Power Allocation Robust to Time-Varying Wireless Channels in Femtocell Networks

Zhixin Liu, Jinle Wang, Yuanqing Xia, Rongfei Fan, Hai Jiang, *Senior Member, IEEE,* and Hongjiu Yang

***Abstract*—We investigate the power allocation problem in a two-tier femtocell network including a macrocell and multiple femtocells. Due to shadowing and fading effects, at each cell (macrocell or femtocell), the power levels of desired signal and interference signals vary with time. Under this circumstance, one method to achieve transmission quality in the cells is to first get channel state information of all desired links and interference links and then perform power allocation. This method has very large communication and computation overhead. In this work, we focus on power allocation in which the transmit power levels of the users in the cells do not need to change when the wireless channels fluctuate. We formulate a power allocation problem subject to bounded outage probability in any cell (macrocell or femtocell). The formulated problem has probabilistic constraints. It is hard to have closed-form expressions of the probabilistic constraints. To solve this, we propose novel transformations of the constraints, based on which we obtain constraints in the format of worst-case value-at-risk. Such constraints can be converted to convex constraints, and thus, the research problem is transformed to a convex problem. For the convex problem, we provide an iterative algorithm that converges quickly. We also investigate the case when the channel gain distributions are unknown. Our proposed schemes have the merits of very small communication and computation overhead, and are particularly useful in fast fading environments.**

***Index Terms*—Femtocell, optimization, robustness.**

I. INTRODUCTION

Femtocell networks have the advantages of extending the cellular network coverage and providing indoor users with reliable and high-data-rate wireless access. Femtocells in the coverage of a macrocell share the wireless spectrum of the macrocell. Due to the random deployment of femtocells, there may exist severe cross-tier interference between macrocell and femtocells and co-tier interference among femtocells. Therefore, to combat the interference, power allocation is a critical issue for femtocell networks, and has received a lot of research attention recently.

Manuscript received August 3, 2014; revised January 24, 2015 and March 16, 2015; accepted March 29, 2015. The associate editor coordinating the review of this paper and approving it for publication was Dr. J. Deng. This work was supported in part by the National Natural Science Foundation of China (Project numbers 61473247), by the National Basic Research Program of China (973 Program, project number 2012CB720000), and by the Natural Sciences and Engineering Research Council of Canada. Z. Liu, J. Wang, and H. Yang are with Institute of Electrical Engineering, Yanshan University, Qinhuangdao, Hebei 066004, China (e-mail: lzxauto@ysu.edu.cn, wangjinleysu@163.com, yanghongjiu@ysu.edu.cn). Y. Xia is with the School of Automation, Beijing Institute of Technology, Beijing 100081, China, and with Institute of Electrical Engineering, Yanshan University, Qinhuangdao, Hebei 066004, China (e-mail: xia_yuanqing@bit.edu.cn). R. Fan is with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: fanrongfei@bit.edu.cn). H. Jiang is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2V4, Canada (e-mail: hai1@ualberta.ca).

To achieve power allocation, traditionally the channel state information (CSI) of the desired links and interference links is obtained first, and then the transmit power levels in the femtocells and/or macrocell are managed such that the transmission quality in the cells is guaranteed. However, since it is hard (if not impossible) to get perfect CSI, robust power allocation has been receiving increasing attention, in which "robust" means the power allocation still works if there exists some level of uncertainty in CSI. The work in [1] investigates uplink power control of a two-tier femtocell network, in which only path loss attenuation is considered (without fading). Due to the possible inaccuracy in measuring the path distance of desired and interference links, it is assumed that there is a bounded error in distance estimation, and a robust power allocation problem is formulated and solved, which takes into account quality-of-service of macrocell and femtocell users and power efficiency of femtocell users. The work in [2] assumes there is an uncertainty bound for channel gain estimation of desired links and interference links in downlink of a two-tier femtocell network. Power control and beamforming problems are solved, in which the proposed algorithm has two steps. In the first step, the beamformer in the macrocell is determined, and the power allocation in macrocell and femtocells are jointly determined. In the second step, the beamformers in the femtocells are determined in a distributed manner. The work in [3] considers uplink power allocation of a femtocell network by assuming an uncertainty bound for CSI estimation. A hierarchical game is formulated and solved by iteratively performing two sub-games until convergence: the first sub-game is for power allocation in femtocells given power allocation of the macrocell's sub-game, and the second sub-game is for power allocation of the macrocell given power allocation of the femtocells' sub-game. The work in [4] considers downlink transmission in a two-tier femtocell network. The femto-base stations have imperfect estimation of the instantaneous channel gain information, and it is assumed that the difference of the estimated channel gain from the exact channel gain is a zero-mean unit-variance complex Gaussian random variable. Power allocation and beamforming are combined so as to mitigate interference among femto-users and macro-users. The effect of channel uncertainty is analyzed, based on which the transmit power level is determined to guarantee spectrum efficiency of indoor edge femto-user and the beam weight is determined from a minimum mean square error (MMSE) criterion. The work in [5] considers a single femtocell in a macrocell. Imperfect CSI is assumed (e.g., considering that the channel CSI estimation has errors, and the errors have known distributions). The power of the macrocell user is given, and the femtocell user needs to minimize its power consumption such that the quality of its transmission is guaranteed in a probabilistic sense and the interference level to the macrocell user is also bounded in a probabilistic sense. By using semidefinite relaxation (SDR) and/or approximations, the formulated problems are relaxed into convex semidefinite programming problems. The work in [6] deals with power control in an uplink code-division multiple access network. A base station collects instantaneous CSI from each user to itself. The CSI estimations have errors, and the errors for all users' channels are independent and identically distributed within a closed region. To achieve robustness to CSI uncertainty, a probabilistic constraint is adopted, i.e., the outage probability of each user is bounded. The work in [6] targets at solving the formulated minimum power consumption problem by reduced complexity. After approximations of the probabilistic constraint, the formulated problem is transformed to a second order cone programming problem or a linear programming problem, both of which are convex problems. The work in [7] investigates an underlay cognitive radio network, in which a secondary user can share the spectrum with primary users if its interference to each primary user is bounded by a pre-defined threshold. The instantaneous CSI estimation from secondary transmitters to their receivers is perfect, the estimation of instantaneous interference from primary users to secondary users is perfect, while the instantaneous CSI estimation from each secondary transmitter to primary users has bounded errors. To achieve robustness, the interference to primary users is required to be always below a given threshold when the estimation errors are within the bounded region. For two different formations of the bounded region of errors, the secondary throughput optimization problem is solved. Tradeoff between robust interference control and secondary throughput is further investigated, by keeping the probability of violating the interference requirement below a given threshold.

In these existing robust power allocation schemes, it is still required that the instantaneous CSI of the links (desired links and interference links) should be measured. In a two-tier femtocell network, femtocells are usually isolated from each other [8]. It may be costly, or even impossible, to get instantaneous CSI of interference links. And the transmit power levels in the macrocell and femtocells need to change frequently as the channels vary, resulting in high communication and computation overhead. Further, estimation of instantaneous CSI of all links and subsequent data transmissions should be done within the channel coherence time, which means that existing robust schemes may not work in a fast fading environment. To solve this problem, our target is robust power allocation with much less communication and computation overhead. Our robustness is in the sense that we do not need to measure instantaneous CSI of any link and we do not need to change transmit power levels of the users in the cells when the wireless channels vary with time (due to shadowing and fading).

The contributions of this paper are as follows. 1) The formulated problem has probabilistic constraints. It is hard to have closed-form expressions of the probabilistic constraints. To solve this, we propose novel transformations of the probabilistic constraints, based on which we obtain constraints in the format of worst-case value-at-risk. Such constraints can be converted to convex constraints, and thus, the research problem is transformed to a convex problem. 2) We prove that the convex problem satisfies some special features, and thus, can be solved by an iterative algorithm that converges quickly. 3) We further investigate the research problem in a more practical setting when the channel gain distributions are unknown. 4) Our solutions are robust to channel fluctuations, with very small communication and computation overhead.

The rest of the paper is organized as follows. The system model is given, and the research problem is formulated and solved in Section II for the cases when the channel gain distributions are known and unknown. The performance of our solutions are evaluated by simulations in Section III, followed by conclusions in Section IV.

II. Problem Formulation, Transformations, and Iterative Algorithm

A. System Model

Consider a two-tier femtocell network, which includes a macrocell and a number, N, of femtocells inside the coverage region of the macrocell. The radius of the macrocell or a femtocell is R_m or R_f, respectively, with a macrocell base station (BS) or a femtocell BS located at the center. Uplink transmissions in the macrocell and femtocells are considered (note that downlink transmissions can be treated similarly). Consider a target frequency band, which is used by a user in each of the $N+1$ cells. For presentation simplicity, the macrocell is called cell 0, and the N femtocells are called cell $1, 2, ..., N$, respectively. In cell i ($i = 0, 1, ..., N$), the user that uses the target frequency band and the BS are called user i and BS i, respectively. Denote $\mathcal{N} \stackrel{\triangle}{=} \{0, 1, 2, ..., N\}$.

Denote $g_{i,j}$, $i, j \in \mathcal{N}$, as the channel gain from user j to BS i, which includes path loss attenuation with pass loss exponent α, log-normal distributed shadowing (i.e., the logarithm [to base 10] of the shadowing follows normal distribution with mean zero and variance σ^2), and Rayleigh fading.

