Reinforcement Learning-Based Interference Control for Ultra-Dense Small Cells

Hailu Zhang*, Minghui Min*, Liang Xiao*, Sicong Liu*, Peng Cheng†, Mugen Peng‡
*Dept. of Communication Engineering and Key Laboratory of Digital Fujian on IoT Communication, Architecture and Security Technology (IoTCAS), Xiamen University, Xiamen, China. Email: {liusc, lxiao}@xmu.edu.cn
†State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou, China
‡School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China

***Abstract*—The densification deployment of small cells emerging into 5G cellular networks can achieve high capacity, but is faced with the challenge of how to manage energy consumption and inter-cell interference well in time-varying channels. In this paper, we propose a reinforcement learning based downlink power control algorithm to manage interference for the ultra-dense small cell networks. More specifically, base stations of the small cells use Q-learning to select the downlink transmit powers. A transfer learning method called hotbooting is applied to further accelerate the learning speed and save the energy consumption based on the estimated user density without being aware of the network and channel model of the other small cells. Simulation results demonstrate this scheme significantly improves the network throughput and saves the energy consumption compared with the benchmark, a data-driven based transmission power adaptation scheme.**

***Index Terms*—Ultra-dense small cells, interference, energy consumption, power control, reinforcement learning.**

I. Introduction

The fifth generation (5G) mobile communication systems can apply ultra-dense small cells with low-cost and low-power cellular base stations (BSs) to improve the user capacity if the interference in the ultra-dense cell deployments is well managed. Therefore, power control is critical for interference management and energy saving in ultra-dense small cell systems [1]. A power control algorithm as proposed in [2] uses the power update function and noncooperative game theory to choose the transmit power for each user in the target cell to mitigate the inter-cell interference. A mean-filed approach based power control as presented in [3] uses Lax-Friedrichs scheme and Lagrange relaxation to choose the downlink transmit power for saving the energy consumption and mitigating the interference in a highly dense network. A cooperative optimal power control scheme in [4] applies the cost index and quadratic programming framework for interference management in cellular networks.

The work of Liang Xiao was supported in part by NSFC under Grant 61671396, 91638204, 61731012, 61371081, 61533015, and 61728101, in part by the Open Research Project of the State Key Laboratory of Industrial Control Technology, Zhejiang University under Grant ICT180036 and in part by the State Major Science and Technology Special Project under Grant 2016ZX03001020-006. (*Corresponding author: Sicong Liu*).

Nevertheless, the interference control in the ultra-dense small cell systems has to address the huge signaling and computation overhead to collect the communication and interference information from the small cells, the time-varying radio channels and the dynamic user density in each small cell.

The power control process in the ultra-dense small cell system can be formulated as a Markov decision process (MDP), in which future network state is independent of the previous state for the given current power control policy and the inter-cell interference of the small cell system. Thus, a BS can use reinforcement learning (RL) such as Q-learning for interference management without knowing the network and channel model of the other small cells.

In this paper, we propose a hotbooting Q based downlink power control algorithm for outdoor ultra-dense small cell systems. Based on the hotbooting technique, the Q-value is initialized with the relay power control experiences in similar 5G communication scenarios [5] to save the random explorations at the beginning of the process and accelerate the learning speed. Each small cell BS chooses the downlink transmit power based on the state of the time slot, which consists of the local user density and the signal-to-interference-plus-noise ratio (SINR) of the signals fed back by the users. This algorithm maintains a Q-function or expected long-term discount utility for each state and action pair via iterative Bellman equations. Each BS evaluates the utility that depends on the SINR of the signals from the target BS to served users, the transmit cost of the target cell and the interference from the target BS to the users in the other cells, which is the basis to update the Q-values in each time slot. Simulation results show that this algorithm reduces the inter-cell interference, increases the system throughput and saves the energy consumption compared with the data-driven transmit power adaptation (TPA) scheme as presented in [6].

The reminders of this paper are organized as follows. We review the related work in Section II and present the system model in Section III. We present the hotbooting Q based downlink power control algorithm in Section IV. Simulation results are given in Section V, followed by the conclusion in Section VI.

