \sigma_w^2$. All the results are averaged over 10000 tests using Monte-Carlo simulations.

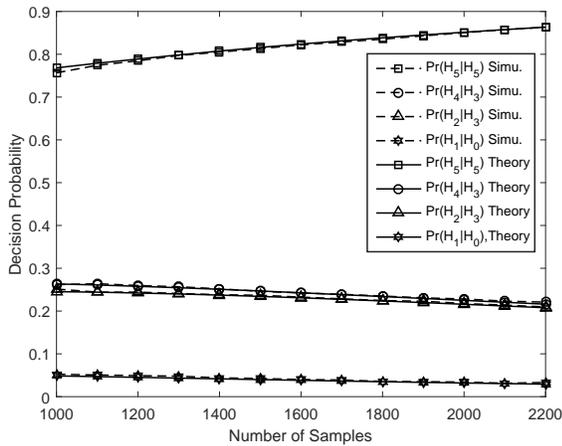
A. Decision Probability

The decision probability versus the number of samples of the proposed OSC and GSC algorithms are shown in Fig. 3(a) and Fig. 3(b), respectively. In the simulation, the number of antennas M at cognitive small cell is set to be 4, and the noise correlation coefficient ρ is set to be 0.2. The average SNR is chosen as -8dB . The number of samples are arranging from 100 to 1000 in evaluating OSC algorithm and from 1000 to 2200 in evaluating GSC algorithm. From both Fig. 3(a) and Fig. 3(b), it is seen that the simulation results match the theoretical results well. Furthermore, the correct decision probabilities raises while the error decision probabilities reduces with the increase of the number of samples. This indicates that both the detection and classification performance of both the proposed algorithms can be improved when the number of samples becomes large. Additionally, even when the number of samples lows to 100, the probability $\text{Pr}(\mathcal{H}_1|\mathcal{H}_0)$ of the OSC algorithm still approaches to zero. It suggests that if the

$$\mathcal{D}(\mathcal{H}_i) = \begin{cases} \mathbb{T}_{\text{GSC}} \in (\eta^*, \min_{1 < j \leq L} \Xi(1, j)), & i = 1; \\ \mathbb{T}_{\text{GSC}} \in (\max\{\eta^*, \max_{1 \leq j < i} \Xi(i, j)\}, \min_{i < j \leq L} \Xi(i, j)), & 1 < i < L; \\ \mathbb{T}_{\text{GSC}} \in (\max\{\eta^*, \max_{1 \leq j < L} \Xi(L, j)\}, +\infty), & i = L. \end{cases} \quad (59)$$



(a) Decision probability for the OSC algorithm



(b) Decision probability for the GSC algorithm

Fig. 3. Decision probability vs number of samples with $M = 4$, $\rho = 0.2$, and $\text{SNR} = -8\text{dB}$.

cognitive small cell has priori knowledge about the signaling features of the transmitted signal, the false alarm probability can be controlled below a very low level even when the number of samples is extremely low, which is very important for the cognitive small cell networks.

B. Detection and Classification Probability

The detection and classification probability versus the number of samples of the proposed OSC and GSC algorithms are shown in Fig. 4(a) and Fig. 4(b), respectively. In the simulation, we set the average SNR to be -20dB and -15dB for the OSC algorithm, and -12dB and -8dB for the GSC algorithm. In Fig. 4(a) and Fig. 4(b), it is shown that the detection and classification probabilities of both the proposed algorithms