B. Optimization Problem and Transformations

In the two-tier network considered, users in the macrocell and femtocells jointly allocate their transmit power such that the communication quality in each cell is guaranteed. For the communication of user i ($\in \mathcal{N}$) to its BS, the signal to interference plus noise ratio (SINR) is given as

$$\gamma_i = \frac{p_i g_{i,i}}{\sum_{j=0, j\neq i}^{N} p_j g_{i,j} + \delta^2} \tag{1}$$

where p_l is the transmit power of user $l \in \mathcal{N}$, and δ^2 is the variance of background noise.

The target SINR in cell i is given as Γ_i. To provide a certain level of quality-of-service in each cell, the outage probability (i.e., the probability that γ_i is less than Γ_i) should be bounded by $\varepsilon_i \in (0, 1)$, given as

$$\Pr\left(\gamma_i \geq \Gamma_i\right) \geq 1 - \varepsilon_i \tag{2}$$

where $\Pr(\cdot)$ means probability of an event. The expression of γ_i in (1) involves multiple random variables $g_{i,j}$'s ($i, j \in \mathcal{N}$), which makes it difficult to get a closed-form expression of $\Pr(\gamma_i \geq \Gamma_i)$. Even if such a closed-form expression is available, it is very unlikely for constraint (2) to be a convex constraint, which makes it difficult to solve the research problem subject to constraint (2). To address this challenge, we have the following mathematical manipulations to transform constraint (2).

Let

$$\mathbf{y}_i = -\Gamma_i \left[p_0, p_1, \cdots, p_{i-1}, 0, p_{i+1}, \cdots, p_N, \delta^2\right]^T$$

in which superscript T denotes transpose operation. Define random vector

$$\mathbf{r}_i = \left[\frac{g_{i,0}}{g_{i,i}}, \frac{g_{i,1}}{g_{i,i}}, \cdots, \frac{g_{i,i-1}}{g_{i,i}}, 0, \frac{g_{i,i+1}}{g_{i,i}}, \cdots, \frac{g_{i,N}}{g_{i,i}}, \frac{1}{g_{i,i}}\right]^T.$$

Then (2) can be rewritten as

$$\Pr\left(-\mathbf{y}_i^T \mathbf{r}_i \leq p_i\right) \geq 1 - \varepsilon_i. \tag{3}$$

Our objective is to find power allocation of $p_0, p_1, ..., p_N$ in the cells such that the total power consumption, given as $\sum_{i=0}^{N} p_i$, is minimized subject to constraint (3) for all cells, in which channel gain $g_{i,j}$, $i, j \in \mathcal{N}$, fluctuates (including path loss, shadowing, and fading) as described in Section II-A.

To solve the above problem, the major challenge lies in the time-varying fluctuations of channel gains $g_{i,j}$'s. A possible solution is as follows, referred to as an *ideal scheme* with perfect CSI: for each fading block in which the channel gains keep unchanged, a central controller (e.g., the macrocell BS) collects instantaneous values of $g_{i,j}$'s for all $i, j \in \mathcal{N}$, and performs power allocation to minimize the total power consumption such that the achieved SINR in each cell is not less than the target value; then the central controller sends the power allocation results to all the cells via a control channel. The major drawback of the ideal scheme is that the communication overhead (for collecting instantaneous CSI of all desired links and interference links) and computational overhead (for solving the power allocation problem for each fading block) are both very large. Therefore, it is desired to have power allocation with much less communication and computation overhead.

Accordingly, we select to satisfy constraint (3) in an "imaginary" more general case as follows. For constraint (3), denote the mean and covariance matrix of vector $\mathbf{r}_i$ as $\mathbf{m}_{\mathbf{r}_i} = \mathbb{E}(\mathbf{r}_i)$ and $\mathbf{V}_{\mathbf{r}_i} = \mathrm{Cov}(\mathbf{r}_i)$, respectively. Here $\mathbb{E}(\cdot)$ represents expectation, and $\mathrm{Cov}(\cdot)$ is the covariance matrix of a vector. Random vector $\mathbf{r}_i$ has a distribution, which fully depends on the channel gain distribution as described in Section II-A, although it is hard to have a closed-form expression of the distribution of $\mathbf{r}_i$. In this work, we consider an imaginary more general case in which constraint (3) is guaranteed when vector $\mathbf{r}_i$ has mean the same as $\mathbf{m}_{\mathbf{r}_i}$ and covariance matrix the same as $\mathbf{V}_{\mathbf{r}_i}$ but with *arbitrary distribution*, denoted as $\mathbf{r}_i \sim (\mathbf{m}_{\mathbf{r}_i}, \mathbf{V}_{\mathbf{r}_i})$. In other words, we select to satisfy

$$\inf_{\mathbf{r}_i \sim (\mathbf{m}_{\mathbf{r}_i}, \mathbf{V}_{\mathbf{r}_i})} \Pr\left(-\mathbf{y}_i^T \mathbf{r}_i \leq p_i\right) \ \geq \ 1 - \varepsilon_i. \tag{4}$$

It can be seen that the constraint in (4) has the same format of the worst-case value-at-risk (VaR) introduced in [9]. Here the term "worst-case" is in the sense that $\Pr\left(-\mathbf{y}_i^T \mathbf{r}_i \leq p_i\right) \geq 1 - \varepsilon_i$ can be guaranteed even if $\mathbf{r}_i$ has an adverse distribution in the imaginary general case.

Next we give expressions of the elements of $\mathbf{m}_{\mathbf{r}_i}$ and $\mathbf{V}_{\mathbf{r}_i}$. For presentation simplicity, the first row or column of a matrix is called row 0 or column 0, respectively, and the first element of a vector is call element 0. Denote m_j ($j \in \mathcal{N} \cup \{N+1\}$) as the jth element of $\mathbf{m}_{\mathbf{r}_i}$.[1] Denote $v_{j,k}$ ($j, k \in \mathcal{N} \cup \{N+1\}$) as the element of $\mathbf{V}_{\mathbf{r}_i}$ at the jth row and kth column.

Denote $d_{i,j}$ as the distance from user j to BS i. The following lemma for elements in $\mathbf{m}_{\mathbf{r}_i}$ is in order.

Lemma 1: Let $\omega = \sqrt{\sigma^2 + 5.57^2}$ dB; $\mu_j = 10\alpha \lg d_{i,i} - 10\alpha \lg d_{i,j}$ for $j \in \mathcal{N}, j \neq i$; and $\mu_{N+1} = 10\alpha \lg d_{i,i} + 2.5$. We have

$$\begin{aligned}
&m_j = 0, \text{ if } j = i;\\
&m_j = 10^{\frac{\mu_j}{10}} \exp\left[\left(\tfrac{\ln 10}{10}\right)^2 \omega^2\right], \text{ for } j \in \mathcal{N}, j \neq i;\\
&m_{N+1} = 10^{\frac{\mu_{N+1}}{10}} \exp\left[0.5\left(\tfrac{\ln 10}{10}\right)^2 \omega^2\right].
\end{aligned}$$

Proof: See Appendix A. ■

For $\mathbf{V}_{\mathbf{r}_i}$, we have $v_{j,k} = 0$ when $j = i$ or $k = i$. Next we give expressions of other elements in $\mathbf{V}_{\mathbf{r}_i}$. First we have the following lemma for diagonal elements.

Lemma 2: We have

$$\begin{aligned}
v_{j,j} = 10^{\frac{\mu_j}{5}} &\left(\exp\left[4\left(\tfrac{\ln 10}{10}\right)^2 \omega^2\right] - \exp\left[2\left(\tfrac{\ln 10}{10}\right)^2 \omega^2\right]\right)\\
&\text{for } j \in \{0, 1, ..., i-1, i+1, ..., N\};\\
v_{N+1,N+1} = 10^{\frac{\mu_{N+1}}{5}} &\left(\exp\left[2\left(\tfrac{\ln 10}{10}\right)^2 \omega^2\right]\right.\\
&\left. - \exp\left[\left(\tfrac{\ln 10}{10}\right)^2 \omega^2\right]\right)
\end{aligned}$$

Proof: See Appendix C. ■

Next we give expressions of non-diagonal elements in $\mathbf{V}_{\mathbf{r}_i}$. First, we give the expression of $\bar{g}_{i,j}$ that is the mean value of $g_{i,j}$. Similar to the derivation of m_j ($j \in \mathcal{N}, j \neq i$) shown in proof of Lemma 1, for $j \neq i$, we have

$$\bar{g}_{i,j} = 10^{\frac{-2.5 - 10\alpha \lg d_{i,j}}{10}} \exp\left[0.5\left(\frac{\ln 10}{10}\right)^2 \omega^2\right].$$

Then for $j, k \in \{0, 1, ..., i-1, i+1, ..., N\}$ and $j \neq k$, we have

$$\begin{aligned}
v_{j,k} &= v_{k,j} = \mathbb{E}\left(\tfrac{g_{i,j}}{g_{i,i}} \cdot \tfrac{g_{i,k}}{g_{i,i}}\right) - \mathbb{E}\left(\tfrac{g_{i,j}}{g_{i,i}}\right)\mathbb{E}\left(\tfrac{g_{i,k}}{g_{i,i}}\right)\\
&= \mathbb{E}(g_{i,j})\mathbb{E}(g_{i,k})\mathbb{E}\left(\tfrac{1}{g_{i,i}^2}\right) - \mathbb{E}(g_{i,j})\mathbb{E}(g_{i,k})\left[\mathbb{E}\left(\tfrac{1}{g_{i,i}}\right)\right]^2\\
&= \bar{g}_{i,j}\bar{g}_{i,k} v_{N+1,N+1}.
\end{aligned}$$

And similarly, we have

$$v_{j,N+1} = v_{N+1,j} = \bar{g}_{i,j} v_{N+1,N+1}.$$

Now we have expressions of elements in $\mathbf{m}_{\mathbf{r}_i}$ and $\mathbf{V}_{\mathbf{r}_i}$. According to [9], constraint (4) can be converted to

$$\sqrt{\frac{1-\varepsilon_i}{\varepsilon_i}} \sqrt{\mathbf{y}_i^T \mathbf{V}_{\mathbf{r}_i} \mathbf{y}_i} - \mathbf{y}_i^T \mathbf{m}_{\mathbf{r}_i} \leq p_i. \tag{5}$$

[1] Note that for presentation simplicity, here we omit subscript 'i' from notation m_j and some other subsequent notations.