978-1-5386-4727-1/18/$31.00 ©2018 IEEE

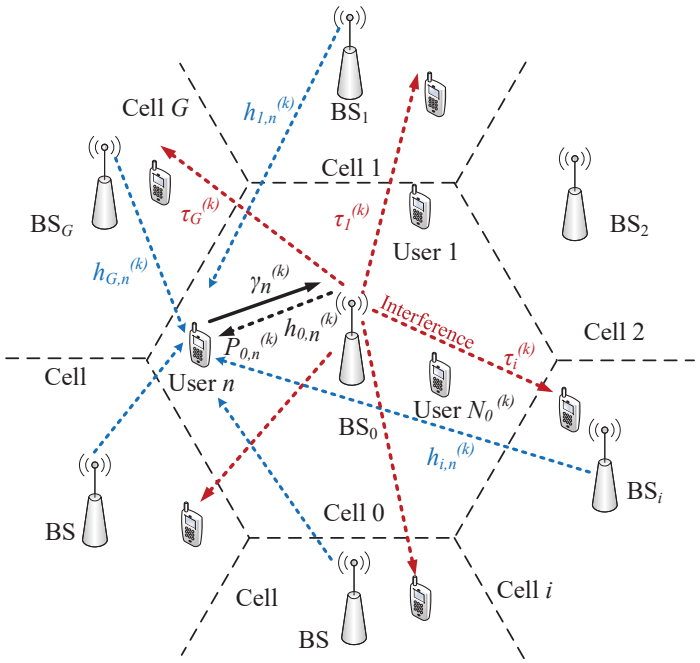


Fig. 1: Illustration of the interference mitigation in a ultra-dense small cell system, in which the BS of the target small cell with $N_0^{(k)}$ mobile users chooses its $P_{0,n}^{(k)}$ to mitigate the interference to the neighboring G small cells at time slot k, and the user n returns the estimated SINR $\gamma_n^{(k)}$ to the BS.

II. Related Work

Power control is a key technique to reduce the inter-cell interference in ultra-dense small cell systems. In particular, a mixes-integer programming based power control scheme as investigated in [7] combines the user association to reduce the interference in millimeter-wave systems. A dynamic pricing based power control scheme as proposed in [8] can reach the Nash equilibrium of the non-cooperative game. A data-driven BS power control scheme as proposed in [6] uses statistics analysis approach to save the energy cost for ultra-dense small cell system.

Reinforcement learning approaches have been applied in the power control problem of wireless networks. For instance, the downlink power control and rate adaptation scheme as presented in [9] uses Lagrange duality theory and artificial neural network to improve the resource allocation utility. A centralized Q-learning with compact state representation algorithm is investigated in [10] where the network controller solves the optimal traffic offloading strategy based on the traffic observations to minimize the energy cost and maintain the quality-of-service. A fuzzy Q-learning based power control scheme as proposed in [11] formulates the inter-cell interference coordination issue as a cooperative MultiAgent control problem to improve the performance of the cellular systems. A RL based decentralized power control strategy is proposed in [12], in which small cells jointly estimate time-average performance and optimize probability distributions for interference management in closed-access small cell networks. A dynamic Q-learning based interference coordination algorithm as proposed in [13] smartly offloads traffic to open access picocells and then improves the system throughput.

III. System Model

The BS of the target small cell (i.e., BS_0) is assumed to $N_0^{(k)}$ mobile users at time slot k, and its signals can reach G neighboring cells in the small cell system as shown in Fig. 1. Mobile user n measures the bit error rate (BER) of the message, then estimates the SINR of the signals, denoted by $\gamma_n^{(k-1)}$, and returns such information to the serving BS.

Orthogonal Frequency Division Multiple Access technology has been selected for ultra-dense small cell networks. Equipped with multiple isotropic antennae, the BS_0 assigns transmission channel bandwidth B to users in the cell. Similar to [14], the cellular system has reciprocal radio channel model, and thus the BS_0 can estimate the downlink channel state to the user n, denoted by $h_{0,n}^{(k)}$. Similarly, the channel gain between the other small cell i and the user n at time slot k is denoted by $h_{i,n}^{(k)}$. For simplicity, we assume a constant noise power denoted by σ at each user.