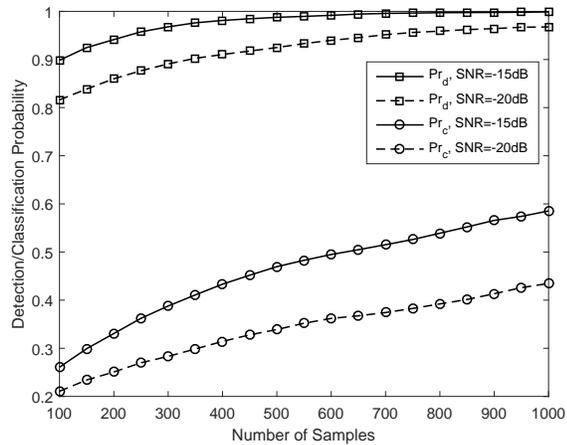
improve as the number of samples increases. Additionally, the detection probability is much higher than the classification probability for both algorithms. This is because that even when the macrocell channel is detected to be occupied, the cognitive small cell might make mistakes in classifying the transmission powers on the occupied channel. It indicates that in low SNR region, the occupation of the macrocell channel can be correctly detected, while the exact transmission power is difficult to be distinguished. Hence, in the low SNR region, the cognitive small cell is suggested to adopt overlay model rather than underlay model for limiting the interference introduced to the macro cell. Moreover, the gaps between the detection probability and classification probability decrease for both algorithms when the number of samples becomes larger or the SNR becomes higher. This indicates that when the sensing condition becomes better, the mistakes made in classifying the transmission powers gradually decrease and only the classification probability can effectively evaluate the detection and classification capability.

C. Impact of the Number of Antennas

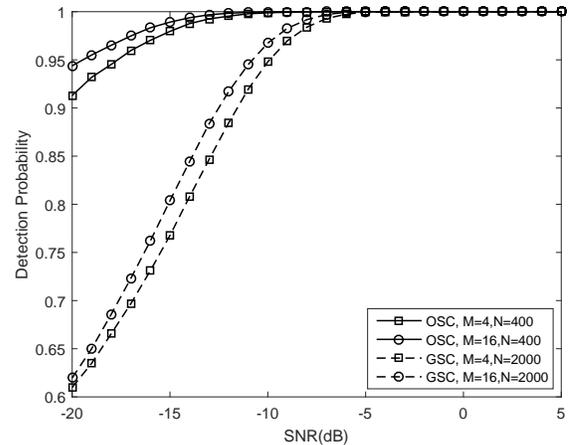
The impact of the number of antennas on the detection probability and classification probability are shown in Fig. 5(a) and Fig. 5(b), respectively. In the simulation, we set the noise correlation coefficient ρ to be 0.2. The number of samples N is chosen as 400 and 2000, respectively. It is seen that the detection and classification performance improves with the increase of the number of antennas for both the proposed algorithms. This is due to the fact that both the proposed OSC and GSC algorithms can exploit the space diversity gain of the multiple antennas. The amount of the improvement on classification performance of the OSC algorithm is about 2dB, when the number of antennas M increases from 4 to 16. As for the GSC algorithm, the amount of the improvement on classification performance is around 1dB, when M increases from 4 to 16. Increasing the number of antennas at cognitive small cell base station can improve the detection and classification performance, while it also leads to an increase on the cost and complexity of hardware devices. Therefore, a compromise between the performance and the complexity exists for the cognitive small cell networks.

D. Impact of the Noise Correlation

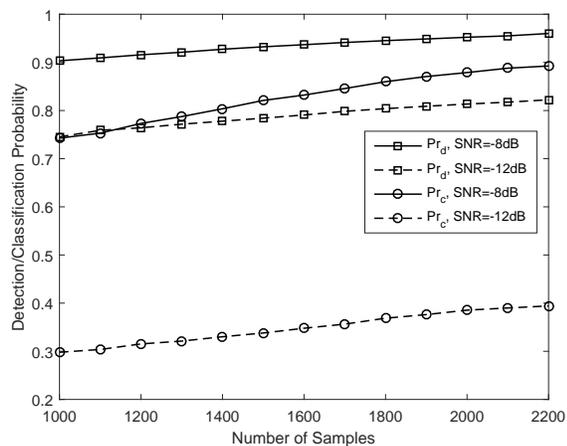
The impact of the noise correlation on the detection probability and classification probability are shown in Fig. 6(a) and Fig. 6(b). In the simulation, we set the number of antennas M at cognitive small cell base station to be 4. The number of samples N are set to 100 in evaluating OSC algorithm and 1000 for GSC algorithm, respectively. It can be seen that the detection and classification probabilities of both the proposed