Therefore, the power allocation problem of the cells can be formulated as

$$\min \sum_{i=0}^{N} p_i \quad \text{s.t.} \begin{cases} \text{(i)} & \text{Constraint } (5),\ i \in \mathcal{N} \\ \text{(ii)} & 0 \le p_i \le p_{i,\max},\ i \in \mathcal{N} \end{cases} \tag{6}$$

where $p_{i,\max}$ is the maximum allowable transmit power of user i.

We take the solution of (6) as the power allocation for the considered two-tier femtocell network. The solution is conservative. However, since we do not have instantaneous CSI in (6), we do not need to collect instantaneous CSI. This also means that the solution of (6) is *robust* since we do not need to change the transmit power levels in the cells when the channels fluctuate with time. Overall, the major advantages of the solution of (6) are the very small communication and computation overhead, and the robustness to channel fluctuations.

C. The Case when the Mean and Covariance Matrix of $\mathbf{r}_i$ Need to be Estimated

In the power allocation problem in (6), we should know $\mathbf{m}_{\mathbf{r}_i}$ and $\mathbf{V}_{\mathbf{r}_i}$, the mean and covariance matrix of vector $\mathbf{r}_i$. When the channel gain distributions as described in Section II-A are known, $\mathbf{m}_{\mathbf{r}_i}$ and $\mathbf{V}_{\mathbf{r}_i}$ can be derived, as we have done in Section II-B. However, it is possible that the channel gain distributions are not known, and thus, $\mathbf{m}_{\mathbf{r}_i}$ and $\mathbf{V}_{\mathbf{r}_i}$ need to be estimated based on sampled data. In this subsection, we consider this case, and give a new transformation of the constraint in (4).

Suppose we have L samples for $\mathbf{r}_i$,[2] denoted as $\mathbf{r}_{i,1}, \mathbf{r}_{i,2}, ..., \mathbf{r}_{i,L}$. The estimated mean of $\mathbf{r}_i$ can be written as

$$\hat{\mathbf{m}}_{\mathbf{r}_i} = \frac{1}{L}\sum_{l=1}^{L} \mathbf{r}_{i,l}$$

and the unbiased estimation of covariance matrix of $\mathbf{r}_i$ can be written as

$$\hat{\mathbf{V}}_{\mathbf{r}_i} = \frac{1}{L-1}\sum_{l=1}^{L} \left(\mathbf{r}_{i,l} - \hat{\mathbf{m}}_{\mathbf{r}_i}\right) \times \left(\mathbf{r}_{i,l} - \hat{\mathbf{m}}_{\mathbf{r}_i}\right)^T.$$

Then the focus is on how to transform the constraint in (4) given $\hat{\mathbf{m}}_{\mathbf{r}_i}$ and $\hat{\mathbf{V}}_{\mathbf{r}_i}$, $i \in \mathcal{N}$.

Define a new random variable $z_i \triangleq -\mathbf{y}_i^T \mathbf{r}_i$. Accordingly, z_i has L samples, denoted as $z_{i,l} = -\mathbf{y}_i^T \mathbf{r}_{i,l}, l = 1, 2, ..., L$. We denote the real mean and variance of z_i as m_{z_i} and v_{z_i}, respectively, and denote the estimated mean and estimated variance of z_i as $\hat{\mathrm{m}}_{z_i} = (1/L)\sum_{l=1}^{L} z_{i,l}$ and $\hat{\mathrm{v}}_{z_i} = \big(1/(L-1)\big)\sum_{l=1}^{L}\left(z_{i,l} - \hat{\mathrm{m}}_{z_i}\right)^2$, respectively. The estimated mean and variance of $z_i, i \in \mathcal{N}$ can be obtained as:

$$\begin{aligned} \hat{\mathrm{m}}_{z_i} &= \tfrac{1}{L}\textstyle\sum_{l=1}^{L} z_{i,l} = \tfrac{1}{L}\sum_{l=1}^{L}\left(-\mathbf{y}_i^T \mathbf{r}_{i,l}\right) \\ &= -\mathbf{y}_i^T \left(\tfrac{1}{L}\textstyle\sum_{l=1}^{L} \mathbf{r}_{i,l}\right) = -\mathbf{y}_i^T \hat{\mathbf{m}}_{\mathbf{r}_i}, \end{aligned} \tag{7}$$

[2]Samples of $\mathbf{r}_i$ can be obtained from the sampled $g_{i,j}$'s, $j \in \mathcal{N}$.

$$\begin{aligned} \hat{\mathrm{v}}_{z_i} &= \tfrac{1}{L-1}\textstyle\sum_{l=1}^{L}\left(z_{i,l} - \hat{\mathrm{m}}_{z_i}\right)^2 \\ &= \tfrac{1}{L-1}\textstyle\sum_{l=1}^{L}\left(-\mathbf{y}_i^T \mathbf{r}_{i,l} + \mathbf{y}_i^T \hat{\mathbf{m}}_{\mathbf{r}_i}\right)^2 \\ &= \tfrac{1}{L-1}\textstyle\sum_{l=1}^{L}\left(-\mathbf{y}_i^T \mathbf{r}_{i,l} + \mathbf{y}_i^T \hat{\mathbf{m}}_{\mathbf{r}_i}\right)\left(\mathbf{r}_{i,l}^T\left(-\mathbf{y}_i\right) + \hat{\mathbf{m}}_{\mathbf{r}_i}^T \mathbf{y}_i\right) \\ &= \left(-\mathbf{y}_i^T\right)\left(\tfrac{1}{L-1}\textstyle\sum_{l=1}^{L}(\mathbf{r}_{i,l} - \hat{\mathbf{m}}_{\mathbf{r}_i})\left(\mathbf{r}_{i,l} - \hat{\mathbf{m}}_{\mathbf{r}_i}\right)^T\right)\left(-\mathbf{y}_i\right) \\ &= \mathbf{y}_i^T \hat{\mathbf{V}}_{\mathbf{r}_i} \mathbf{y}_i. \end{aligned} \tag{8}$$

Thus, constraint (4) given $\hat{\mathbf{m}}_{\mathbf{r}_i}$ and $\hat{\mathbf{V}}_{\mathbf{r}_i}$ is equivalent to the following constraint

$$\inf_{z_i \sim\ (\mathrm{m}_{z_i}, \mathrm{v}_{z_i})} \Pr\left(z_i \le p_i\right) \ \ge\ 1 - \varepsilon_i \tag{9}$$

given $\hat{\mathrm{m}}_{z_i}$ and $\hat{\mathrm{v}}_{z_i}$ as shown in (7) and (8), respectively.

Expression (9) can be rewritten as

$$\sup_{z_i \sim\ (\mathrm{m}_{z_i}, \mathrm{v}_{z_i})} \Pr\left(\frac{z_i - \hat{\mathrm{m}}_{z_i}}{\sqrt{\hat{\mathrm{v}}_{z_i}}} \ge \frac{p_i - \hat{\mathrm{m}}_{z_i}}{\sqrt{\hat{\mathrm{v}}_{z_i}}}\right) \ \le\ \varepsilon_i. \tag{10}$$

Further define new random variable

$$u_i = \sqrt{\frac{L}{L-1}}\frac{\left(z_i - \hat{\mathrm{m}}_{z_i}\right)}{\sqrt{\hat{\mathrm{v}}_{z_i}}}, i \in \mathcal{N}$$

and define

$$k_i = \sqrt{\frac{L}{L-1}}\frac{\left(p_i - \hat{\mathrm{m}}_{z_i}\right)}{\sqrt{\hat{\mathrm{v}}_{z_i}}}, i \in \mathcal{N}.$$

For the region of k_i, we have the following lemma.

Lemma 3: With ε_i close to 0 and L large enough, we should have $k_i > 1$, $i \in \mathcal{N}$.

Proof: We use proof by contradiction.

