



King's Research Portal

Document Version
Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Zhang, X., Nakhai, M. R., Zheng, G., Lambotharan, S., & Ottersten, B. (in press). A Calibrated Learning Approach to Distributed Power Allocation in Small Cell Networks. *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2019)*.

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

A CALIBRATED LEARNING APPROACH TO DISTRIBUTED POWER ALLOCATION IN SMALL CELL NETWORKS

Xinruo Zhang[†] Mohammad Reza Nakhai^{*} Gan Zheng[†] Sangarapillai Lambotharan[†] Björn Ottersten[‡]

[†] Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, UK

^{*} Department of Informatics, Centre for Telecommunications Research, King's College London, UK

[‡] Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, Luxembourg

ABSTRACT

This paper studies the problem of max-min fairness power allocation in distributed small cell networks operated under the same frequency bandwidth. We introduce a calibrated learning algorithm to account for the dynamism in wireless channels and achieve max-min user fairness in the long run. Provided that the small base stations (SBSs) serve their own users simultaneously and no prior knowledge of forthcoming user demand is available at the SBSs, the aim of the proposed algorithm is to allow SBSs to gradually improve their forecast of the possible transmit power levels of other SBSs and make best decisions based on the predicted results at individual time slots. Simulation results validate that in terms of achieving max-min signal-to-interference-plus-noise ratio, the proposed distributed design outperforms two benchmark schemes and achieves a similar performance as compared to the optimal centralized design.

Index Terms— small cell networks, distributed power allocation, online learning, calibration

1. INTRODUCTION

Cellular networks have experienced an explosive increase in the number of mobile subscribers in the last decade, and the global mobile data traffic is predicted to reach 48.3 Exabytes per month by 2021 [1]. Hyper-dense deployment of small cell base stations (SBSs) underlying the existing macrocell cellular networks is considered a promising technique to meet the requirements of the mounting growth of mobile data traffic [2]. However, such network densification in cellular networks with limited licensed spectrum will result in increasing intercell interference (ICI). Cloud radio access networks that allow centralized radio resource coordination across multiple cells, have been widely investigated in the literature for ICI management and resource allocation [3]. However, the requirement of high-capacity fronthaul links to support the immense traffic demands in such a centralized implementation necessitates the distributed operation of SBSs that can be organized and coordinated autonomously. Conventional literature either models the distributed resource allocation prob-

lem as a noncooperative game and solves it via iterative processes [4–7], or uses iterative inter-base station (BS) fronthaul information exchange such as the subgradient method in [8–10] and the alternating direction method of multipliers technique in [10, 11] to schedule transmit power and manage ICI among BSs. These designs, nevertheless, assume that the channel remains invariant until the iterations are completed or convergence is achieved, which is not very practical. Furthermore, some types of noncooperative games, e.g., the Stackelberg game in [6], require the players to alternately make moves, which may not be suitable for the practical scenario where BSs perform simultaneous transmissions.

This paper focuses on the design of an intelligent distributed power allocation mechanism in small cell networks operated under the same spectrum that maximizes the averaged minimum signal-to-interference-plus-noise ratio (SINR) in the long run. A calibrated learning based distributed multi-armed bandit (MAB) approach is developed for the individual SBSs to learn during the operation, i.e., find a trade-off between exploring the environment and exploiting the current knowledge of the environment, so as to account for the dynamism in wireless channel conditions. In contrast to the existing distributed power allocation designs that iteratively exchange side information among SBSs within a channel coherence time, this paper considers a more realistic scenario, where SBSs serve their own users simultaneously and only exchange the past power information at the end of each time instance. The proposed algorithm is designed for the SBSs to progressively and distributively calibrate their anticipations of the possible transmit power levels of the other SBSs, and react with the best response to the predicted results.