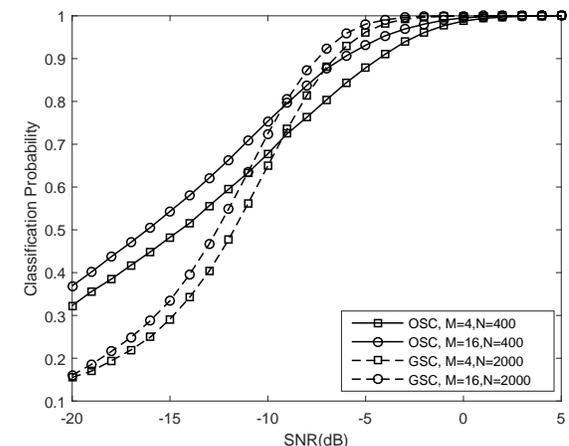
(a) Detection and classification probability for the OSC algorithm



(a) The impact of number of antennas on detection probability



(b) Detection and classification probability for the GSC algorithm



(b) The impact of number of antennas on classification probability

Fig. 4. Detection and classification probability vs number of samples with $M = 4$ and $\rho = 0.2$.

Fig. 5. The impact of number of antennas on the proposed OSC and GSC algorithms with $\rho = 0.2$.

algorithms improve with the increase of noise correlation coefficient. This is because that to de-correlate the spatially correlated noise will create a linear transformation on the received signal. The effect of this linear transformation is to introduce different weight values on each antenna. Moreover, the weight values is related to noise correlation coefficient, where a higher correlation coefficient results in a higher weight value.

E. Comparison Between OSC and GSC Algorithms

The average number of samples required to achieve a desired detection and classification probability of the OSC and GSC algorithms are shown in Table I and Table II. In this simulation, we set the number of antennas M at the cognitive small cell base station to be 4 and the noise correlation coefficient ρ to be 0.2. The average SNR is chosen as -12dB for evaluating the detection performance and -10dB for evaluating the classification performance, respectively. From Table I and Table II, it can be seen that, to achieve a same desired detection probability or classification probability, less samples are required for the OSC algorithm compared to the GSC

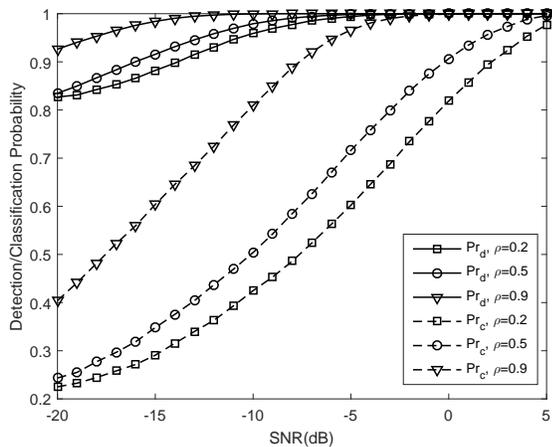
TABLE I
THE AVERAGE NUMBER OF SAMPLES REQUIRED TO ACHIEVE THE DESIRED DETECTION PROBABILITY

Pr_d	80%	85%	90%	92%	95%	97%
OSC algorithm	12	25	65	85	100	165
GSC algorithm	1360	1780	4570	6250	9800	12500

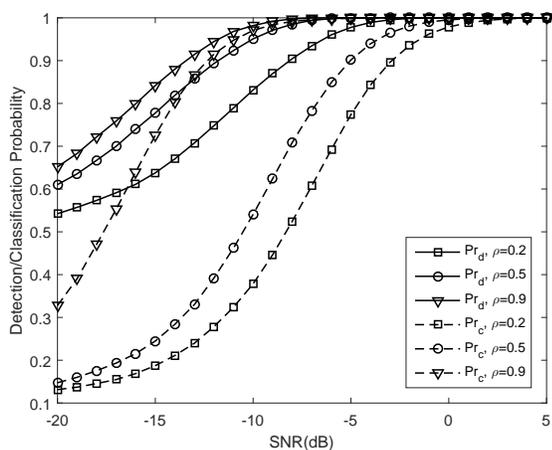
algorithm. This indicates that by leveraging the priori knowledge about the signaling features of the transmitted signal, the cognitive small cell can more fleetly and effectively detect the occupation status of the macrocell channel and classify the transmission power of MBS. In other words, the energy consumption to perform the channel states classification can be dramatically reduced, such that more energy can be reserved to transmit data for the cognitive small cell, if the total available energy is limited.

