Learning Based Power Control for mmWave Massive MIMO Against Jamming

Zhongcheng Xiao*, Bin Gao†, Sicong Liu‡, Liang Xiao*‡

*School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China
†Key Laboratory of Underwater Acoustic Signal Processing of Ministry of Education, Southeast University, Nanjing, China
‡Department of Communication Engineering, Xiamen University, Xiamen, China. Email: liusc@xmu.edu.cn

***Abstract*—Millimeter-wave (mmWave) massive multiple-input multiple-output (MIMO) systems have to address smart jammers that use smart radio devices to choose the jamming policy with the goal of interrupting the ongoing transmissions. In this paper, we propose a reinforcement learning based power control strategy for the downlink mmWave massive MIMO systems. More specifically, we present a fast policy hill-climbing based power control algorithm for a base station to choose the transmit power over multiple antennas. Based on the signal-to-interference-plus-noise ratio (SINR) of the signals and the jamming strength, we evaluate the impact of the number of transmit antennas on the communication performance. Simulation results verify that the proposed schemes can increase the average SINR, sum data rate and the utility of the mmWave massive MIMO against smart jamming compared with the benchmark strategy.**

***Index Terms*—mmWave, massive MIMO, jamming, power control, reinforcement learning.**

I. Introduction

With the high improvements in the capacity and the potential bandwidth, Millimeter-wave (mmWave) massive multiple-input multiple-output (MIMO) is regard as one of the key technologies in the 5G cellular communication systems [1] [2] [3]. A switch and inverter based hybrid precoding architecture proposed in [4] can be used for the systems to reduce the hardware cost and energy consumption. Instead of using phase shifters, the analog part of the proposed architecture is realized by a small number of energy-efficient switch and inverters.

However, the mmWave massive MIMO has to address jamming attacks [5] [6], especially smart jammers that use the smart radio device such as universal software radio peripherals to choose the jamming policy such as the jamming power and channels based on the ongoing transmission status and the channel states [7]. Otherwise, jammers can interrupt the ongoing transmission of the base station (BS) to reduce the network throughput and exhaust the batteries of the mobile devices. By reducing the signal-to-interference-plus-noise ratio (SINR) of the signals, the jammers can significantly decrease the data rates of the mmWave massive MIMO transmission and even result in the denial of the service attacks.

Game theory on the power control can be used to analyze the wireless security under smart jamming with the known jamming model and channel model [8] [9]. For instance, an stackelberg equilibrium is developed in [8] for the BS to derive the optimal transmit power so as to improve the SINR of the users with a low transmission cost. However, in the dynamic interactions of the anti-jamming system, it's hard for the BS to know about the jamming model and the channel model, and the BS can't adjust the transmit power in time to improve the performance of anti-jamming [10].

The power control process in the repeated interactions between the BS and the smart jammer is a Markov decision process (MDP), thus the reinforcement learning algorithm could be used to improve the performance of anti-jamming on the basis of the communication quality without knowing the jamming model and the channel model. A transfer learning technique that is called hotbooting technique as developed in [11] is proposed for the BS to accelerate the initial learning based on the anti-jamming power control experiences.

The policy hill-climbing (PHC) algorithm [12] which keeps not only the Q-function but also the current mixed strategy is a model-free reinforcement learning technique. It chooses the action based on the mixed strategy, and updates the Q-function and mixed strategy in a time slot. More specifically, in the PHC power control strategy, the BS chooses the transmit power according to the mixed strategy with the given system state, i.e., the users' last SINR and the last jamming power to send the signals to the users. The moment when the users receive the signals, they estimate the communication quality and send the SINR to the BS.

Simulations are carried out to evaluate our algorithm, and the results show that the fast PHC based power control strategy we proposed improves the performance of anti-jamming. The average SINR, sum data rate of the users and the utility of the BS increase with time slot, and converge to a higher level compared with those of the benchmark strategy RCRA proposed in [13]. In addition, with the increase of the antennas, the average SINR, sum data rate and the utility show significantly improvement.