2. SYSTEM MODEL AND PROBLEM FORMULATION

This paper considers a distributed downlink small cell network consists of N_b SBSs, indexed by $\mathcal{L}_b = \{1, \dots, N_b\}$, that simultaneously serve their own users, over a shared spectrum. Let the time horizon T be divided into discrete time slots and indexed as $\mathcal{T} = \{1, \dots, T\}$. Each time slot cor-

responds to a channel coherence time and the channel is assumed to vary across time slots but remains constant within each time slot. It is assumed that the individual SBS has to make its decision, i.e., selects a transmit power level from an index set $\mathcal{A} = \{1, \dots, A\}$ of discrete transmit power levels, independently, to serve its user during time slot t , $t \in \mathcal{T}$, and can communicate with each other via capacity-constrained inter-SBS communication links at the end of each time slot.

2.1. Downlink Transmission Model

Let the channel gain between SBS b , $b \in \mathcal{L}_b$ and its own user at the t -th time slot, $t \in \mathcal{T}$, be denoted as $\Psi_{bb}^{[t]}$. Also let $P_b^{[t]} \in \mathcal{A}$ be the transmit power from SBS b to its own user. The SINR for the user associated with SBS b at time slot t is given by

$$\text{SINR}_b^{[t]} = \frac{P_b^{[t]} |\Psi_{bb}^{[t]}|^2}{\sum_{b' \in \mathcal{L}_b, b' \neq b} P_{b'}^{[t]} |\Psi_{b'b}^{[t]}|^2 + \sigma_b^2}, \quad (1)$$

where the numerator of (1) denotes the desired signal for the user, the terms in the denominator of (1) denote, respectively, the ICI caused by other SBSs and the additive white Gaussian noise with noise variance of σ_b^2 .

2.2. Problem Formulation

Due to the fact that the individual SBSs are autonomous decision makers in the considered distributed small cell network, it is evident that the presence of interference, coupling with the decision of individual SBSs, raises a conflict: the individual SBSs tries to transmit at its maximal transmit power level to improve its utility, while the ICI incurred by doing so may degrade the performance of SBSs. Thus, we formulate the power allocation for distributed small cell networks as a long-term max-min user fairness problem to optimize the power allocation decisions of SBSs as well as account for the ICI and the randomness in wireless channel conditions, as

$$\max_{\{P_b^{[t]}\}} \left\{ \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \min_b (\text{SINR}_b^{[t]}) \right\}. \quad (2)$$

Since the system is operated in a distributed manner and the SBSs have to make decisions simultaneously, the key issue to be addressed for the individual SBSs is the acquisition or the reliable prediction of its opponents' decisions. In the next section, the ϵ -calibrated forecaster [12] will be introduced for each SBS to gradually improve its accuracy in predicting the opponents' decisions. It will be followed by the proposed calibrated learning algorithm that accounts for the dynamism in wireless channels and allows the individual SBSs to select the most appropriate transmit power level based on the output of the forecaster, such that the minimum SINR can be maximized in the long run.

3. CALIBRATED LEARNING ALGORITHM

3.1. ϵ -Calibrated Forecaster for the Opponents' Actions

Having a calibrated forecasting scheme not only allows the players to act with the best response to the predicted future events, but also enables multiple players to converge to a reasonable joint play in some cases [12]. The definition of ϵ -calibration can be explained as follows. Let us consider a forecaster playing a game against the environment, where each SBS has a finite set \mathcal{A} of possible outcomes. Define $\mathcal{P} = \{\mathbf{p}_1, \dots, \mathbf{p}_{N_\epsilon}\}$ as the N_ϵ set of candidate probability values over \mathcal{A} . Suppose that at each time slot t , $t \in \mathcal{T}$, a forecaster outputs its forecasts of the environment, given by one of the candidate probability values over each SBS's outcome, i.e., $\mathbf{p}_b^t = \{p_1^t, \dots, p_A^t\} \in \mathcal{P}$ for SBS b at time t , whilst the environment simultaneously selects outcomes $\{a_b^t\}_b \in \mathcal{A}$. Let us denote by $\mathcal{K} = \{1, \dots, N_\epsilon\}$ the indexes of $\{\mathbf{p}_1, \dots, \mathbf{p}_{N_\epsilon}\}$ and $\mathbf{p}_1, \dots, \mathbf{p}_{N_\epsilon}$ the center of N_ϵ balls with radius of ϵ . Given $\epsilon > 0$, the sequence of forecasts is called ϵ -calibrated if