VI. CONCLUSION AND FUTURE WORKS

In this paper, we have investigated the channel states classification in cognitive small cell networks, considering multiple transmission powers. With the priori known signaling features,



(a) The impact of noise correlation on the OSC algorithm



(b) The impact of noise correlation on the GSC algorithm

Fig. 6. The impact of noise correlation on the proposed OSC and GSC algorithms.

TABLE II
THE AVERAGE NUMBER OF SAMPLES REQUIRED TO ACHIEVE THE DESIRED CLASSIFICATION PROBABILITY

Pr_c	50%	60%	70%	80%	90%	95%
OSC algorithm	200	350	610	880	1640	2350
GSC algorithm	1200	1880	3460	5570	9200	15400

we have proposed a coherent classification algorithm (i.e., OSC algorithm). The OSC algorithm can achieve a desired detection and classification performance with relatively less samples, which is crucial to shorten the spectrum sensing period. Moreover, for the unknown signaling features, we have proposed a non-coherent classification algorithm (i.e., GSC algorithm) which is much simple to implement, due to the lower hardware requirements. We have derived the optimal sensing threshold as well as closed-form decision thresholds for analyzing the detection and classification performance. With the proposed algorithms, the cognitive small cell can classify the exact transmission power of MBS when it has multiple different transmission powers, which is of significance to improve the spectrum efficiency and mitigate the inter-cell

interference.

For the future work, we will investigate cooperative channel states classification for the cognitive small cell networks, where cognitive small cell base station and cognitive mobile users cooperatively detect the occupation of macrocell channel and classify the transmission power on the occupied channel.

APPENDIX A PROOF OF LEMMA 2

When the case $\Pr(\mathcal{H}_i) = \Pr(\mathcal{H}_j)$ is considered, $\Xi(i, j)$ can be written as

$$\Xi(i, j) = N \ln \left[\frac{\det(P_i \mathbf{h} \mathbf{h}^H + \sigma_w^2 \mathbf{\Lambda})}{\det(P_j \mathbf{h} \mathbf{h}^H + \sigma_w^2 \mathbf{\Lambda})} \right] \frac{(\sigma_w^2 + P_i \mu)(\sigma_w^2 + P_j \mu)}{P_i - P_j} \quad (61)$$

Using Sylvester's determinant theorem [33] with a more general case, for \mathbf{A} , an $m \times n$ matrix, \mathbf{B} , an $n \times m$ matrix, and any invertible $m \times m$ matrix \mathbf{Z} , there is

$$\det(\mathbf{Z} + \mathbf{A} \mathbf{B}) = \det(\mathbf{Z}) \det(\mathbf{I}_n + \mathbf{B} \mathbf{Z}^{-1} \mathbf{A}). \quad (62)$$

then, we have

$$\det(P_i \mathbf{h} \mathbf{h}^H + \sigma_w^2 \mathbf{\Lambda}) = (\sigma_w^2)^M \det(\mathbf{\Lambda}) \left(1 + \frac{P_i}{\sigma_w^2} \mathbf{h}^H \mathbf{\Lambda}^{-1} \mathbf{h}\right). \quad (63)$$

Therefore, $\Xi(i, j)$ can be further simplified as

$$\Xi(i, j) = N \ln \left[\frac{\sigma_w^2 + P_i \mu}{\sigma_w^2 + P_j \mu} \right] \frac{(\sigma_w^2 + P_i \mu)(\sigma_w^2 + P_j \mu)}{P_i - P_j}. \quad (64)$$

Define $\xi = \frac{\sigma_w^2 + P_i \mu}{\sigma_w^2 + P_j \mu} = \frac{\alpha}{\beta}$, $\xi \in (0, 1) \cup (1, \infty)$, we have

$$\Xi(i, j) = N \ln(\xi) \frac{\alpha \mu}{\xi - 1}. \quad (65)$$

The partial derivative of $\Xi(i, j)$ over P_j can be computed as

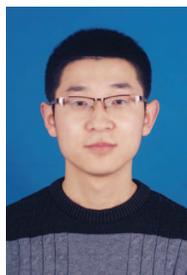
$$\frac{\partial \Xi(i, j)}{\partial P_j} = \frac{\partial \Xi(i, j)}{\partial \xi} \cdot \frac{\partial \xi}{\partial P_j} = N \mu^2 \frac{\ln(\xi) + 1/\xi - 1}{\xi^2 (\xi - 1)^2}. \quad (66)$$

It can be verified that $\ln(\xi) + 1/\xi - 1 > 0$ for $\xi \in (0, 1) \cup (1, +\infty)$. Hence, $\Xi(i, j)$ is an increasing function over P_j for any given $P_i, \forall i, j \in 1, 2, \dots, L$.