We first prove that $k_i < 0$ leads to a contradiction. Suppose $k_i < 0$, i.e., $p_i < \hat{\mathrm{m}}_{z_i}$, then

$$\begin{aligned} &\sup\nolimits_{z_i \sim\ (\mathrm{m}_{z_i}, \mathrm{v}_{z_i})} \Pr\left(\tfrac{z_i - \hat{\mathrm{m}}_{z_i}}{\sqrt{\hat{\mathrm{v}}_{z_i}}} \ge \tfrac{p_i - \hat{\mathrm{m}}_{z_i}}{\sqrt{\hat{\mathrm{v}}_{z_i}}}\right) \\ \ge\ &\sup\nolimits_{z_i \sim\ (\mathrm{m}_{z_i}, \mathrm{v}_{z_i})} \Pr\left(\tfrac{z_i - \hat{\mathrm{m}}_{z_i}}{\sqrt{\hat{\mathrm{v}}_{z_i}}} \ge \tfrac{\hat{\mathrm{m}}_{z_i} - \hat{\mathrm{m}}_{z_i}}{\sqrt{\hat{\mathrm{v}}_{z_i}}}\right) \\ =\ &\sup\nolimits_{z_i \sim\ (\mathrm{m}_{z_i}, \mathrm{v}_{z_i})} \Pr\left(z_i \ge \hat{\mathrm{m}}_{z_i}\right) \\ \ge\ &\sup\nolimits_{z_i \sim\ (\mathrm{m}_{z_i}, \mathrm{v}_{z_i}),\ z_i \text{ is symmetrically distributed}} \Pr\left(z_i \ge \hat{\mathrm{m}}_{z_i}\right). \end{aligned} \tag{11}$$

When the random variable z_i is symmetrically distributed, we have

$$\sup_{z_i \sim\ (\mathrm{m}_{z_i}, \mathrm{v}_{z_i}),\ z_i \text{ is symmetrically distributed}} \Pr\left(z_i \ge \mathrm{m}_{z_i}\right) = \frac{1}{2}. \tag{12}$$

When L is large enough, $\hat{\mathrm{m}}_{z_i}$ approaches to m_{z_i}, and therefore, there exists a small ϵ such that

$$\begin{aligned} &\sup_{z_i \sim\ (\mathrm{m}_{z_i}, \mathrm{v}_{z_i}),\ z_i \text{ is symmetrically distributed}} \Pr\left(z_i \ge \hat{\mathrm{m}}_{z_i}\right) \\ \ge &\sup_{z_i \sim\ (\mathrm{m}_{z_i}, \mathrm{v}_{z_i}),\ z_i \text{ is symmetrically distributed}} \Pr\left(z_i \ge \mathrm{m}_{z_i}\right) - \epsilon. \end{aligned} \tag{13}$$

From (11), (12), and (13), it can be seen that

$$\sup_{z_i \sim\ (\mathrm{m}_{z_i}, \mathrm{v}_{z_i})} \Pr\left(\frac{z_i - \hat{\mathrm{m}}_{z_i}}{\sqrt{\hat{\mathrm{v}}_{z_i}}} \ge \frac{p_i - \hat{\mathrm{m}}_{z_i}}{\sqrt{\hat{\mathrm{v}}_{z_i}}}\right) \ge \frac{1}{2} - \epsilon$$

which contradicts (10) when ε_i is close to 0.

Next we prove that $0 \le k_i \le 1$ also leads to a contradiction.

Suppose $0 \leq k_i \leq 1, i \in \mathcal{N}$. By the definition of k_i and (7) and (8), we have

$$-\mathbf{y}_i^T \hat{\mathbf{m}}_{\mathbf{r}_i} \leq p_i \leq \sqrt{\frac{L-1}{L}} \sqrt{\mathbf{y}_i^T \hat{\mathbf{V}}_{\mathbf{r}_i} \mathbf{y}_i} - \mathbf{y}_i^T \hat{\mathbf{m}}_{\mathbf{r}_i}. \quad (14)$$

When L is large, $\hat{\mathbf{m}}_{\mathbf{r}_i}$ and $\hat{\mathbf{V}}_{\mathbf{r}_i}$ can almost approximate $\mathbf{m}_{\mathbf{r}_i}$ and $\mathbf{V}_{\mathbf{r}_i}$ perfectly. Replace $\hat{\mathbf{m}}_{\mathbf{r}_i}$ and $\hat{\mathbf{V}}_{\mathbf{r}_i}$ with $\mathbf{m}_{\mathbf{r}_i}$ and $\mathbf{V}_{\mathbf{r}_i}$ in (14), we get

$$-\mathbf{y}_i^T \mathbf{m}_{\mathbf{r}_i} \leq p_i \leq \sqrt{\frac{L-1}{L}} \sqrt{\mathbf{y}_i^T \mathbf{V}_{\mathbf{r}_i} \mathbf{y}_i} - \mathbf{y}_i^T \mathbf{m}_{\mathbf{r}_i}. \quad (15)$$

When $\hat{\mathbf{m}}_{\mathbf{r}_i}$ and $\hat{\mathbf{V}}_{\mathbf{r}_i}$ can almost approximate $\mathbf{m}_{\mathbf{r}_i}$ and $\mathbf{V}_{\mathbf{r}_i}$ perfectly, the constraint (5) should also hold. From (15) and (5), it can be seen that

$$0 \leq \sqrt{\frac{1-\varepsilon_i}{\varepsilon_i}} \leq \sqrt{\frac{L-1}{L}} \quad (16)$$

which leads to

$$\frac{1}{2} \leq \frac{1}{2-\frac{1}{L}} \leq \varepsilon_i \leq 1. \quad (17)$$

The inequality in (17) contradicts the fact that ε_i is close to zero.

As a summary, $k_i < 0$ and $0 \leq k_i \leq 1$ both lead to contradiction. Therefore, we should have $k_i > 1$. This completes the proof. ∎

Denote the sampled sequence of random variable u_i as

$$u_{i,l} = \sqrt{\frac{L}{L-1}} \frac{(z_{i,l} - \hat{\mathrm{m}}_{z_i})}{\sqrt{\hat{\mathrm{v}}_{z_i}}}, l = 1, 2, ..., L$$

then it can be seen that

$$\sum_{l=1}^{L} u_{i,l} = 0 \quad (18)$$

and

$$\sum_{l=1}^{L} u_{i,l}^2 = L. \quad (19)$$

According to [10], for the sequence of $u_{i,l}, l = 1, 2, ..., L$ satisfying the equations (18) and (19), for $k_i > 1$, we have

$$\Pr\left(u_i^2 \geq k_i^2\right) \leq \frac{1}{L} \left\lfloor \left(\frac{L}{k_i^2}\right) \right\rfloor$$

in which $\lfloor \cdot \rfloor$ means floor function. So

$$\begin{aligned} \Pr\left(u_i \geq k_i\right) &\leq \Pr\left(u_i^2 \geq k_i^2\right) \\ &\leq \tfrac{1}{L} \left\lfloor \left(\tfrac{L}{k_i^2}\right) \right\rfloor \leq \tfrac{1}{L} \left(\tfrac{L}{k_i^2}\right) = \tfrac{1}{k_i^2}. \end{aligned} \quad (20)$$

Recall that constraint (4) is equivalent to constraint (10), which is further equivalent to $\sup_{z_i \sim (\mathrm{m}_{z_i}, \mathrm{v}_{z_i})} \Pr\left(u_i \geq k_i\right) \leq \varepsilon_i$. From (20), if $1/k_i^2 \leq \varepsilon_i$, then constraint (10) is satisfied, and thus, constraint (4) is satisfied. Therefore, we try to satisfy $1/k_i^2 \leq \varepsilon_i$, which is equivalent to (based on definition of k_i and (7) and (8))

$$\sqrt{\frac{L-1}{L\varepsilon_i}} \sqrt{\mathbf{y}_i^T \hat{\mathbf{V}}_{\mathbf{r}_i} \mathbf{y}_i} - \mathbf{y}_i^T \hat{\mathbf{m}}_{\mathbf{r}_i} \leq p_i. \quad (21)$$

So the power allocation problem with sampled data $\hat{\mathbf{m}}_{\mathbf{r}_i}$ and $\hat{\mathbf{V}}_{\mathbf{r}_i}$, $i \in \mathcal{N}$ can be given as

$$\begin{aligned} &\min \sum_{i=0}^{N} p_i \\ &\text{s.t.} \begin{cases} \text{(i)} & \text{Constraint } (21), \ i \in \mathcal{N} \\ \text{(ii)} & 0 \leq p_i \leq p_{i,\max}, \ i \in \mathcal{N}. \end{cases} \end{aligned} \quad (22)$$

Problem (22) has the same structure as problem (6). And constraint (21) has similar format to constraint (5). Therefore, similar to the solution of problem (6), the solution of problem (22) is also robust, and has the advantages of very small communication and computation overhead.

D. *Iterative Algorithm*

The problem (6) and problem (22) are convex problems, and can be solved by a Lagrangian decomposition method. Since it is possible that problem (6) and problem (22) are not feasible (i.e., it might be impossible that all constraints are satisfied), the Lagrangian decomposition method may converge to a point that violates the constraints [11]. If this happens, call admission control is needed to remove some femtocell users.

As an alternative solution, next we give an iterative algorithm to solve problem (6), and perform call admission control if necessary. Since problem (22) has a similar structure to problem (6), it can be treated similarly.