$$\limsup_{T \rightarrow \infty} \sum_{k=1}^{N_\epsilon} \left\| \frac{1}{T} \sum_{t=1}^T \mathbf{1}_{\{k_b^t=k\}} (\mathbf{p}_k - \delta_{a_b^t}) \right\| \leq \epsilon, \quad (3)$$

where $\delta_{a_b^t} \in \{0, 1\}$ is the dirac distribution on the environment's outcomes $\{a_b^t\}_b \in \mathcal{A}$, and $\mathbf{1}_{k_b^t}$ denotes an indicator function that returns one if the prediction of the probability of power levels for a particular SBS at time t is \mathbf{p}_k and zero otherwise. Following the similar procedure as in [12], let us define the target set C as a subset of the ϵ -ball around $(\mathbf{0}, \dots, \mathbf{0})$ for the calibration norm, as

$$C = \{\underline{\mathbf{x}} = (\mathbf{x}_1, \dots, \mathbf{x}_{N_\epsilon}) \in \mathbb{R}^{AN_\epsilon} : \sum_{i=1}^{N_\epsilon} \|\mathbf{x}_i\| \leq \epsilon, \mathbf{x}_i \in \mathbb{R}^A\} \quad (4)$$

and let us define the vector-valued pay-off function as

$$\mathbf{m}(k_b^t, a_b^t) = (\mathbf{0}, \dots, \mathbf{0}, \mathbf{p}_{k_b^t} - \delta_{a_b^t}, \mathbf{0}, \dots, \mathbf{0}), k_b^t \in \mathcal{K}, a_b^t \in \mathcal{A}, \quad (5)$$

where $\mathbf{m}(k_b^t, a_b^t) \in \mathbb{R}^{AN_\epsilon}$ contains one non-zero vector element of \mathbb{R}^A located at the k_b^t -th position given by the difference between the predicted probability value $\mathbf{p}_{k_b^t}$ and the dirac distribution of the opponent's true action $\delta_{a_b^t}$, and $N_\epsilon - 1$ zero vector elements elsewhere. According to Blackwell's approachability theorem [13], the condition of ϵ -calibration in (3) is equivalent to the following statement: the closed convex set C is approachable by the vector-valued regrets, i.e.,

$$\begin{aligned} \bar{\mathbf{m}}_T &= \frac{1}{T} \sum_{t=1}^T \mathbf{m}(k_b^t, a_b^t) = \\ &= \frac{1}{T} \left(\sum_{t=1}^T \mathbf{1}_{\{k_b^t=1\}} (\mathbf{p}_1 - \delta_{a_b^t}), \dots, \sum_{t=1}^T \mathbf{1}_{\{k_b^t=N_\epsilon\}} (\mathbf{p}_{N_\epsilon} - \delta_{a_b^t}) \right), \end{aligned} \quad (6)$$

and there exists a prediction distribution φ_b^t over $\{\mathbf{p}_1, \dots, \mathbf{p}_{N_\epsilon}\}$, such that for $\forall a_b^t \in \mathcal{A}$,

$$(\bar{\mathbf{m}}_{t-1} - \prod_C(\bar{\mathbf{m}}_{t-1})) \cdot (\mathbf{m}(\varphi_b^t, a_b^t) - \prod_C(\bar{\mathbf{m}}_{t-1})) \leq 0, \quad (7)$$

where \cdot denotes the inner product and $\prod_C(\bar{\mathbf{m}}_{t-1})$ is the projection of $\bar{\mathbf{m}}_{t-1}$ in l_2 -norm onto C .