The rest of this paper is organized as follows. We present the system model in Section II, and propose a fast PHC-

This work was supported in part by the National Natural Science Foundation of China under Grant 61671396, BK20170765 and the open research fund of the National Mobile Communications Research Laboratory, Southeast University (No.2018D08).

based power control strategy for the mmWave massive MIMO system in Section III. Simulation results are provided in Section IV. In Section V, we conclude this work.

II. System Model

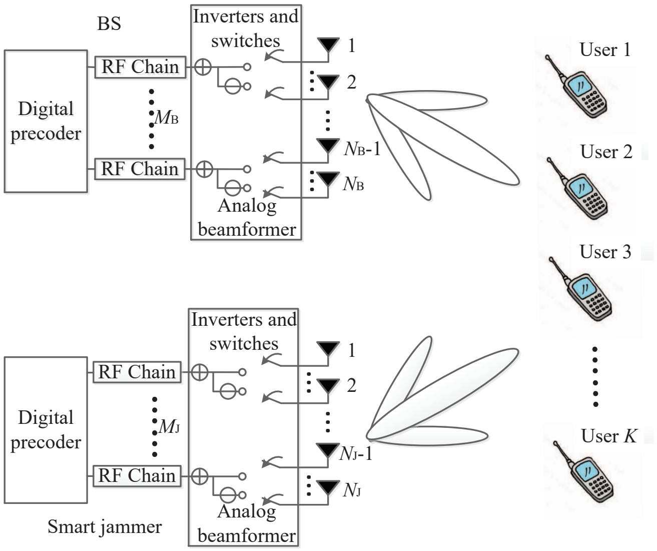


Fig. 1. Power control of a mmWave massive MIMO system against jamming.

A. Network Model

As illustrated in Fig.1, we consider the mmWave massive MIMO system for K users. The BS applies a hybrid precoding architecture [4], M_B RF chains and N_B antennas to send the $K \times 1$ transmission signal vector $\mathbf{s}_{\mathrm{B}}$ with transmit power, denoted by P_B, is given by $P_B = \mathbb{E}[\mathbf{s}_{\mathrm{B}}^T \mathbf{s}_{\mathrm{B}}]$, following the power constraint $\bar{P}_B$, i.e., $0 \leq P_B \leq \bar{P}_B$ to serve all the K single-antenna users. The moment when the K users receive the signals, they estimate the channel quality and the SINR to feed them back to the BS. The transmission cost of the BS is denoted by C_B.

In the mmWave massive MIMO system, there are L_B^k paths for the user k and the BS [14]. The complex gain, the azimuth angle and the elevation angle of departure of the path l with $1 \leq l \leq L_B^k$ for the user k and the BS are denoted by $\alpha_B^{l,k}$, $\phi_B^{l,k}$, $\theta_B^{l,k}$, respectively. We apply an geometric channel model proposed in [15] for user k to depict the characteristics of the mmWave massive MIMO channel which is quite different from that of the conventional wireless networks. Let $\mathbf{a}_{\mathrm{B}}(\phi, \theta)$ denotes the array steering vector that can be calculated through [4], the $N_B \times 1$ mmWave channel vector, i.e., $\mathbf{h}_{\mathrm{B}}^k$, between user k and the BS can be written as

$$\mathbf{h}_{\mathrm{B}}^k = \sqrt{\frac{N_B^k}{L_B^k}} \sum_{l=1}^{L_B^k} \alpha_{\mathrm{B}}^{l,k} \mathbf{a}_{\mathrm{B}}(\phi_B^{l,k}, \theta_B^{l,k}). \tag{1}$$

For simplicity, the channel matrix between the K users and the BS is represented by $\mathbf{H}_{\mathrm{B}} = [\mathbf{h}_{\mathrm{B}}^k]_{1 \leq k \leq K}^H$.