It is easy to see, by Lagrangian method, that, if problem (6) is feasible, the solution satisfies

$$p_i = \min\left\{\sqrt{\frac{1-\varepsilon_i}{\varepsilon_i}} \cdot \sqrt{\mathbf{y}_i^T \mathbf{V}_{\mathbf{r}_i} \mathbf{y}_i} - \mathbf{y}_i^T \mathbf{m}_{\mathbf{r}_i}, p_{i,\max}\right\}$$

for all $i \in \mathcal{N}$.

Define vector $\mathbf{p} = [p_0, p_1, \cdots, p_N]$, $\mathbf{p}_{\max} = [p_{0,\max}, p_{1,\max} \cdots, p_{N,\max}]$, and $\mathbf{I}(\mathbf{p}) = [I_0(\mathbf{p}), I_1(\mathbf{p}) \cdots, I_N(\mathbf{p})]$, where

$$I_i(\mathbf{p}) = \sqrt{\frac{1-\varepsilon_i}{\varepsilon_i}} \cdot \sqrt{\mathbf{y}_i^T \mathbf{V}_{\mathbf{r}_i} \mathbf{y}_i} - \mathbf{y}_i^T \mathbf{m}_{\mathbf{r}_i}, \ i \in \mathcal{N}. \quad (23)$$

Then if problem (6) is feasible, its solution satisfies $\mathbf{p} = \mathbf{A}(\mathbf{p}) \overset{\triangle}{=} \min\{\mathbf{I}(\mathbf{p}), \mathbf{p}_{\max}\}$. Here $\min(\mathbf{a}, \mathbf{b})$ means a vector whose kth element is the minimum one of the kth elements of vectors **a** and **b**. We use $\mathbf{a} \succeq \mathbf{b}$ (or $\mathbf{a} \succ \mathbf{b}$) to represent that each element in vector **a** is not less than (or larger than) the corresponding element in vector **b**.

We consider the following iterative algorithm:

$$\mathbf{p}(t+1) = \mathbf{A}(\mathbf{p}(t))$$

in which $\mathbf{p}(t)$ is $\mathbf{p}$ in the tth iteration. In other words, at each iteration $t = 1, 2, \cdots$, user i updates its transmit power level by $p_i(t+1) = A_i(\mathbf{p}(t))$, i.e.,

$$p_i(t+1) = \min\left\{I_i(\mathbf{p}(t)), p_{i,\max}\right\}. \quad (24)$$

Here $A_i(\mathbf{p}(t))$ is the ith element of $\mathbf{A}(\mathbf{p}(t))$.

Next, we show that the iterative algorithm converges to optimal solution of problem (6) if problem (6) is feasible.

Definition [12]: a function **f(p)** is *standard* if it has three properties: 1) positivity, which means $\mathbf{f(p)} \succ \mathbf{0}$ (here **0** is a vector with all elements being 0); 2) monotonicity, which means $\mathbf{f}(\mathbf{p}^\dagger) \succeq \mathbf{f}(\mathbf{p})$ if $\mathbf{p}^\dagger \succeq \mathbf{p}$; 3) scalability, which means $\varrho\mathbf{f}(\mathbf{p}) \succ \mathbf{f}(\varrho\mathbf{p})$ for $\varrho > 1$.

Next we check the three properties for functions **I(p)** and $\mathbf{p}_{\max}$.

For **I(p)**:

- Positivity: All elements of $-\mathbf{y}_i$ and $\mathbf{m}_{\mathbf{r}_i}$ are non-negative, and the last element of them are both positive. So $-\mathbf{y}_i^T\mathbf{m}_{\mathbf{r}_i} > 0$. We also have $\sqrt{\mathbf{y}_i^T\mathbf{V}_{\mathbf{r}_i}\mathbf{y}_i} = \sqrt{(-\mathbf{y}_i)^T\mathbf{V}_{\mathbf{r}_i}(-\mathbf{y}_i)} > 0$, because all elements of $-\mathbf{y}_i$ are non-negative with the last element being positive, and all elements of $\mathbf{V}_{\mathbf{r}_i}$ are non-negative with the element on the last row and last column being positive. Thus, we have
$$I_i(\mathbf{p}) = \sqrt{\frac{1-\varepsilon_i}{\varepsilon_i}} \cdot \sqrt{\mathbf{y}_i^T\mathbf{V}_{\mathbf{r}_i}\mathbf{y}_i} - \mathbf{y}_i^T\mathbf{m}_{\mathbf{r}_i} > 0.$$
- Monotonicity: it is not difficult to find out that $I_i(\mathbf{p})$ is an increasing function of p_j, $j \neq i$. Therefore, for $\mathbf{p}^\dagger \succeq \mathbf{p}$, we have $I_i(\mathbf{p}^\dagger) \geq I_i(\mathbf{p})$.
- Scalability: For $\varrho > 1$, $I_i(\varrho\mathbf{p})$ can be expressed as
$$I_i(\varrho\mathbf{p}) = \sqrt{\frac{1-\varepsilon_i}{\varepsilon_i}} \cdot \sqrt{\mathbf{w}_i^T\mathbf{V}_{\mathbf{r}_i}\mathbf{w}_i} + \mathbf{w}_i^T\mathbf{m}_{\mathbf{r}_i} \tag{25}$$
in which
$$\mathbf{w}_i = \Gamma_i\left[\varrho p_0, \varrho p_1, \cdots, \varrho p_{i-1}, 0, \varrho p_{i+1}, \cdots, \varrho p_N, \delta^2\right]^T.$$
Define
$$\mathbf{x}_i = \Gamma_i\left[\varrho p_0, \varrho p_1, \cdots, \varrho p_{i-1}, 0, \varrho p_{i+1}, \cdots, \varrho p_N, \varrho\delta^2\right]^T.$$
All elements in $\mathbf{w}_i$ and $\mathbf{x}_i$ are non-negative, and the last elements of them are positive. The only difference between $\mathbf{w}_i$ and $\mathbf{x}_i$ is that the last element of $\mathbf{x}_i$ is larger. Since all elements in $\mathbf{m}_{\mathbf{r}_i}$ are non-negative, and its last element is positive, we have $\mathbf{w}_i^T\mathbf{m}_{\mathbf{r}_i} < \mathbf{x}_i^T\mathbf{m}_{\mathbf{r}_i}$. Since all elements of $\mathbf{V}_{\mathbf{r}_i}$ are non-negative, and its element on the last row and last column is positive, we have $\mathbf{w}_i^T\mathbf{V}_{\mathbf{r}_i}\mathbf{w}_i < \mathbf{x}_i^T\mathbf{V}_{\mathbf{r}_i}\mathbf{x}_i$. Together with (25), we have
$$\begin{aligned} I_i(\varrho\mathbf{p}) &< \sqrt{\tfrac{1-\varepsilon_i}{\varepsilon_i}} \cdot \sqrt{\mathbf{x}_i^T\mathbf{V}_{\mathbf{r}_i}\mathbf{x}_i} + \mathbf{x}_i^T\mathbf{m}_{\mathbf{r}_i} \\ &\overset{(a)}{=} \sqrt{\tfrac{1-\varepsilon_i}{\varepsilon_i}} \cdot \sqrt{(\varrho\mathbf{y}_i)^T\mathbf{V}_{\mathbf{r}_i}(\varrho\mathbf{y}_i)} - \varrho\mathbf{y}_i^T\mathbf{m}_{\mathbf{r}_i} \\ &= \varrho\left(\sqrt{\tfrac{1-\varepsilon_i}{\varepsilon_i}} \cdot \sqrt{\mathbf{y}_i^T\mathbf{V}_{\mathbf{r}_i}\mathbf{y}_i} - \mathbf{y}_i^T\mathbf{m}_{\mathbf{r}_i}\right) \\ &\overset{(b)}{=} \varrho I_i(\mathbf{p}) \end{aligned} \tag{26}$$
in which (a) is due to the fact $\mathbf{x}_i = -\varrho\mathbf{y}_i$, and (b) is due to expression of $I_i(\mathbf{p})$ given in (23). So **I(p)** has the property of scalability.

For $\mathbf{p}_{\max}$, define $\mathbf{H}(\mathbf{p}) \triangleq \mathbf{p}_{\max}$. Here $\mathbf{H}(\mathbf{p})$ is a constant vector, and is always equal to $\mathbf{p}_{\max}$ for whatever $\mathbf{p}$. Apparently $\mathbf{H}(\mathbf{p})$ has the properties of positivity and monotonicity. For $\varrho > 1$, since $\mathbf{H}(\mathbf{p})$ is a constant vector, we have $\mathbf{H}(\varrho\mathbf{p}) = \mathbf{p}_{\max}$. So we have $\varrho\mathbf{H}(\mathbf{p}) = \varrho\mathbf{p}_{\max} = \varrho\mathbf{H}(\varrho\mathbf{p}) \succ \mathbf{H}(\varrho\mathbf{p})$. So $\mathbf{H}(\mathbf{p})$ has the property of scalability.

Therefore, functions $\mathbf{I}(\mathbf{p})$ and $\mathbf{H}(\mathbf{p}) \triangleq \mathbf{p}_{\max}$ are both standard. And apparently, function $\mathbf{A}(\mathbf{p}) = \min\{\mathbf{I}(\mathbf{p}), \mathbf{p}_{\max}\}$ is also standard. Since $\mathbf{A}(\mathbf{p})$ is standard and problem (6) is convex, it can be concluded [12] that iterative algorithm (24), i.e., $p_i(t+1) = A_i(\mathbf{p}(t)) \triangleq \min\{I_i(\mathbf{p}(t)), p_{i,\max}\}$, converges to the optimal solution of problem (6) if problem (6) is feasible.