Recall from Section 2 that we denote by $\mathcal{A} = \{1, \dots, A\}$ the finite index set of discrete transmit power levels or actions of SBSs with cardinality of A in the considered scenario. Each SBS acts as a forecaster, the opponents act as the environment and the actions of the opponents can be regarded as the other SBSs transmitting and serving their own users at certain discrete power levels. The procedure of the ϵ -calibrated forecaster at each SBS to predict each of the other SBSs' actions is summarized in Algorithm 1 below.

Algorithm 1 ϵ -calibrated forecaster

- 1: **Input:** current time-slot t , ϵ^t , discount factor β , actual arm chosen by the opponent at time $t-1$, i.e., a_b^{t-1} .
 - 2: Update the averaged vector-valued regrets up to time $t-1$ as per (6), as $\bar{\mathbf{m}}_{t-1} = \frac{1}{t-1} \sum_{t'=1}^{t-1} \mathbf{m}(k_b^{t'}, a_b^{t'}) \beta^{(t-t')}$.
 - 3: Calculate the projection $\prod_C(\bar{\mathbf{m}}_{t-1})$, by obtaining the optimal solution $\mathbf{x}^* \in \mathbb{R}^{AN_\epsilon}$ of the following problem

$$\min_{\mathbf{x}} \|\mathbf{x} - \bar{\mathbf{m}}_{t-1}\|_2^2$$
 s.t. $x_i \geq 0$, $1 \leq i \leq AN_\epsilon$ and $\sum_{i=1}^{AN_\epsilon} x_i \leq \epsilon^t$.
 - 4: Optimize the prediction distribution φ_b^t , i.e.,

$$\operatorname{argmin}_{\varphi_b^t} \max_{a_b^t \in \mathcal{A}} (\bar{\mathbf{m}}_{t-1} - \prod_C(\bar{\mathbf{m}}_{t-1})) \cdot \mathbf{m}(\varphi_b^t, a_b^t),$$
 based on the multiplicative weights algorithm [14], such that for $\forall a_b^t \in \mathcal{A}$, (7) is satisfied.
 - 5: **Output** the forecast result \mathbf{p}_b^t according to the optimal prediction distribution φ_b^{t*} over $\{\mathbf{p}_1, \dots, \mathbf{p}_{N_\epsilon}\}$.
-

3.2. Calibrated Learning Algorithm for Power Allocation

Let us consider an MAB problem that models a system of A arms whose expected rewards are i.i.d. over time with unknown means. The objective of an MAB problem is to maximize the accumulated reward over time via exploring the environment to find profitable actions, while exploiting current knowledge to make the empirically best decisions among a set of actions [15]. The max-min SINR problem investigated in this paper can be regarded as a distributed MAB problem, where each SBS acts as an agent and a forecaster, A transmit power levels correspond to A actions or arms, and the instantaneous reward of individual power levels chosen by SBS b can be defined as $\text{SINR}_b^{[t]}$. The proposed distributed calibrated learning based power allocation algorithm is governed by a trade-off between exploring different arms that might yield a better accumulated reward in the presence of wireless channel random variations, and exploiting the best arm