Upon receiving the feedback control information from the $N_0^{(k)}$ users at time slot k, the BS_0 formulates the SINR vector denoted by $\boldsymbol{\gamma}^{(k-1)} = [\gamma_1^{(k-1)}, \cdots, \gamma_{N_0}^{(k-1)}]$. The user density of the G neighboring cells changes over time, i.e., the number of the active users in cell i at time slot k denoted by $\rho_i^{(k)}$. According to [15], the user density $\rho_i^{(k)}$ is assumed to follow the independently and identically distributed two-dimensional

TABLE I: Summary of symbols and notations

$N_0^{(k)}$	Num. users in the target small cell at time slot k
$\rho_0^{(k)}$	User density of the target small cell
ϕ_0	Area of the target small cell
$\boldsymbol{\gamma}^{(k)}$	SINR from the mobile users
P^M	Maximum BS transmit power
L	Num. the feasible transmit power levels
$\mathbf{P}^{(k)}$	Transmit power of the BS
$\boldsymbol{\Omega}$	Feasible BS transmit power set
α	Learning rate in Algorithm 1
β	Discount factor in Algorithm 1
$u^{(k)}$	Utility of the BS
$\mathcal{C}_s$	Transmit cost per unit power of the BS
B	Downlink bandwidth for a user

Poisson point process. The number of users $N_0^{(k)}$ of the target cell with area ϕ_0 is given by

$$\Pr\{N_0^{(k)} = \lambda|\phi_0\} = \frac{(\rho_0^{(k)}\phi_0)^\lambda}{\lambda!} e^{-\rho_0^{(k)}\phi_0}. \quad (1)$$

To minimize the interference, the target BS uses the average interference factor denoted by $\tau_i^{(k)}$, to judge its interference to non-served users in the other cell i. According to [16] and [17], the average interference factor $\tau_i^{(k)}$ at time slot k is given by:

$$\tau_i^{(k)} = \frac{g_i^{(k)}\sqrt{\eta}}{l_i^{(k)}\sqrt{|G+1|}}, \quad (2)$$

where η is the small cell density and $\eta/|G+1|$ is the normalization factor. The $g_i^{(k)}$ denotes the large-scale fading gain from the BS_0 to other small cell i, and the $l_i^{(k)}$ is the path loss from the BS_0 to other small cell i at time slot k.

To this end, the BS_0 has to choose its transmit power denoted by $P_{0,n}^{(k)}$ from the action set $\boldsymbol{\Omega} = [lP^M/L]_{0\le l\le L}$, where P^M is the maximum downlink transmit power of the BS and L is the number of the feasible transmit power level. The interference power to the user n from the BS in the surrounding cell i at time slot k is denoted by $P_{i,n}^{(k)}$.

For ease of reference, we summarize our commonly used notation in Table 1.

IV. Hotbooting Q based Power Control Algorithm

In this section, we propose a hotbooting Q based power control scheme to manage interference and reduce energy cost, in which each BS exploits the SINR sent by users and estimates user density to achieve an optimal power control solution via trial without knowledge of the network and channel model.

In the dynamic power control process, the BS in the target cell estimates the user density $\rho_0^{(k)}$ of the target small cell by statistic law according to Eq. (1) at time slot k. Meanwhile the BS requires the served users to provide the SINR. Once receiving the BS's request, all served users send the SINR $\boldsymbol{\gamma}^{(k-1)}$ to the BS at that time. Thus, the state observed by the BS at time slot k, denoted by $\mathbf{s}^{(k)}$, consists of the current user density and the previous SINR of the users, i.e., $\mathbf{s}^{(k)} = \left[\rho_0^{(k)}, \boldsymbol{\gamma}^{(k-1)}\right]$.

As power control decision of the BS has impacts on the future state of the target small cell, the power control process of the BS in the dynamic game can be formulated as a MDP. Therefore, the BS can apply RL techniques such as hotbooting Q to derive the optimal strategy via trials without knowledge of the network and channel model. The hotbooting technique is used to initialize the Q-value with the power control experiences in similar environments to save the random explorations at the beginning of the interference control process and then accelerate the learning speed [18]. The output of the hotbooting Q technique, i.e., Q^* is the initial Q-value.