Consider that the iterative algorithm (24) converges to a convergence point $\tilde{\mathbf{p}} = [\tilde{p}_0, \tilde{p}_1, ..., \tilde{p}_N]$. $\tilde{p}_i$ should be equal to either $I_i(\tilde{\mathbf{p}})$ or $p_{i,\max}$, whichever is less. If $\tilde{p}_i = I_i(\tilde{\mathbf{p}})$ for all $i \in \mathcal{N}$, which also implies $\tilde{p}_i \leq p_{i,\max}$, it can be seen that all constraints of problem (6) are satisfied, since $I_i(\tilde{\mathbf{p}})$ is the left handside of constraint (5) when the users use power levels $\tilde{p}_0, \tilde{p}_1, ..., \tilde{p}_N$. However, for a cell, say cell i, if we have $\tilde{p}_i = p_{i,\max} < I_i(\tilde{\mathbf{p}})$ (i.e., $\tilde{p}_i$ converges to $p_{i,\max}$), then we say *power saturation* happens at cell i, and call user i a *saturated user*. In this situation, $\tilde{p}_i < I_i(\tilde{\mathbf{p}})$ means that at the convergence point, constraint (5) is not satisfied for cell i, which implies that $p_{i,\max}$ is not sufficiently large compared with interference. Since constraint (5) is transformed from constraint (2), power saturation at cell i means that user i's quality-of-service cannot be guaranteed, i.e., the outage probability of user i cannot be bounded below ε_i. So the problem (6) is infeasible, and thus, the system cannot accommodate all the users simultaneously. Therefore, call admission control should be performed to remove some users from the system such that the optimization problem (6) with the remaining users becomes feasible.

For call admission control, we have the following observations.

- If all saturated users are removed from the system, then problem (6) with the remaining users becomes feasible. The reason is as follows. For a non-saturated user (i.e., a user that is not a saturated user), say user i, constraint (5) is satisfied at the convergence point, that is
$$I_i([\tilde{p}_0, \tilde{p}_1, ..., \tilde{p}_N]) \leq \tilde{p}_i. \tag{27}$$
If all saturated users are removed, which is equivalent to setting the power levels of the saturated users in (27) to be zero, the inequality (27) still holds, since $I_i([\tilde{p}_0, \tilde{p}_1, ..., \tilde{p}_N])$ is an increasing function of $\tilde{p}_j$ ($j = 0, 1, 2, ..., N, j \neq i$). Therefore, if we only keep the non-saturated users, then constraint (5) is still satisfied for all the non-saturated users. Thus, problem (6) with only the non-saturated users is feasible.
- It is possible that problem (6) becomes feasible if part of the saturated users are removed. We use the following example to illustrate. Assume there are two saturated users, called user 1 and user 2. Consider that user 1 is removed and all other users keep using power levels at convergence point $\tilde{\mathbf{p}}$. For the constraint (5) of user 2 (which was originally not satisfied), removal of user 1 makes the left handside of (5) become smaller, and thus, it is possible that constraint (5) of user 2 becomes satisfied. In other words, it may not be necessary to remove all saturated users to make problem (6) feasible.

Based on the above observations, when there is only one saturated user and the user is a femtocell user, then we only need to remove the saturated user, which makes problem (6) become feasible. When there are multiple saturated users and all of them are femtocell users, we can remove the saturated

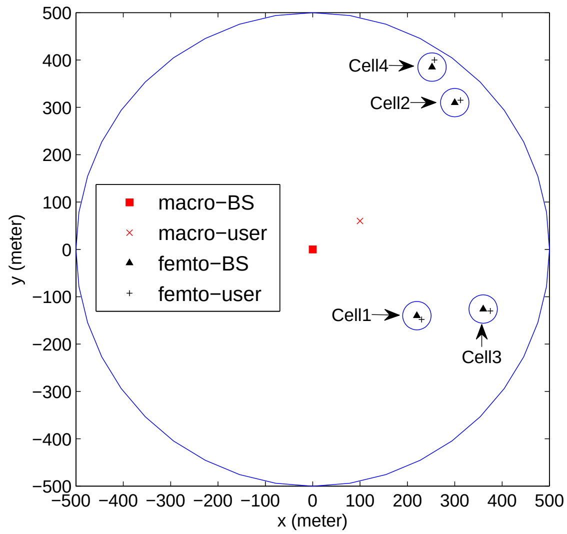

Fig. 1. The simulated two-tier femtocell network.

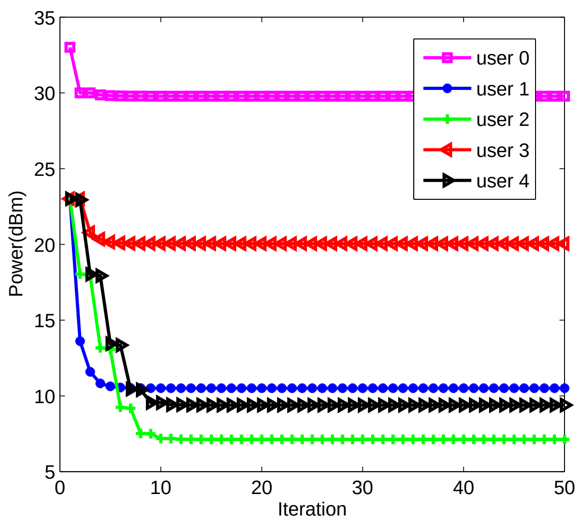

Fig. 2. Convergence performance of the iterative algorithm (24).

users one after another until problem (6) becomes feasible (i.e., power saturation does not happen at any remaining cell). On the other hand, if the macrocell user is a saturated user, it may not be good to remove the macrocell user, since the macrocell user should have a higher priority. Then we can remove the femtocell users one after another, based on descending order of their interference level to the macrocell user, until problem (6) becomes feasible.

Once the power allocation and/or the call admission decision are made, the results are sent to all the cells via, for example, the control channel of the macrocell.

III. Performance Evaluation

Here we present simulation results. Consider one macrocell and four femtocells as shown in Fig. 1. The positions of the five BSs are (0,0), (220, -140), (300, 310), (360, -126), (252, 400) (unit: meter). The positions of the five users are: $(100, 60), (225, -143), (304, 315), (378, -128)$, $(257, 407)$. $R_m = 500$m, $R_f = 30$m, $p_{0,\max} = 33$ dBm (2 W), $p_{i,\max} = 23$ dBm (200 mW), $i = 1, 2, 3, 4$. $\delta^2 = 10^{-9}$, $\Gamma_0 = 5$dB, $\Gamma_i = 15$ dB, $i = 1, 2, 3, 4$. $\alpha = 4$, $\sigma = 6.5$dB. $\varepsilon_i = \varepsilon$, $i = 0, 1, 2, 3, 4$.

For our proposed scheme with known channel gain distributions, Fig. 2 shows the transmit power levels p_i's of the iterative algorithm in (24) when $\varepsilon = 0.2$. The initial power level of cell i is $p_{i,\max}$. It can be seen that the iterative algorithm converges within approximately 10 iterations.

We simulate the performance of our proposed scheme with known channel gain distributions and our proposed scheme without channel gain distribution information but using sampled data (30 samples are used for channel gain of each link). We change the value of ε from 0.1 to 0.4. For each specific ε value, we first use the iterative algorithm in (24) to calculate the power allocation of the users, then with the power allocation, we run Monte-Carlo simulation using MATLAB to simulate the average outage probabilities of the users when the channels (desired channels and interference channels) vary as described in Section II-A. As a comparison, we also simulate the ideal scheme introduced in Section II-B and a naive scheme. The naive scheme has the average channel gain information of each link. By assuming the real channel gains take the values of the average channel gains, the naive scheme derives the transmit power levels of all the users such that the SINR of each user is not less than the corresponding target value. Then the derived transmit power levels are used in the simulation in which the channels vary as described in Section II-A.

The simulation results are shown in Tables I and II. Table I shows the outage probabilities of the five users in our proposed scheme with known channel gain distributions (represented by "proposed (known distributions)"), our proposed scheme with sampled data (represented by "proposed (sampled data)"), and the naive scheme. The ideal scheme does not have outage. It can be seen that for the proposed schemes with known channel gain distributions and with sampled data, the outage probability of each user is bounded by the threshold value ε, which demonstrates the robustness of the power allocation to time varying wireless channels. Since the naive scheme is based only on average channel gain information, the outage probability of each user is large. Table II shows the power consumption of the four schemes. It is clear that the naive scheme has the least power consumption since it is based only on average channel gain information. The ideal scheme has more power consumption than the naive scheme, but with less power consumption than the proposed schemes.

Next we compare the communication overhead and computation overhead of our proposed schemes with the ideal scheme.

Communication overhead: For the ideal scheme, for each fading block, the instantaneous channel gains of all desired links and interference links need to be measured, and the decision of power allocation needs to be delivered to all users. This can be achieved by $2(N + 1) + 1$ message exchanges.