Algorithm 2 Distributed Calibrated Learning Main Algorithm at SBS b

- 1: **Initialize:** $t = 1$; total no. of time slots T , ϵ^t , $\gamma_{t=1} = 1$.
 - 2: **While** $t \neq T$ **do**
 - 3: with probability of γ_t : **Exploration Stage**
-The SBS transmits to its user at a random power level.
 - 4: with probability of $1 - \gamma_t$: **Exploitation Stage**:
-Receive the information of actual arm chosen by the opponents, i.e., $\{a_b^{t-1}\}_b$, as well as the average mean rewards of the opponents, i.e., $\bar{\boldsymbol{\mu}}_{b'}^{[t-1]} = \{\bar{\mu}_{b'1}, \dots, \bar{\mu}_{b'A}\}$, $b' \neq b, b' \in \mathcal{L}_b$ of previous time slot.
-Receive prediction of the opponents' actions for the current time slot, i.e., $\mathbf{p}_b^t = \{p_{a'}^t\}_{a' \in \mathcal{A}}$, from Algorithm 1.
-Calculate its estimated mean reward as $\hat{\boldsymbol{\mu}}_b^{[t]} = \{\sum_{a'=1}^A p_{a'}^t \overline{\text{SINR}}_b^{[t]}(a_b^t)\}_{a_b^t \in \mathcal{A}}$.
-Transmit at a power level associated with the highest max-min reward, as $k_b^t = \operatorname{argmax}_{a_b^t \in \mathcal{A}} \min(\hat{\boldsymbol{\mu}}_b^{[t]}, \{\bar{\boldsymbol{\mu}}_{b'}^{[t-1]}\})$.
 - 5: Observe the true associated reward $\text{SINR}_b^{[t]}$.
 - 6: Average $\bar{\boldsymbol{\mu}}_b^{[t]}$ over accumulated number of time slots, as $\bar{\boldsymbol{\mu}}_b^{[t]} = \frac{\sum_{t'=1}^{t-1} \bar{\boldsymbol{\mu}}_b^{[t']} \beta^{(t-t')} + \mathbf{1}_{k_b^t} \text{SINR}_b^{[t]}}{t}$.
 - 7: increment the time slot count $t = t + 1$.
 - 8: **end while**
-

so far that maximizes the accumulated reward at the individual SBSs. The details of the proposed calibrated learning algorithm to be executed at the individual SBSs are described in Algorithm 2. Note that the only information an SBS needs to share with other SBSs at the end of time slot t for the proposed design is its actual chosen arm, i.e., a_b^t , and a vector of its averaged mean reward, i.e., $\bar{\boldsymbol{\mu}}_b^{[t]}$, which is a light overhead as compared to that of the conventional iterative distributed power allocation designs [8–11].

4. SIMULATION RESULTS

Consider a distributed small cell network consists of $N_b = 2$ adjacent SBSs. Two users are randomly scheduled in the vicinity of the boundaries of the small cells to account for the worst-case ICI. Each SBS can transmit at $\{1, 2\}$ W discrete transmit power levels. The channel gain $\Psi_{bb}^{[t]}$ is scaled by $G_a L_{bb} \sigma_F^2 e^{-0.5 \frac{(\sigma_s \ln 10)^2}{100}}$ [8], where $L_{bb} = 125.2 + 36.3 \log_{10}(d)$ denotes the path loss model over a distance of d km between SBS b and its own user, the antenna gain is $G_a = 15$ dBi and the log-normal shadowing standard deviation is $\sigma_s = 8$ dB. The candidate probability values are set to be $\{\mathbf{p}_1, \dots, \mathbf{p}_{N_\epsilon}\} = \{(0, 1), \dots, (1, 0)\}$ with $N_\epsilon = 21$ over $A = 2$ actions. The performance of the proposed design is evaluated with $F = 13$ frames and 100 time slots per frame. In order to demonstrate the advantages of our proposed calibrated learning based power allocation design, a random

power allocation design that randomly selects transmit power level from \mathcal{A} at individual time slots as well as a max power allocation design that always chooses the maximum transmit power level, have been set as the benchmark schemes. Since the distributed algorithm may be suboptimal, a centralized max-min SINR design is employed as a performance upper bound.