TABLE I
SUMMARY OF SYMBOLS AND NOTATIONS

K	Number of the users
$M_{B/J}$	Number of the RF chains at the BS/jammer
$\mathbf{s}_{B/J}$	$K \times 1$ transmission/jamming signal vector
$N_{B/J}$	Number of the BS/jammer antennas
$P_{B/J}$	Transmit power of the BS/jammer
$C_{B/J}$	Cost coefficient of the BS/jammer
$\mathbf{h}_{\mathrm{B}}^k$	Channel vector between the BS and the user k
$\mathbf{h}_{\mathrm{J}}^k$	Channel vector between the jammer and user k
$\mathbf{H}_{\mathrm{B/J}}$	Channel matrix of the BS/jammer
$\bar{\mathbf{f}}_w^{\mathbf{R}}$	Analog beamformer of the BS on the wth sub antenna array
$\mathbf{F}_{\mathrm{R/D}}$	Analog RF precoder/digital baseband precoder for the BS

A switch and inverter (SI) based hybrid precoding architecture [4] is considered here for the BS to reduce the hardware cost and the energy consumption of the mmWave massive MIMO system. Each RF chain of the BS is only connected to a sub antenna array with $W_B = N_B/M_B$ (assumed to be an integer) antennas instead of all N_B antennas. The $W_B \times 1$ analog beamformer of the BS on the wth sub antenna array is denoted by $\bar{\mathbf{f}}_w^{\mathbf{R}}$ with $1 \leq w \leq W_B$. Let $\mathbf{F}_{\mathrm{R}}$ denotes the $N_B \times M_B$ analog RF precoder for the BS. According to the SI-based architecture, $\mathbf{F}_{\mathrm{R}}$ should satisfy the hardware constraints as

$$\mathbf{F}_{\mathrm{R}} = \begin{bmatrix} \bar{\mathbf{f}}_1^{\mathbf{R}} & 0 & \cdots & 0 \\ 0 & \bar{\mathbf{f}}_2^{\mathbf{R}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \bar{\mathbf{f}}_{M_B}^{\mathbf{R}} \end{bmatrix}_{N_B \times M_B} . \tag{2}$$

Since only inverters and switches are used, the N_B nonzero elements of $\mathbf{F}_{\mathrm{R}}$ are set to satisfy the combinational constraint of $\left(-1/\sqrt{N_B}, 1/\sqrt{N_B}\right)$.

The $M_B \times K$ digital baseband precoder for the BS is denoted by $\mathbf{F}_{\mathrm{D}}$ and the kth column of $\mathbf{F}_{\mathrm{D}}$ is denoted by $\mathbf{f}_k^{\mathrm{D}}$. An adaptive cross-entropy based hybrid precoding scheme [4] is proposed here for the BS to calculate $\mathbf{F}_{\mathrm{D}}$ and $\mathbf{F}_{\mathrm{R}}$. The K-dimensional noise vector, denoted by $\mathbf{n}$ is assume to be additive white Gaussian with $\mathbf{n} \in \mathcal{CN}(0, \mathbf{I})$.

B. Jamming Model

The smart jammer is assumed to have the same precoding architecture with the BS and it applies N_J antennas and M_J RF chains to send the jamming signals, denoted by $\mathbf{s}_{\mathrm{J}}$, at the same frequency with the BS in order to block the communication between the BS and the users. More specifically, at time slot n, it chooses the jamming power, denoted by $P_J^{(n)} \geq 0$, based on the ongoing transmission status and the channel states to decrease the SINR of the mmWave massive MIMO system. The jamming cost of the smart jammer is denoted by C_J.