- The first $N + 1$ messages are as follows: Each of the

	user 0	user 1	user 2	user 3	user 4
Proposed (known distributions)	{0.001 0.003 0.006 0.008}	{0.002 0.004 0.007 0.008}	{0.002 0.003 0.004 0.008}	{0.002 0.004 0.009 0.013}	{0.002 0.005 0.006 0.014}
Proposed (sampled data)	{0.001 0.002 0.005 0.006}	{0.002 0.003 0.006 0.008}	{0.003 0.003 0.006 0.01}	{0.002 0.004 0.011 0.014}	{0.003 0.004 0.005 0.012}
Naive	0.597	0.623	0.605	0.631	0.625

TABLE I

SIMULATED OUTAGE PROBABILITIES OF THE USERS IN DIFFERENT SCHEMES WHEN ε={0.1, 0.2, 0.3, 0.4}.

	user 0	user 1	user 2	user 3	user 4
Proposed (known distributions)	{31.666 29.784 28.641 27.746}	{14.336 10.505 8.102 6.249}	{12.200 7.131 4.788 3.115}	{23.010 20.053 17.932 16.352}	{14.560 9.395 7.138 5.604}
Proposed (sampled data)	{32.852 30.172 29.114 28.275}	{15.961 12.043 9.650 7.771}	{10.355 6.007 4.133 2.757}	{23.010 19.804 17.794 16.231}	{13.569 10.857 8.585 6.076}
Naive	7.003	-22.019	-20.805	-7.260	-16.960
Ideal	17.374	-3.079	-2.761	4.607	-2.195

TABLE II

SIMULATED POWER CONSUMPTION (UNIT: DBM) OF THE USERS IN DIFFERENT SCHEMES WHEN ε={0.1, 0.2, 0.3, 0.4}.

$N+1$ users sends a pilot message in turn, and based on reception of the pilot messages, each BS measures the channel gains from the sending users to itself.

- The subsequent $N+1$ messages are as follows: the $N+1$ BSs send their measured instantaneous channel gain information to the central controller in turn.
- The last message is as follows: Based on collected instantaneous channel gain information, the central controller makes a decision on the transmit power levels of the users and announces the decision.

Therefore, the communication overhead is $2(N+1)+1$ message exchanges per fading block. For the proposed scheme with known channel gain distributions, when a user joins the system, the user sends to the central controller a message containing information of its location and its BS's location, and the central controller runs the iterative algorithm to decide on the transmit power levels of the users and announces a message containing the power allocation decision. So the communication overhead is two message exchanges per user joining the system. For our proposed scheme with sampled data, we need $2(N+1)+1$ message exchanges per user joining the system. The first $N+1$ message exchanges are the $N+1$ users' pilot messages. For each pilot message, each BS takes a number of samples. When the $N+1$ pilot messages are completed, each BS, say BS i, can estimate the mean and covariance matrix of $\mathbf{r}_i$. Then all the $N+1$ BSs send their estimation to the central controller in turn (totally $N+1$ messages). The central controller runs the iterative algorithm, and announces the power allocation decision to all users (one message). So the communication overhead is $2(N+1)+1$ message exchanges per user joining the system. For the proposed schemes with known channel gain distributions and with sampled data, once a user joins the system, we do not need message exchanges for the fading blocks.

Computation overhead: The ideal scheme needs to solve a power allocation problem per fading block. The power allocation is a convex problem. The proposed schemes with known channel gain distributions and with sampled data need to solve problem (6) and problem (22), respectively, per user joining the system. Both problems are convex and can be solved by using the iterative algorithm (24). Once a user joins the system, we do not need computation for the fading blocks.

Overall, it can be seen that the proposed schemes have much less communication overhead and much less computation overhead, and are particularly useful in fast fading environments.

IV. CONCLUSION

In this paper, by proposing novel transformations, we have solved the power allocation problem in a two-tier femtocell network, and the solution is robust to time-varying channels in macrocell and femtocells. An iterative algorithm has been presented to find the power allocation. We have also investigated the case when the channel gain distributions are not known, and derived robust power allocation solution. Simulation results demonstrate that our schemes have strong robustness to channel fluctuations, with very small communication and computation overhead.

APPENDIX

A. Proof of Lemma 1

It is apparent that $m_j = 0$ when $j = i$. Next we focus on the expression of m_j when $j \in \mathcal{N},\ j \neq i$.

The probability density function of $g_{i,j}$ $(i, j \in \mathcal{N})$ can be approximated as [13], [14]

$$f(g_{i,j}) = \frac{10/\ln 10}{g_{i,j}\sqrt{2\pi\omega^2}} \times \exp\left[-\frac{(10\lg g_{i,j}+2.5+10\alpha\lg d_{i,j})^2}{2\omega^2}\right] \quad (28)$$

where 'lg' means logarithm to base 10, and $\omega^2=\sigma^2+5.57^2$. Define $\tau_j \triangleq g_{i,j}/g_{i,i}$, $j\in\mathcal{N}, j\neq i$. The probability density function of τ_j can be derived as (the proof is given in Appendix B)

$$f(\tau_j)=\frac{10/\ln 10}{\tau_j\sqrt{2\pi\kappa^2}}\cdot\exp\left[-\frac{(10\lg\tau_j-\mu_j)^2}{2\kappa^2}\right] \tag{29}$$

with $\mu_j=10\alpha\lg d_{i,i}-10\alpha\lg d_{i,j}$ and $\kappa^2=2\omega^2$.

Based on (29), for $j\in\mathcal{N}, j\neq i$, we have

$$\begin{aligned} m_j &= \mathbb{E}(\tau_j)=\int_0^{+\infty}\tau_j f(\tau_j)d\tau_j\\ &= \int_0^{+\infty}\frac{10/\ln 10}{\sqrt{2\pi\kappa^2}}\cdot\exp\left[-\frac{(10\lg\tau_j-\mu_j)^2}{2\kappa^2}\right]d\tau_j\\ &\overset{(c)}{=} \frac{1}{\sqrt{2\pi\kappa^2}}\cdot 10^{\frac{\mu_j}{10}}\cdot\int_{-\infty}^{+\infty}10^{\frac{x_j}{10}}\exp\left[-\frac{x_j^2}{2\kappa^2}\right]dx_j\\ &\overset{(d)}{=} \exp\left[\frac{\left(\kappa^2\frac{\ln 10}{10}\right)^2}{2\kappa^2}\right]\cdot 10^{\frac{\mu_j}{10}}\\ &\quad\times\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi\kappa^2}}\exp\left[-\frac{\left[x_j-\kappa^2\frac{\ln 10}{10}\right]^2}{2\kappa^2}\right]dx_j\\ &\overset{(e)}{=} 10^{\frac{\mu_j}{10}}\exp\left[\left(\frac{\ln 10}{10}\right)^2\omega^2\right] \end{aligned} \tag{30}$$

in which $\mathbb{E}(\cdot)$ means expectation, (c) comes from defining $x_j=10\lg\tau_j-\mu_j$, (d) comes from

$$10^{\frac{x_j}{10}}=\exp\left[\frac{\ln 10}{10}x_j\right]$$

and (e) comes from

$$\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi\kappa^2}}\exp\left[-\frac{\left[x_j-\kappa^2\frac{\ln 10}{10}\right]^2}{2\kappa^2}\right]dx_j=1$$

and $\kappa^2=2\omega^2$.

Similarly, we can get the expression of m_{N+1}.

B. Derivation of (29)

For presentation simplicity, denote $g_{i,j}$, $g_{i,i}$, and τ_j as X, Y, and Z, respectively. So we have $Z=X/Y$. From (28), we can have the probability density function of X and Y as

$$f_X(x)=\frac{10/\ln 10}{x\sqrt{2\pi\omega^2}}\cdot\exp\left[-\frac{(10\lg x-\lambda_j)^2}{2\omega^2}\right] \tag{31}$$

$$f_Y(y)=\frac{10/\ln 10}{y\sqrt{2\pi\omega^2}}\cdot\exp\left[-\frac{(10\lg y-\lambda_i)^2}{2\omega^2}\right] \tag{32}$$

where $\lambda_j=-2.5-10\alpha\lg d_{i,j}$ and $\lambda_i=-2.5-10\alpha\lg d_{i,i}$. Then after some math manipulation, the probability density function of Z is given as

$$f_Z(z)=\int_0^{+\infty}yf_X(yz)f_Y(y)dy. \tag{33}$$

Replacing (31) and (32) in (33), we have

$$\begin{aligned} f_Z(z)=\frac{10/\ln 10}{z\sqrt{2\pi\omega^2}}\cdot\int_0^{+\infty}\exp\left[-\frac{(10\lg(yz)-\lambda_j)^2}{2\omega^2}\right]\\ \times\frac{10/\ln 10}{y\sqrt{2\pi\omega^2}}\cdot\exp\left[-\frac{(10\lg y-\lambda_i)^2}{2\omega^2}\right]dy. \end{aligned}$$