Based on the system state, the BS in the target cell chooses the transmit power for N_0 users at time slot k, denoted by $\mathbf{P}^{(k)} = \left[P_{0,1}^{(k)}, P_{0,2}^{(k)}, \ldots, P_{0,N_0}^{(k)}\right]$ and sends a message to each user n with $P_{0,n}^{(k)}$. The BS then evaluates its utility obtained at the time slot k, denoted by $u^{(k)}$ based on the user density at time slot k, the current SINR of users, the transmit power chose by the BS and given by

$$u^{(k)} = \sum_{n=1}^{N_0^{(k)}} \gamma_n^{(k)} - \mathcal{C}_s \sum_{n=1}^{N_0^{(k)}} P_{0,n}^{(k)} \left(\sum_{i=1}^{G} \tau_i^{(k)} N_i^{(k)} + 1\right), \quad (3)$$

where the first term represents the SINR sent by the served users at time slot k, $N_i^{(k)}$ is the number of users in the other cell i. The second term stands for the energy consumption of the target cell and the interference from the target cell to non-served users in the other cells. $\mathcal{C}_s$ is the transmit cost per unit power of the BS.

The power control process with hotbooting Q is based on the learning rate, denoted by $\alpha \in (0, 1]$, which shows the weight of the current experience. The discount factor $\beta \in [0, 1]$ indicates the uncertainty of the regarding the future utility. The Q-function of the transmit power vector $\mathbf{P}^{(k)}$ at state $\mathbf{s}^{(k)}$ is denoted by $Q\left(\mathbf{s}^{(k)}, \mathbf{P}^{(k)}\right)$ and is updated according to iterative Bellman equation as follows:

$$Q\left(\mathbf{s}^{(k)}, \mathbf{P}^{(k)}\right) \leftarrow (1-\alpha) Q\left(\mathbf{s}^{(k)}, \mathbf{P}^{(k)}\right) + \alpha\left(u^{(k)} + \beta V\left(\mathbf{s}^{(k+1)}\right)\right) \quad (4)$$

$$V(\mathbf{s}^{(k)}) \leftarrow \max_{\mathbf{P}\subseteq\boldsymbol{\Omega}} Q\left(\mathbf{s}^{(k)}, \mathbf{P}\right), \quad (5)$$

where the value function $V(\mathbf{s}^{(k)})$ is the maximal Q-function over the feasible power control scheme at state $\mathbf{s}^{(k)}$.

The BS in the target cell applies the ϵ-greedy policy to determine the optimal transmit power that maximizes the utility with a high probability $1-\epsilon$, and chooses the suboptimal transmit power with a small probability ϵ to avoid staying in

the local maximum, i.e.,

$$\Pr\left(\mathbf{P}^{(k)}=\mathbf{\Theta}\right)=\begin{cases}1-\epsilon, & \mathbf{\Theta}=\arg\max\limits_{\mathbf{P}\in\mathbf{\Omega}} Q(\mathbf{s}^{(k)},\mathbf{P})\\ \frac{\epsilon}{|L|}, & \text{o.w.}\end{cases} \tag{6}$$

The detailed power control processes with hotbooting Q algorithm is summarized in algorithm 1.

Algorithm 1 Hotbooting Q based BS interference control Algorithm for ultra-dense small cell systems

1: Initialize α, β, $\mathbf{\Omega}$, $\mathbf{Q}=\mathbf{Q}^*$, $\mathbf{V}=\mathbf{0}$, and $\boldsymbol{\gamma}^{(k-1)}=\mathbf{0}$
2: **for** $k=1,2,...$ **do**
3: Estimate the current user density $\rho^{(k)}$ of the target small cell via Eq. (1)
4: Receive the SINR $\boldsymbol{\gamma}^{(k-1)}$ from the served users
5: Obtain the current system state $\mathbf{s}^{(k)}=\left[\rho^{(k)},\boldsymbol{\gamma}^{(k-1)}\right]$ of the target small cell
6: Select the transmit power $\mathbf{P}^{(k)}\in\mathbf{\Omega}$ via Eq. (6) for the served users
7: Evaluate the total transmit cost of target small cell and the interference to non-served users
8: Receive the current SINR $\boldsymbol{\gamma}^{(k)}$ from the served users
9: Evaluate utility $u^{(k)}$ via Eq. (3)
10: Update $Q(\mathbf{s}^{(k)},\mathbf{P}^{(k)})$ via Eq. (4)
11: Update $V\left(\mathbf{s}^{(k)}\right)$ via Eq. (5)
12: **end for**