REFERENCES

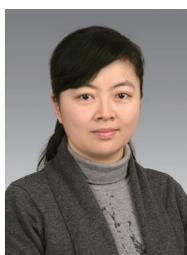
- [1] N. Zhang, S. Zhang, S. Wu, J. Ren, J. W. Mark, and X. Shen, "Beyond coexistence: Traffic steering in lte networks with unlicensed bands," *IEEE Wireless Commun.*, vol. 23, no. 6, pp. 40–46, Dec. 2016.
- [2] U. Siddique, H. Tabassum, E. Hossain, and D. I. Kim, "Wireless backhauling of 5G small cells: challenges and solution approaches," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 22–31, Oct. 2015.
- [3] C. Jiang, H. Zhang, Y. Ren, and H. H. Chen, "Energy-efficient non-cooperative cognitive radio networks: micro, meso, and macro views," *IEEE Commun. Mag.*, vol. 52, no. 7, pp. 14–20, July 2014.
- [4] L. Huang, G. Zhu, and X. Du, "Cognitive femtocell networks: an opportunistic spectrum access for future indoor wireless coverage," *IEEE Wireless Commun.*, vol. 20, no. 2, pp. 44–51, Apr. 2013.
- [5] X. Hong, J. Wang, C. X. Wang, and J. Shi, "Cognitive radio in 5G: a perspective on energy-spectral efficiency trade-off," *IEEE Commun. Mag.*, vol. 52, no. 7, pp. 46–53, July 2014.
- [6] L. Yang, S. H. Song, and K. B. Letaief, "Optimal overlay cognitive spectrum access with F-ALOHA in macro-femto heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1323–1335, Feb. 2016.
- [7] 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, Secondquarter 2015.