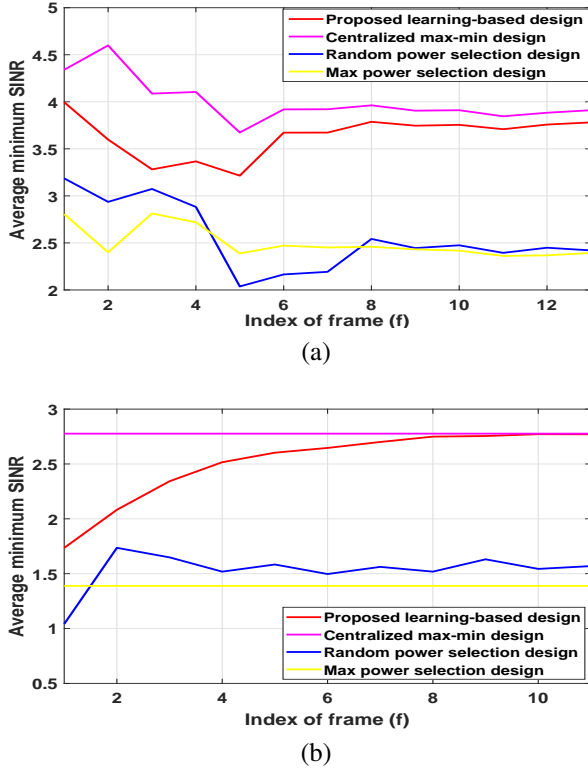


Fig. 1. Min SINR of various power allocation designs for (a) highly uncertain environment, (b) near-static environment.

Fig. 1 compares the average minimum SINR for various power allocation designs for highly uncertain wireless channel variations, i.e., Fig. 1(a), and near-static channel variations, i.e., Fig. 1(b), respectively. As can be seen from Fig. 1, the proposed design progressively converges to the centralized design for the case of static environment and has a slightly degraded performance for the case of uncertain environment, whilst the performance gap between the proposed design and the optimal centralized design is narrowed with increasing number of frames. In addition, the proposed design outperforms two benchmark schemes in both cases. This is due to the fact that in the benchmark schemes, the individual SBS autonomously selects its transmit power level without taking into account the impact it may have on its counterparts as well as the variations in wireless channel conditions. As can be observed from Fig. 1(a), when the wireless channel varies rapidly, the proposed design achieves a performance

Table 1. Reward table for SBSs at the final time slot

SBS1 \ SBS2	Arm1	Arm2
Arm1	2.4286/ 9.6637	1.2146/ 20.3274
Arm2	4.8572/ 6.8347	2.4291/ 13.6694

improvement of approximately 50 percent from the 6th frame onwards, as compared to the benchmark schemes.

Table I provides possible SINRs that could be achieved for various combinations of chosen arms for the two SBSs at the final time slot. One may observe from the table that the maximization of the minimum SINR of 4.8572 is achieved when SBS1 and SBS2 transmit, respectively, at power level 2 and power level 1. Fig. 2 presents the individual SBS's

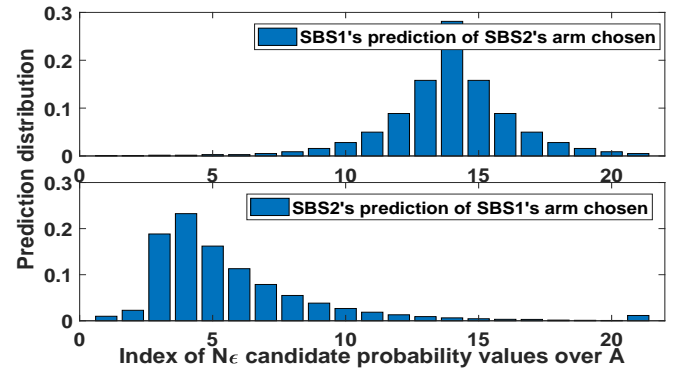


Fig. 2. Prediction distribution of N_ϵ candidate probability values at final time slot.

prediction distribution, i.e., φ_b^t , of candidate probability values $\{p_1, \dots, p_{N_\epsilon}\}$ over its opponent's arm chosen at the final time slot. It is obvious from the figure that SBS1 and SBS2 are highly likely to be expected by their opponents to choose arm 2 and arm 1, respectively. This conclusion is in agreement with Table I, where SBS1 choosing arm 2 whilst SBS2 selecting arm 1 leads to the maximization of minimum SINR.