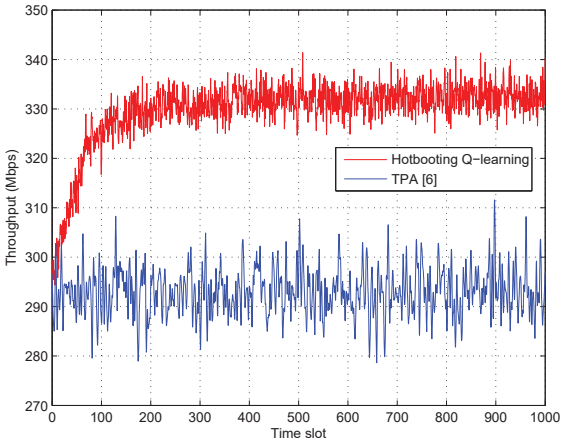
V. Simulation Results

Simulations are performed to evaluate the interference control strategy in an ultra-dense small cell network. In the simulations, the target cell is interfered by the mobile users in the neighboring 6 small cells. If not specified otherwise, we set the $L=4$, $\alpha=0.5$, $\beta=0.85$, $\epsilon=0.1$, $C_s=12$, and $B=10$ MHz to achieve good power control performance according to the experiments not presented in this paper. Each active user in the target small cell estimates the SINR with quadrature phase-shift keying based on the measured BER. Thus, the network throughput of the target small cell at time slot k is given by

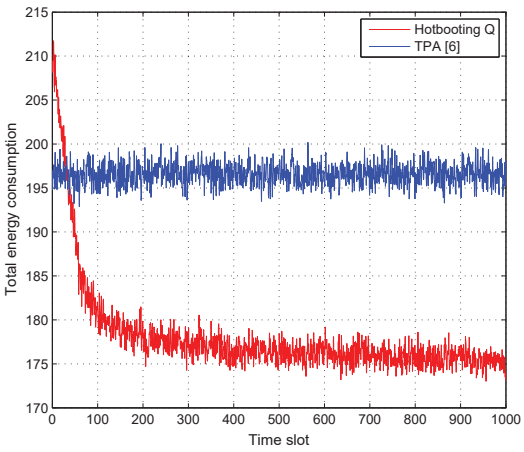
$$R^{(k)}=\sum_{n=1}^{N_0^{(k)}}\frac{B}{N_0^{(k)}}\log_2\left(1+\frac{P_{0,n}^{(k)}h_{0,n}^{(k)}}{\sigma+\sum_{i=1}^{G}P_{i,n}^{(k)}h_{i,n}^{(k)}}\right). \tag{7}$$

As a benchmark, the data-driven TPA scheme proposed in [6] is considered in which each BS adaptively adjusts the downlink transmit power of users based on the edge reference signal received power in each small cell.

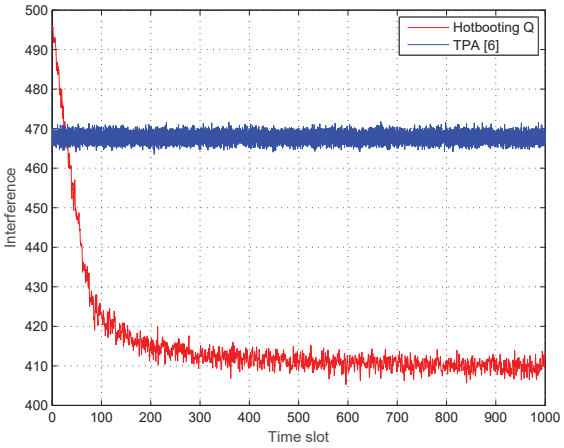
As shown in Fig. 2, before 50 time slots, the performance of the TPA scheme is better than that of our proposed scheme. This is because the TPA scheme dynamically adjusts the transmit power to maintain a stable and optimal state of the system based on the received signal strength of cell edge. While the BS based on our proposed scheme is still in the exploratory stage.



(a) Throughput of the target small cell.