- [8] J. Zou, H. Xiong, D. Wang, and C. W. Chen, "Optimal power allocation for hybrid overlay/underlay spectrum sharing in multiband cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 4, pp. 1827–1837, May 2013.
- [9] X. Jiang, K. K. Wong, Y. Zhang, and D. Edwards, "On hybrid overlay-underlay dynamic spectrum access: Double-threshold energy detection and markov model," *IEEE Trans. Veh. Technol.*, vol. 62, no. 8, pp. 4078–4083, Oct. 2013.
- [10] F. Jasbi and D. K. C. So, "Hybrid overlay/underlay cognitive radio network with MC-CDMA," *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 2038–2047, Apr. 2016.
- [11] S. Zhang, N. Zhang, S. Zhou, J. Gong, Z. Niu, and X. Shen, "Energy-aware traffic offloading for green heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1116–1129, May 2016.
- [12] P. Mach and Z. Becvar, "Energy-aware dynamic selection of overlay and underlay spectrum sharing for cognitive small cells," *IEEE Trans. Veh. Technol.*, vol. 66, no. 5, pp. 4120–4132, May 2017.
- [13] B. Ma, M. H. Cheung, V. W. S. Wong, and J. Huang, "Hybrid overlay/underlay cognitive femtocell networks: A game theoretic approach," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3259–3270, June 2015.
- [14] Q. Xu, X. Li, H. Ji, and X. Du, "Energy-efficient resource allocation for heterogeneous services in ofdma downlink networks: Systematic perspective," *IEEE Trans. Veh. Technol.*, vol. 63, no. 5, pp. 2071–2082, June 2014.
- [15] H. Zhang, S. Chen, X. Li, H. Ji, and X. Du, "Interference management for heterogeneous networks with spectral efficiency improvement," *IEEE Wireless Commun.*, vol. 22, no. 2, pp. 101–107, Apr. 2015.
- [16] "User equipment (ue) radio transmission and reception," 3GPP, Tech. Rep. v.10.3.0, Release 10, June 2011.
- [17] "Requirements for further advancements for evolved universal terrestrial radio access (e-utra) (lte-advanced)," 3GPP, Tech. Rep. v.10.0.0, Mar. 2011.
- [18] S. K. Sharma, S. Chatzinotas, and B. Ottersten, "Eigenvalue-based sensing and snr estimation for cognitive radio in presence of noise correlation," *IEEE Trans. Veh. Technol.*, vol. 62, no. 8, pp. 3671–3684, Oct. 2013.
- [19] P. Urriza, E. Rebeiz, and D. Cabric, "Multiple antenna cyclostationary spectrum sensing based on the cyclic correlation significance test," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 11, pp. 2185–2195, Nov. 2013.
- [20] H. Xie, F. Gao, S. Zhang, and S. Jin, "A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3170–3184, Apr. 2017.
- [21] T. Li and A. Nehorai, "Maximum likelihood direction finding in spatially colored noise fields using sparse sensor arrays," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 1048–1062, Mar. 2011.
- [22] M. Viberg, P. Stoica, and B. Ottersten, "Maximum likelihood array processing in spatially correlated noise fields using parameterized signals," *IEEE Trans. Signal Process.*, vol. 45, no. 4, pp. 996–1004, Apr. 1997.
- [23] F. Digham, M. S. Alouini, and M. K. Simon, "On the energy detection of unknown signals over fading channels," *IEEE Trans. Commun.*, vol. 55, no. 1, pp. 21–24, Jan. 2007.
- [24] H. Shokri-Ghadikolaei and C. Fischione, "Analysis and optimization of random sensing order in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 5, pp. 803–819, May 2015.
- [25] A. Taherpour, M. Nasiri-Kenari, and S. Gazor, "Multiple antenna spectrum sensing in cognitive radios," *IEEE Trans. Wireless Commun.*, vol. 9, no. 2, pp. 814–823, Feb. 2010.
- [26] R. Zhang, J. L. Teng, Y.-C. Liang, and Y. Zeng, "Multi-antenna based spectrum sensing for cognitive radios: a glrt approach," *IEEE Trans. Commun.*, vol. 58, no. 1, pp. 84–88, Jan. 2010.
- [27] P. Wang, J. Fang, N. Han, and H. Li, "Multiantenna-assisted spectrum sensing for cognitive radio," *IEEE Trans. Veh. Technol.*, vol. 59, no. 4, pp. 1791–1800, May 2010.
- [28] Y. Liu, G. Wang, M. Xiao, and Z. Zhong, "Spectrum sensing and throughput analysis for cognitive two-way relay networks with multiple transmit powers," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 11, pp. 3038–3047, Nov. 2016.
- [29] Z. Li, D. Wang, P. Qi, and B. Hao, "Maximum-eigenvalue-based sensing and power recognition for multiantenna cognitive radio system," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8218–8229, Oct. 2016.
- [30] N. Zhang, N. Cheng, N. Lu, H. Zhou, J. W. Mark, and X. Shen, "Risk-aware cooperative spectrum access for multi-channel cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 3, pp. 516–527, Mar. 2014.
- [31] S. L. Loyka, "Channel capacity of mimo architecture using the exponential correlation matrix," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 369–371, Sep. 2001.
- [32] S. Chatzinotas, M. A. Imran, and R. Hoshydar, "On the multicell processing capacity of the cellular mimo uplink channel in correlated rayleigh fading environment," *IEEE Trans. on Wireless Commun.*, vol. 8, no. 7, pp. 3704–3715, July 2009.
- [33] D. S. Bernstein, *Matrix mathematics: theory, facts, and formulas*. Princeton University Press, 2009.
- [34] H. Xie, B. Wang, F. Gao, and S. Jin, "A full-space spectrum-sharing strategy for massive mimo cognitive radio systems," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 10, pp. 2537–2549, Oct. 2016.
- [35] C. Liu, H. Li, J. Wang, and M. Jin, "Optimal eigenvalue weighting detections for multi-antenna cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2083–2096, Apr. 2016.
- [36] A. Celik and A. E. Kamal, "Green cooperative spectrum sensing and scheduling in heterogeneous cognitive radio networks," *IEEE Trans. Cogn. Commun. Netw.*, vol. 2, no. 3, pp. 238–248, Sep. 2016.
- [37] W. W. Hager, "Updating the inverse of a matrix," *SIAM review*, vol. 31, no. 2, pp. 221–239, 1989.