Let $b=10\lg y-\lambda_i$. So we have $y=10^{\frac{b+\lambda_i}{10}}$, and $dy=10^{\frac{b}{10}}\cdot 10^{\frac{\lambda_i}{10}}\cdot\frac{\ln 10}{10}\cdot db$, which lead to

$$\begin{aligned} f_Z(z) &= \tfrac{10/\ln 10}{z\sqrt{2\pi\omega^2}}\cdot\int_{-\infty}^{+\infty}\exp\left[-\tfrac{(b+10\lg z+\lambda_i-\lambda_j)^2}{2\omega^2}\right]\\ &\qquad\times\tfrac{1}{\sqrt{2\pi\omega^2}}\cdot\exp\left[-\tfrac{b^2}{2\omega^2}\right]db\\ &= \tfrac{10/\ln 10}{z\sqrt{2\pi\omega^2}}\cdot\exp\left[-\tfrac{a^2}{4\omega^2}\right]\\ &\qquad\times\int_{-\infty}^{+\infty}\tfrac{1}{\sqrt{2\pi\omega^2}}\exp\left[-\tfrac{(b+\frac{1}{2}a)^2}{\omega^2}\right]db.\\ &= \tfrac{10/\ln 10}{z\sqrt{2\pi\omega^2}}\cdot\exp\left[-\tfrac{a^2}{4\omega^2}\right]\cdot\tfrac{1}{\sqrt{2}}\\ &\qquad\times\int_{-\infty}^{+\infty}\tfrac{1}{\sqrt{2\pi\vartheta^2}}\exp\left[-\tfrac{(b-(-\frac{1}{2}a))^2}{2\vartheta^2}\right]db.\\ &= \tfrac{10/\ln 10}{z\sqrt{2\pi\omega^2}}\cdot\exp\left[-\tfrac{a^2}{4\omega^2}\right]\cdot\tfrac{1}{\sqrt{2}} \end{aligned}$$

in which the second equality comes from denoting $a=10\lg z+\lambda_i-\lambda_j$, the third equality comes from denoting $\vartheta^2=\omega^2/2$, and the last equality comes from

$$\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi\vartheta^2}}\exp\left[-\frac{(b-(-\frac{1}{2}a))^2}{2\vartheta^2}\right]db=1.$$

Applying $a=10\lg z+\lambda_i-\lambda_j$, $\lambda_i=-2.5-10\alpha\lg d_{i,i}$, and $\lambda_j=-2.5-10\alpha\lg d_{i,j}$, we get

$$f_Z(z)=\frac{10/\ln 10}{z\sqrt{2\pi\kappa^2}}\cdot\exp\left[-\frac{(10\lg z-\mu_j)^2}{2\kappa^2}\right]$$

where $\kappa^2=2\omega^2$ and $\mu_j=10\alpha\lg d_{i,i}-10\alpha\lg d_{i,j}$.

C. Proof of Lemma 2

For $j=0,1,2,i-1,i+1,\cdots,N$, we have $v_{j,j}=\mathbb{E}\left(\tau_j^2\right)-\{\mathbb{E}(\tau_j)\}^2=\mathbb{E}\left(\tau_j^2\right)-(m_j)^2$.

From (29), $\mathbb{E}\left(\tau_j^2\right)$ can be calculated as follows.

$$\begin{aligned} &\mathbb{E}\left(\tau_j^2\right)=\int_0^{+\infty}\tau_j^2 f(\tau_j)d\tau_j\\ &=\int_0^{+\infty}\tau_j\cdot\tfrac{10/\ln 10}{\sqrt{2\pi\kappa^2}}\cdot\exp\left[-\tfrac{(10\lg\tau_j-\mu_j)^2}{2\kappa^2}\right]d\tau_j\\ &\overset{(f)}{=}\tfrac{1}{\sqrt{2\pi\kappa^2}}\cdot 10^{\frac{2\mu_j}{10}}\cdot\int_{-\infty}^{+\infty}10^{\frac{2x_j}{10}}\exp\left[-\tfrac{x_j^2}{2\kappa^2}\right]dx_j\\ &\overset{(g)}{=}\tfrac{1}{\sqrt{2\pi\cdot(2\kappa)^2}}\cdot 10^{\frac{2\mu_j}{10}}\int_{-\infty}^{+\infty}10^{\frac{y_j}{10}}\exp\left[-\tfrac{(y_j)^2}{2\cdot(2\kappa)^2}\right]dy_j\\ &\overset{(h)}{=}\tfrac{1}{\sqrt{2\pi\cdot(2\kappa)^2}}\cdot 10^{\frac{2\mu_j}{10}}\int_{-\infty}^{+\infty}\exp\left[-\tfrac{(y_j)^2}{2\cdot(2\kappa)^2}+\tfrac{\ln 10}{10}y_j\right]dy_j\\ &=10^{\frac{2\mu_j}{10}}\cdot\exp\left[2\left(\tfrac{\ln 10}{10}\kappa\right)^2\right]\\ &\qquad\times\int_{-\infty}^{+\infty}\tfrac{1}{\sqrt{2\pi\cdot(2\kappa)^2}}\cdot\exp\left[-\tfrac{(y_j-4\kappa^2\frac{\ln 10}{10})^2}{2\cdot(2\kappa)^2}\right]dy_j\\ &\overset{(i)}{=}10^{\frac{\mu_j}{5}}\exp\left[4\left(\tfrac{\ln 10}{10}\right)^2\omega^2\right] \end{aligned}$$

in which (f) comes from denoting $x_j=10\lg\tau_j-\mu_j$, (g) comes from denoting $y_j=2x_j$, (h) comes from

$$10^{\frac{y_j}{10}}=\exp\left[\frac{\ln 10}{10}y_j\right]$$

and (i) comes from $\kappa^2=2\omega^2$ and

$$\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi\cdot(2\kappa)^2}}\cdot\exp\left[-\frac{(y_j-4\kappa^2\frac{\ln 10}{10})^2}{2\cdot(2\kappa)^2}\right]dy_j=1.$$

Therefore, from (30) we have

$$\begin{aligned} v_{j,j} &= \mathbb{E}\left(\tau_j^2\right) - (m_j)^2 \\ &= 10^{\frac{\mu_j}{5}} \exp\left[4\left(\frac{\ln 10}{10}\right)^2 \omega^2\right] \\ &\qquad - \left(10^{\frac{\mu_j}{10}} \exp\left[\left(\frac{\ln 10}{10}\right)^2 \omega^2\right]\right)^2 \\ &= 10^{\frac{\mu_j}{5}} \left(\exp\left[4\left(\frac{\ln 10}{10}\right)^2 \omega^2\right] - \exp\left[2\left(\frac{\ln 10}{10}\right)^2 \omega^2\right]\right). \end{aligned}$$

Similarly, we can get the expression of $v_{j,j}$ with $j = N+1$.

References

[1] X. Guan, Q. Han, K. Ma, and X. Wang, “Robust uplink power control for co-channel two-tier femtocell networks,” *AEU – Int. J. Electron. Commun.*, vol. 67, no. 6, pp. 504–512, June 2013.

[2] R. Mo, T. Q. S. Quek, and R. W. Heath, “Robust beamforming and power control for two-tier femtocell networks,” in *Proc. IEEE VTC* 2011, pp. 1–5.

[3] Q. Han, B. Yang, X. Wang, K. Ma, C. Chen, and X. Guan, “Hierarchical-game-based uplink power control in femtocell networks,” *IEEE Trans. Veh. Technol.*, vol. 63, no. 6, pp. 2819–2835, July 2014.

[4] D.-C. Oh, H.-C. Lee, and Y.-H. Lee, “Power control and beamforming for femtocells in the presence of channel uncertainty,” *IEEE Trans. Veh. Technol.*, vol. 60, no. 6, pp. 2545–2554, July 2011.

[5] K.-Y. Wang, N. Jacklin, Z. Ding, and C.-Y. Chi, “Robust MISO transmit optimization under outage-based QoS constraints in two-tier heterogeneous networks,” *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, pp. 1883–1897, Apr. 2013.

[6] S. Parsaeefard, A. R. Sharafat, and M. Rasti, “Robust probabilistic distributed power allocation by chance constraint approach,” in *Proc. IEEE PIMRC 2010*, pp. 2162–2167.

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

[8] V. Chandrasekhar, J. G. Andrews, and A. Gatherer, “Femtocell networks: A survey,” *IEEE Commun. Mag.*, vol. 46, no. 9, pp. 59–67, Sept. 2008.

[9] Y.-L. Yu, Y. Li, D. Schuurmans, and C. Szepesvari, “A general projection property for distribution families,” in *23rd Annual Conference on Neural Information Processing Systems (NIPS)*, 2009, pp. 2232–2240.

[10] J. G. Saw, M. C. K. Yang, and T. C. Mo, “Chebysheve inequality with estimated mean and variance,” *The American Statistician*, vol. 38, no. 2, pp. 130–132, May 1984.

[11] I. Mitliagkas, N. D. Sidiropoulos, and A. Swami, “Joint power and admission control for ad-hoc and cognitive underlay networks: Convex approximation and distributed implementation,” *IEEE Trans. Wireless Commun.*, vol. 10, no. 12, pp. 4110–4121, Dec. 2011.

[12] R. D. Yates, “A framework for uplink power control in cellular radio systems,” *IEEE J. Sel. Areas Commun.*, vol. 13, no. 7, pp. 1341–1347, Sep. 1995.

[13] A. M. D. Turkmani, “Probability of error for M-branch macroscopic selection diversity,” *IEE Proceedings I: Communications, Speech and Vision*, vol. 139, no. 1, pp. 71–78, Feb. 1992.

[14] H. Fu and D. I. Kim, “Analysis of throughput and fairness with downlink scheduling in WCDMA networks,” *IEEE Trans. Wireless Commun.*, vol. 5, no. 8, pp. 2164–2174, Aug. 2006.