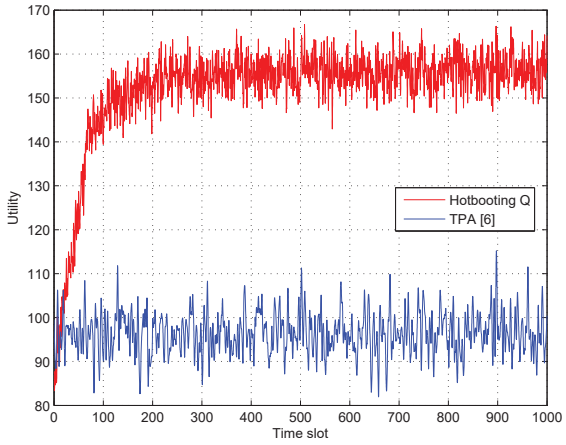


(b) Energy consumption of the target small cell.



(c) Interference of the target small cell to users in the other cells.

After 50 time slots, the proposed power control algorithm based on the hotbooting Q outperforms the TPA based strategy with lower energy cost, less interference, higher system



(d) Utility of the BS.

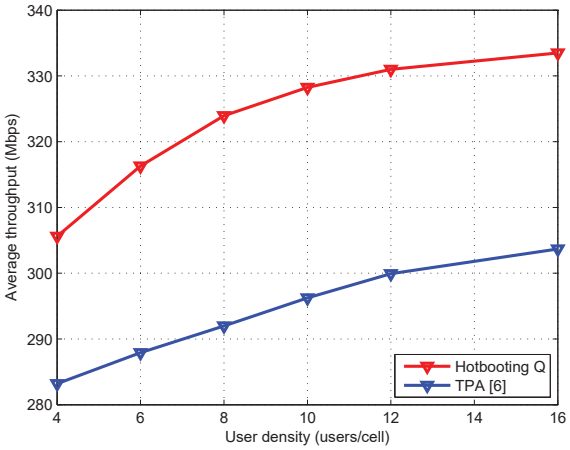
Fig. 2: Performance of the BS power control algorithms in the ultra-dense small cell system, with $G = 6$, $L = 4$, $\alpha = 0.5$, $\beta = 0.85$, $C_s = 12$ and $B = 10$ MHz.

throughput and utility. For instance, as shown in Fig. 2 (a), the throughput of the target small cell increases over time with our proposed scheme, and converges to 335 Mbps after 200 time slots, which is 15.5% higher than that of the TPA strategy, because the BS adjusts the transmit power over time via trials-and-errors.

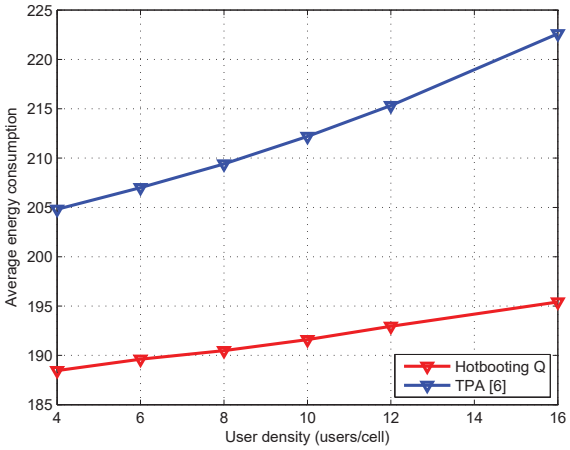
In Fig. 2 (b), due to the optimal downlink transmit power control strategy based on the hotbooting Q, the total energy cost of the target cell decreases by 17.5% after convergence, and its optimal energy cost is 9.32% lower than that of the TPA strategy. Similar to the case of interference as shown in Fig. 2 (c), the hotbooting Q based strategy decreases the interference to users in the other cells by 18.2% after convergence, and the optimal interference of the hotbooting Q based strategy is 12.8% lower than that of the TPA strategy. Consequently, as shown in Fig. 2 (d), the utility of the BS increases quickly after start of the learning process, and converges to 160 which exceeds the benchmark strategy by 60% at time slot 300.