The channel model between the user k and the smart jammer is assume to be a geometric channel model. Similarly, we have L_J^k paths for the users k and the smart jammer. The complex gain, the azimuth angle and the elevation angle of departure for the path i with $1 \leq i \leq L_J^k$ can be denoted by $\alpha_J^{i,k}$, $\phi_J^{i,k}$, $\theta_J^{i,k}$. Therefore, the $N_J \times 1$ mmWave channel vector between the user k and the smart jammer is given by

$$\mathbf{h}_{\mathrm{J}}^k = \sqrt{\frac{N_J^k}{L_J^k}} \sum_{i=1}^{L_J^k} \alpha_J^{i,k} \mathbf{a}_{\mathrm{J}}(\phi_J^{i,k}, \theta_J^{i,k}). \tag{3}$$

The channel matrix between K users and the smart jammer denoted by$\mathbf{H}_{\mathrm{J}}$ is given by $\mathbf{H}_{\mathrm{J}} = [\mathbf{h}_{\mathrm{J}}^k]_{1 \leq k \leq K}^H$.

Similar to the BS, the smart jammer is assume to have the SI based hybrid precoding architecture and applies an adaptive cross-entropy scheme [4] for the architecture to calculate the digital baseband precoder and the analog RF prcoder, denoted by $\mathbf{G}_{\mathrm{D}}$ and $\mathbf{G}_{\mathrm{R}}$. For ease of reference, the commonly used notation is summarized in TABLE 1.

III. Power Control with Reinforcement Learning

In the dynamic mmWave massive MIMO anti-jamming system, on account of the fact that the smart jammer's jamming power is affected by the BS's transmit power, the power control process in the repeated interaction between the BS and the smart jammer could be formulated as an MDP. Thus, the reinforcement learning algorithm could be used to derive the optimal anti-jamming power control strategy via trial and error. In this paper, A hotbooting technique is proposed, which initializes the Q-value based on the training data obtained in advance from large scare experiments in similar scenarios to the reinforcement learning based anti-jamming power control algorithm to improve the performance of resisting smart jamming.

A. Fast Q Based Power Control Algorithm

We propose a fast Q based power control algorithm against jamming. This scheme updates a Q-function and depends on the current state denoted by $\mathbf{s}^{(n)}$ that consists of the SINRs of the signals received by the K users in the last time slot and the previous jamming power.

More specifically, the BS collects the K users' previous SINR, denoted by $\hat{\mathbf{R}}^{(n-1)} = [\hat{\gamma}_k^{(n-1)}]_{1 \leq k \leq K}$ and estimate the previous jamming power, $\hat{P}_J^{(n-1)}$ according to the channel quality and SINRs. The BS chooses the jamming power and all the users' SINR at time slot $n-1$ as the present system state, i.e., $\mathbf{s}^{(n)} = [\hat{P}_J^{(n-1)}, \hat{\mathbf{R}}^{(n-1)}]$. The transmit power $P_B^{(n)}$ at state $\mathbf{s}^{(n)}$ depends on the Q-function $Q(\mathbf{s}^{(n)}, P_B^{(n)})$. On receiving the feedback information from the users, the BS updates the new system state, i.e., $\mathbf{s}^{(n+1)} = [\hat{P}_J^{(n)}, \hat{\mathbf{R}}^{(n)}]$ and evaluates its utility by

$$u_B^{(n)} = \sum_{k=1}^{K} \log_2\left(1 + \hat{\gamma}_k^{(n)}\right) - C_B K P_B^{(n)}. \tag{4}$$

The Q-function is updated in each time as follows

$$\begin{aligned} Q\big(\mathbf{s}^{(n)}, P_B^{(n)}\big) \leftarrow & \big(1-\alpha\big) Q\big(\mathbf{s}^{(n)}, P_B^{(n)}\big) \\ & + \alpha\big(u_B^{(n)} + \delta V\big(\mathbf{s}^{(n+1)}\big)\big) \end{aligned} \tag{5}$$

$$V\big(\mathbf{s}^{(n)}\big) = \max_{P_B \in \Omega} Q\big(\mathbf{s}^{(n)}, P_B\big), \tag{6}$$

where $V(\mathbf{s}^{(n)})$ denotes the maximal Q-function on the feasible actions at state $\mathbf{s}$, $\alpha \in (0,1)$ is the learning factor of the Q-learning algorithm and $\delta \in (0,1)$ is the discount factor indicating the greedy behavior of the BS.