5. CONCLUSION

This paper proposes a calibrated learning based power allocation algorithm in distributed small cell networks, which adapts to the dynamic wireless channel conditions as well as other SBSs' actions, and achieves max-min user fairness in the long run. The proposed design allows the SBSs to gradually improve their prediction on the opponents' behaviour and react with the best response based on the forecast results at individual time slots. Simulation results validate that the proposed distributed design outperforms two benchmark schemes and closely follows the optimal centralized design with very limited amount of side information exchange.

6. REFERENCES

- [1] Cisco, “Cisco visual networking index: Forecast and methodology, 2016-2021,” *White Paper*, Sept 2017.
- [2] F. Al-Turjman, E. Ever, and H. Zahmatkesh, “Small cells in the forthcoming 5g/iot: Traffic modelling and deployment overview,” in *Proceedings Title*. IEEE, Aug 2018.
- [3] P. Rost, C. J. Bernardos, A. D. Domenico, M. D. Girolamo, M. Lalam, A. Maeder, D. Sabella, and D. Wubben, “Cloud technologies for flexible 5g radio access networks,” *IEEE Communications Magazine*, vol. 52, no. 5, pp. 68–76, May 2014.
- [4] Z. Han, D. Niyato, W. Saad, T. Baar, and A. Hjrungnes, *Game Theory in Wireless and Communication Networks: Theory, Models, and Applications*, Cambridge University Press, New York, USA, 2012.
- [5] Z. Han and K. J. R. Liu, “Noncooperative power-control game and throughput game over wireless networks,” *IEEE Transactions on Communications*, vol. 53, no. 10, pp. 1625–1629, Oct. 2005.
- [6] Z. Wang, B. Hu, X. Wang, and S. Chen, “Interference pricing in 5g ultra-dense small cell networks: a stackelberg game approach,” *IET Communications*, vol. 10, no. 15, pp. 1865–1872, Oct. 2016.
- [7] G. Bacci, E. V. Belmega, P. Mertikopoulos, and L. Sanguinetti, “Energy-aware competitive power allocation for heterogeneous networks under qos constraints,” *IEEE Transactions on Wireless Communications*, vol. 14, no. 9, pp. 4728–4742, 2015.
- [8] Z. Xiang, M. Tao, and X. Wang, “Coordinated multicast beamforming in multicell networks,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 1, pp. 12–21, Jan. 2013.
- [9] X. Zhang and M. R. Nakhai, “Robust chance-constrained distributed beamforming for multicell interference networks,” *IEEE International Conference on Communications (ICC)*, pp. 1–6, May 2016.
- [10] O. Tervo, H. Pennanen, D. Christopoulos, S. Chatzinothas, and B. Ottersten, “Distributed optimization for coordinated beamforming in multicell multigroup multicast systems power minimization and sinr balancing,” *IEEE Transactions on Signal Processing*, vol. 66, no. 1, pp. 171–185, Jan. 2018.
- [11] C. Shen, T. H. Chang, K. Y. Wang, Z. Qiu, and C. Y. Chi, “Distributed robust multicell coordinated beamforming with imperfect csi: An admm approach,” *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2988–3003, Feb. 2012.
- [12] S. Mannor and G. Stoltz, “A geometric proof of calibration,” *Mathematics of Operations Research*, vol. 35, no. 4, pp. 721–727, Nov. 2010.
- [13] D. Blackwell, “An analog of the minimax theorem for vector payoffs,” *Pacific Journal of Mathematics*, vol. 6, no. 1, pp. 1–8, 1956.
- [14] Y. Freund and R. E. Schapire, “Adaptive game playing using multiplicative weights,” *Games and Economic Behavior*, vol. 29, pp. 79–103, 1999.
- [15] P. Auer, N. Cesa-Bianchi, and P. Fischer, “Finite-time analysis of the multi-armed bandit problem,” *Machine learning*, vol. 47, no. 2-3, pp. 235–256, May 2002.