We evaluate the impact of the user density in Fig. 3 on the communication performance averaged over 200 learning processes and 1000 time slots. It is shown that the average throughput of the target cell improves with the number of active users per cell, but at a logarithmic speed in both two schemes. For instance, as shown in Fig. 3 (a), the average throughput based our proposed algorithm increases from 306 Mbps to 334 Mbps, if the user density changes from 4 to 16. The growth trend of the average utility is consistent with the average throughput in both two schemes. For example, as shown in Fig. 3 (d), the average utility based our proposed algorithm increases by 31.4%, as the users density changes from 4 to 16, compared with the TPA scheme.

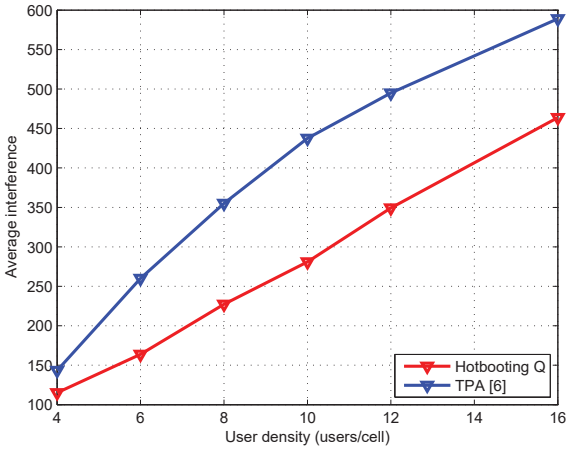
In other words, more users may bring mild throughput and utility improvement for ultra-dense small cells due to more severe inter-cell interference and energy consumption. For example, as shown in Fig. 3 (b), the energy cost of the target



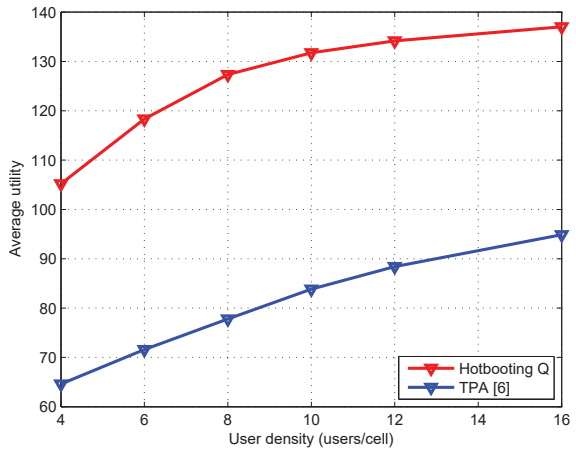
(a) Average throughput of the target small cell.



(b) Average energy consumption of the target small cell.



(c) Average interference of the target small cell.



(d) Average utility of the BS.

Fig. 3: Average performance of the target small cell for different schemes versus the user density over 1000 time slots.

cell increases from 205 to 223 based on the TPA scheme, and as shown in Fig. 3 (c), the interference of the target cell increases from 150 to 558 based on the TPA scheme, if the users density changes from 4 to 16. Therefore, there is a tradeoff among the energy cost, interference and throughput, which can be adjusted by properly setting the user density limit for the small cells.

It also can be seen that our proposed power control scheme outperforms the TPA scheme in [6]. For example, as shown in Fig. 3 (a), the average throughput with our proposed approach is up to 330.9 Mbps, which is about 10.3% higher than that of the TPA strategy if the user density is 14. As shown in Fig. 3 (b), the average energy cost of the target small cell with our proposed approach is about 12.4% lower than that of the TPA scheme if the user density is 16. As shown in Fig. 3 (c), the average interference based on our scheme is 19.2% lower than that of the benchmark strategy as the user density is 16. Consequently, as shown in Fig. 3 (d), the utility of the target BS with our proposed scheme exceeds the TPA strategy by 57.2% if the user density is 12.

VI. Conclusion

In this paper, we propose a reinforcement learning based interference control algorithm for the downlink transmissions in ultra-dense small cell systems. This algorithm reduces the inter-cell interference and saves the energy consumption of the BS without being aware of the network and channel model. Simulation results show this BS power control scheme significantly improves the throughput, reduces the energy consumption and mitigates interference to the other small cells compared with the data-driven TPA strategy in the 5G cellular network. For example, the energy consumption of the target small cell is 17.5% lower and the utility of the BS is 60% higher, compared with the benchmark scheme after 200 time slots.

References

[1] V. Chandrasekhar, J. G. Andrews, T. Muharemovic, Z. Shen, and A. Gatherer, "Power control in two-tier femtocell networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, Aug. 2009.