The ϵ-greedy policy is a decent method for the BS to make the tradeoff between exploitation and exploration during the learning process for avoiding being trapped in the local optimal power control strategy and improving the utility. According to the ϵ-greedy policy, the transmit power that maximizes the Q-function is chosen with a high probability of 1 - ϵ and the other power level are chosen with a small probability, where $\epsilon \in (0,1)$ is a small positive value. Thus, the transmit power of the BS is given by

$$\mathbf{Pr}(P_B = \widetilde{P}_B) = \begin{cases} 1-\epsilon & \widetilde{P}_B = \arg\max\limits_{P_B \in \Omega} Q(\mathbf{s}^{(n)}, P_B) \\ \frac{\epsilon}{|\Omega|-1} & \widetilde{P}_B \neq \arg\max\limits_{P_B \in \Omega} Q(\mathbf{s}^{(n)}, P_B), \end{cases} \tag{7}$$

where $|\Omega|$ is the total number of the actions. The fast Q based mmWave massive MIMO power control is summarized in Algorithm 1.

Algorithm 1 Fast Q Based Power Control.

```
Initialize $\alpha, \delta, \epsilon$, $\hat{P}_J^{(0)}, \hat{\mathbf{R}}^{(0)}$, $Q(\mathbf{s}, P_B) = Q^*(\mathbf{s}, P_B)$ and
  $V(\mathbf{s}) = V^*(\mathbf{s})$
for $n = 1, 2, 3, \cdots$ do
    Update the state $\mathbf{s}^{(n)} = [\hat{P}_J^{(n-1)}, \hat{\mathbf{R}}^{(n-1)}]$
    Choose $P_B^n$ with $\epsilon$-greedy policy via (7)
    Transmit the signals with transmit power $P_B^n$ to
    serve all the $K$ users
    Collect SINR $\hat{\mathbf{R}}^{(n)}$ from the $K$ users
    Evaluate the jamming power $\hat{P}_J^{(n)}$ according to the
    channel quality and the SINR $\hat{\mathbf{R}}^{(n)}$
    Evaluate $u_B^{(n)}$ via (4)
    Update $Q(\mathbf{s}^{(n)}, P_B^{(n)})$ via (5)
    Update $V(\mathbf{s}^{(n)})$ via (6)
end for
```

B. Fast PHC Based Power Control

The fast PHC algorithm which keeps not only the Q-function but also the current mixed strategy (strategy table π) to derive the optimal power control strategy is an extension of the fast Q-learning algorithm. On account of the fast PHC algorithm, the BS chooses the transmit power to send the signals according to the observed system state following the probability distribution table π and improve the policy by increasing the probability that it selects the most valuable action according to a learning rate $\theta \in (0, 1]$.

Similar to the fast Q-learning algorithm, the BS uses the estimated jamming power and all the users' SINR at time slot $n-1$ as the current system state, i.e., $\mathbf{s}^{(n)} = [\hat{P}_J^{(n-1)}, \hat{\mathbf{R}}^{(n-1)}]$. On basis of the system state $\mathbf{s}^{(n)}$, the BS chooses the action $P_B^{(n)}$ with probability $\pi(\mathbf{s}^{(n)}, P_B^{(n)})$. Afterwards, the BS calculates all the users' SINR and estimates the jamming power, which are used as the new system state at time slot $n+1$. Thus, the next system state is denoted by $\mathbf{s}^{(n+1)} = [\hat{P}_J^{(n)}, \hat{\mathbf{R}}^{(n)}]$ and the utility of the system is evaluated by (4).