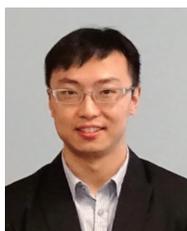
Danyang Wang received the B.S. degree and Ph.D degree in Telecommunications Engineering School from Xidian University, Xi'an, China, in 2012 and 2017, respectively. He is now a postdoctoral research fellow at the State Key Laboratory of Integrated Service Networks, School of Telecommunications Engineering, Xidian University, Xi'an. From September 2016 to August 2017, he visited the Broadband Communications Research (BBCR) Group, Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada.

His research interests are in the area of cognitive radio networks, cooperative spectrum sensing, and non-orthogonal multiple access(NOMA).



Zan Li received the B.S. degree in Communications Engineering from Xidian University, Xi'an, China, in 1998, and the M.S. and Ph.D. degrees in Communication & Information Systems from Xidian University, Xi'an, China, in 2001 and in 2004, respectively. She is currently a Professor of the State Key Laboratory of Integrated Services Networks (ISN) (Xidian University).

Her research interests focus on topics in wireless communications and signal processing, including weak signal detection, spectrum sensing, and cooperative communication.



Ning Zhang (IEEE M15) received his Ph.D degree from the University of Waterloo in 2015. He is now an Assistant Professor in the Department of Computing Science at Texas A&M University-Corpus Christi. Before that, he was a postdoctoral research fellow first at the University of Waterloo and then at the University of Toronto. He was the co-recipient of the Best Paper Awards at IEEE GLOBECOM 2014 and IEEE WCSP 2015.

His current research interests include next generation wireless networks, software defined networking, vehicular networks, and physical layer security.



Huici Wu received her B.S. degree in Information Engineering School from Communication University of China, Beijing, China, in 2013. She is currently pursuing Ph.D. degree with the communications and information systems at Beijing University of Posts and Telecommunications, Beijing, China. From September 2016 to August 2017, she visited the Broadband Communications Research (BBCR) Group, Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada.

Her research interests are in the area of wireless communications and networks, with current emphasis on the CoMP and physical layer security in heterogeneous networks.



Xuemin (Sherman) Shen (IEEE M97-SM02-F09) received the B.Sc.(1982) degree from Dalian Maritime University (China) and the M.Sc. (1987) and Ph.D. degrees (1990) from Rutgers University, New Jersey (USA), all in electrical engineering. He is a University Professor and the Associate Chair for Graduate Studies, Department of Electrical and Computer Engineering, University of Waterloo, Canada. Dr. Shens research focuses on resource management, wireless network security, social networks, smart grid, and vehicular ad hoc and sensor

networks. Dr. Shen served as the Technical Program Committee Chair/Co-Chair for IEEE Globecom16, Infocom14, IEEE VTC10 Fall, and Globecom07, the Symposia Chair for IEEE ICC10, the Tutorial Chair for IEEE VTC'11 Spring and IEEE ICC08, the General Co-Chair for ACM Mobihoc15, Chinacom07 and the Chair for IEEE Communications Society Technical Committee on Wireless Communications. He also serves/served as the Editor-in-Chief for IEEE Internet of Things Journal, IEEE Network, Peer-to-Peer Networking and Application, and IET Communications; a Founding Area Editor for IEEE Transactions on Wireless Communications; an Associate Editor for IEEE Transactions on Vehicular Technology, Computer Networks, and ACM/Wireless Networks, etc.; and the Guest Editor for IEEE JSAC, IEEE Wireless Communications, and IEEE Communications Magazine, etc. Dr. Shen received the Excellent Graduate Supervision Award in 2006, and the Premiers Research Excellence Award (PREA) in 2003 from the Province of Ontario, Canada. He is a registered Professional Engineer of Ontario, Canada, an Engineering Institute of Canada Fellow, a Canadian Academy of Engineering Fellow, a Royal Society of Canada Fellow, and a Distinguished Lecturer of IEEE Vehicular Technology Society and Communications Society.