[2] V. N. Ha and L. B. Le, "Distributed base station association and power control for heterogeneous cellular networks," *IEEE Trans. Vehicular Technology*, vol. 63, no. 1, pp. 282–296, Aug. 2014.

[3] P. Semasinghe and E. Hossain, "Downlink power control in self-organizing dense small cells underlaying macrocells: A mean field game," *IEEE Trans. Mobile Computing*, vol. 15, no. 2, pp. 350–363, Mar. 2016.

[4] A. Sanchez, J. Arauz, J. W. McClure, and Z. Miller, "Cooperative self-organized optimal power control for interference mitigation in femtocell networks," in *Proc. IEEE Commun. Computing (COLCOM)*, pp. 1–6, Cartagena, Colombia, Apr. 2016.

[5] L. Xiao, Y. Li, C. Dai, H. Dai, and H. V. Poor, "Reinforcement learning-based noma power allocation in the presence of smart jamming," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3377–3389, Dec. 2018.

[6] L.-C. Wang, S.-H. Cheng, and A.-H. Tsai, "Bi-son: Big-data self organizing network for energy efficient ultra-dense small cells," in *Proc. IEEE Veh. Technol. Conf. (VTC-Fall), Montreal*, pp. 1–5, QC, Canada, Mar. 2016.

[7] H. Zhang, S. Huang, C. Jiang, K. Long, V. C. Leung, and H. V. Poor, "Energy efficient user association and power allocation in millimeter-wave-based ultra dense networks with energy harvesting base stations," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1936–1947, Jun. 2017.

[8] J. Zheng, Y. Wu, N. Zhang, H. Zhou, Y. Cai, and X. Shen, "Optimal power control in ultra-dense small cell networks: A game-theoretic approach," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4139–4150, Dec. 2017.

[9] E. Ghadimi, F. D. Calabrese, G. Peters, and P. Soldati, "A reinforcement learning approach to power control and rate adaptation in cellular networks," in *Proc. IEEE Int'l Conf. Commun. (ICC)*, pp. 1–7, Paris, France, Jul. 2017.

[10] X. Chen, C. Wu, Y. Zhou, and H. Zhang, "A learning approach for traffic offloading in stochastic heterogeneous cellular networks," in *Proc. IEEE Int'l Conf. Commun. (ICC)*, pp. 3347–3351, London, UK, Sep. 2015.

[11] M. Dirani and Z. Altman, "A cooperative reinforcement learning approach for inter-cell interference coordination in ofdma cellular networks," in *Proc. IEEE Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, pp. 170–176, Avignon, France, Jun. 2010.

[12] M. Bennis, S. M. Perlaza, P. Blasco, Z. Han, and H. V. Poor, "Self-organization in small cell networks: A reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3202–3212, Jun. 2013.

[13] M. Simsek, M. Bennis, and I. Güvenç, "Learning based frequency- and time-domain inter-cell interference coordination in hetnets," *IEEE Trans. Veh. Technol.*, vol. 64, no. 10, pp. 4589–4602, Nov. 2015.

[14] K. Ren, H. Su, and Q. Wang, "Secret key generation exploiting channel characteristics in wireless communications," *IEEE Wireless Commun.*, vol. 18, no. 4, Aug. 2011.

[15] T. X. Brown, "Cellular performance bounds via shotgun cellular systems," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 11, pp. 2443–2455, Nov. 2000.

[16] S. Samarakoon, M. Bennis, W. Saad, M. Debbah, and M. Latva-Aho, "Ultra dense small cell networks: Turning density into energy efficiency," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1267–1280, May. 2016.

[17] J. Cao, T. Peng, Z. Qi, R. Duan, Y. Yuan, and W. Wang, "Interference management in ultra-dense networks: A user-centric coalition formation game approach," *IEEE Trans. Veh. Technol.*, vol. 67, no. 6, Jun. 2018.

[18] L. Xiao, Y. Li, G. Han, H. Dai, and H. V. Poor, "A secure mobile crowdsensing game with deep reinforcement learning," *IEEE Trans. Inform. Forensics and Security*, vol. 13, no. 1, pp. 35–47, Jan. 2017.