The Q-function and the value function are updated in each time slot by (5) and (6), where the learning factor α and the discount factor δ are chosen to control the learning speed of Algorithm 2. The probability-function $\pi(\mathbf{s}^{(n)}, P_B^{(n)})$ is updated by

$$\pi(\mathbf{s}^{(n)}, P_B^{(n)}) = \pi(\mathbf{s}^{(n)}, P_B^{(n)}) + \begin{cases} \theta & \text{if } P_B^{(n)} = \arg\max\limits_{P_B'} Q(\mathbf{s}^{(n)}, P_B') \\ \frac{-\theta}{|\Omega|-1} & \text{otherwise.} \end{cases} \quad (8)$$

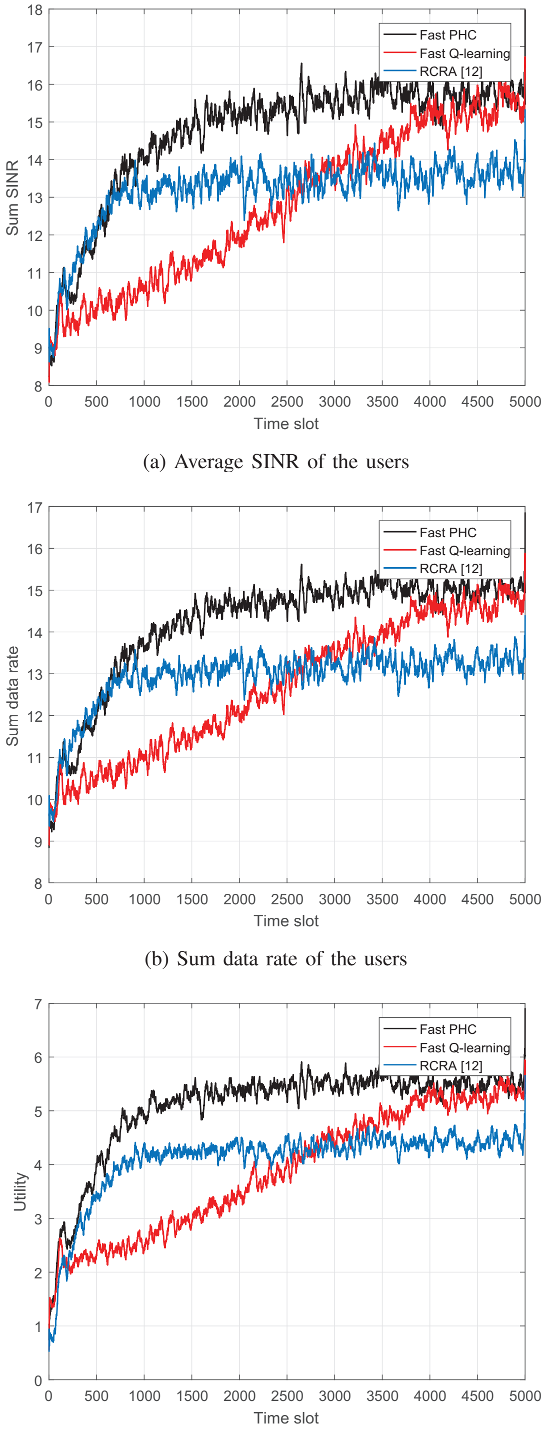
The fast PHC based mmWave massive MIMO power control strategy is summarized in Algorithm 2.

Algorithm 2 Fast PHC based Power Control.

```
Initialize $\alpha, \delta, \theta$, $\hat{P}_J^{(0)}, \hat{\mathbf{R}}^{(0)}$, $Q(\mathbf{s}, P_B) = Q^*(\mathbf{s}, P_B)$, $V(\mathbf{s}) = V^*(\mathbf{s})$ and $\pi(\mathbf{s}, P_B) = \pi^*(\mathbf{s}, P_B)$
for $n = 1, 2, 3, \cdots$ do
    Update the state $\mathbf{s}^{(n)} = [\hat{P}_J^{(n-1)}, \hat{\mathbf{R}}^{(n-1)}]$
    Choose $P_B^{(n)}$ with the probability $\pi(\mathbf{s}^{(n)}, P_B^{(n)})$
    Transmit the signals with transmit power $P_B^{(n)}$ to serve all the $K$ users
    Collect SINR $\hat{\mathbf{R}}^{(n)}$ from the $K$ users
    Evaluate the jamming power $\hat{P}_J^{(n)}$ according to the channel quality and the SINR $\hat{\mathbf{R}}^{(n)}$
    Evaluate $u_B^{(n)}$ via (4)
    Update $Q(\mathbf{s}^{(n)}, P_B^{(n)})$ via (5)
    Update $V(\mathbf{s}^{(n)})$ via (6)
    Update $\pi(\mathbf{s}^{(n)}, P_B^{(n)})$ via (8)
end for
```

IV. Simulation Results

Simulations have been performed to evaluate the mmWave massive MIMO system against smart jamming with a single antennas and the number of transmit antennas ranging between 48 and 256. In the simulations, the BS equiped 16 RF chains and chose the transmit power from 10 levels with tansmission cost 2, to serve 16 users. The learning parameters are chosen as $\alpha = 0.5$, $\delta = 0.85$ and $\theta = 0.05$ to achieve good performance. As a benchmark, the RCRA strategy [13] is evaluated in the same simulation setting with the reinforcement learning algorithm.



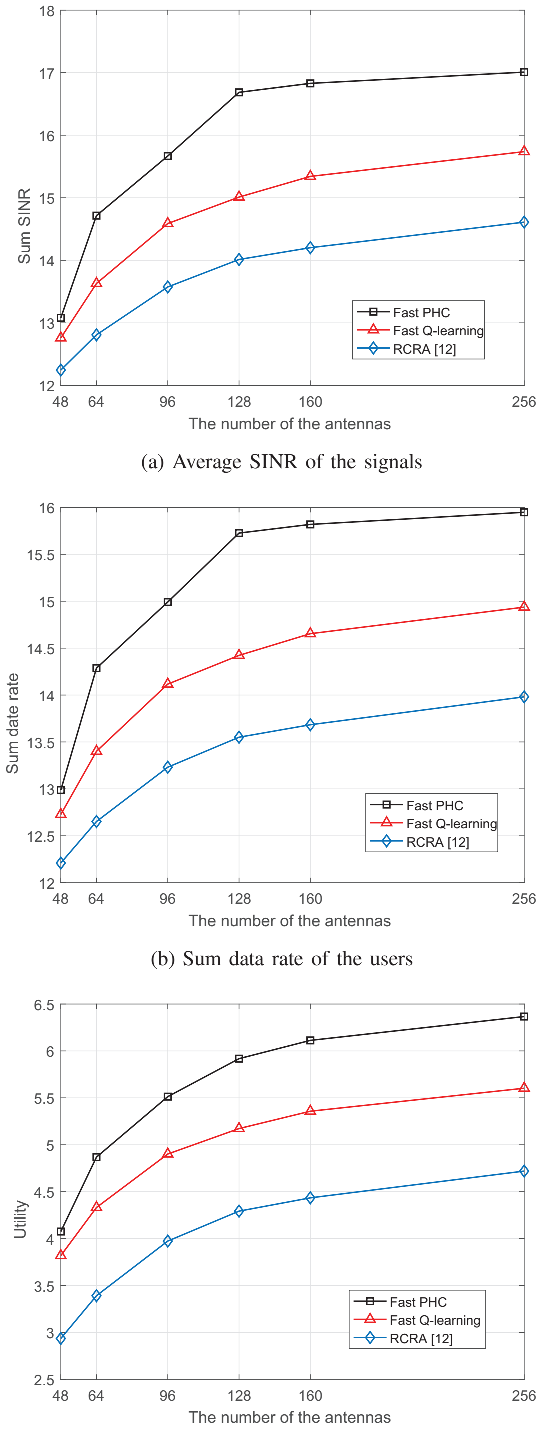
(a) Average SINR of the users

(b) Sum data rate of the users

(c) Utility of the BS

Fig. 2. Performance of the RL based mmWave massive MIMO power control strategies against smart jamming over time with C_B = 1, N_B = 64, M_B = 16 and K = 16.

As shown in Fig.2, the fast PHC based power control strategy which keeps the Q-function and the current mixed strategy exceeds the benchmark scheme RCRA in [13]. For instance, the fast PHC based strategy increases the average



(a) Average SINR of the signals

(b) Sum data rate of the users

(c) Utility of the BS

Fig. 3. Performance of the RL based mmWave massive MIMO power control strategies against smart jamming.

SINR by 14%, improves the sum data rate by 18% and raises the utility to 40% at time slot 1500, compared with RCRA-based power control scheme.

As shown in Fig.3, the average performance over 4000 time slots of the three strategies mentioned above increases with the number of the transmit antennas. In addition, the average performance of the fast PHC power control strategy over different antennas show significant improvement compared to that of the RCRA strategy in [13]. For instance, the average average SINR, sum data rate and utility of the fast PHC based strategy with 256 transmit antennas increase by 31%, 23% and 57%, compared to the system with 48 transmit antennas.

V. Conclusion

In this work, we have proposed a reinforcement learning based mmWave massive MIMO power control scheme to achieve the optimal power allocation policy without being aware of the jamming model and the radio channel model. According to the simulation results, this power control strategy increases the average SINR of the signals, the sum data rate and the utility of the mmWave massive MIMO systems. In addition, the performance gain increases with the number of transmit antennas. For instance, this scheme increases the average SINR. The sum data rate and the utility of the mmWave system with 48 antennas by 16%, 18% and 40%, respectively, compared with RCRA in [13]. The average SINR, sum data rate and utility of the fast PHC further incerase by 31%, 23% and 57%, respectively, for the mmWave systems with 256 antennas.

References

[1] S. Rangan, T. S. Rappaport, and E. Erkip, “Millimeter-wave cellular wireless networks: Potentials and challenges,” *Proc. IEEE Proceedings*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[2] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, “A survey of millimeter wave (mmWave) communications for 5G: Opportunities and challenges,” *Computer Science-Networking and Internet Architeture*, Apr. 2015.

[3] J. C. Li, M. Lei, and F. Gao, “Device-to-device (D2D) communication in mu-MIMO cellular networks,” in *Proc. IEEE Global Commun. Conf*, pp. 3583–3587, Anaheim, CA, Dec. 2012.

[4] X. Gao, L. Dai, Y. Sun, S. Han, and I. Chih-Lin, “Machine learning inspired energy-efficient hybrid precoding for mmwave massive MIMO systems,” in *Proc. IEEE Int'l Conf. Commun. (ICC)*, pp. 1–6, Paris, France, Jul. 2017.

[5] S. Sodagari and T. C. Clancy, “Efficient jamming attacks on MIMO channels,” in *Proc. IEEE Int'l Conf. Commun. (ICC)*, pp. 852–856, Nov. 2012.

[6] Y. Li, L. Shi, P. Cheng, J. Chen, and D. E. Quevedo, “Jamming attacks on remote state estimation in cyber-physical systems: A game-theoretic approach,” *IEEE Trans. Automatic Control*, vol. 60, no. 10, pp. 2831–2836, Oct. 2015.

[7] Y.-C. Tung, S. Han, D. Chen, and K. G. Shin, “Vulnerability and protection of channel state information in multiuser mimo networks,” in *Proc. ACM Comput.Commun. Security. (CCS)*, pp. 775–786, Scottsdale, AR, Nov. 2014.

[8] D. Yang, G. Xue, J. Zhang, A. Richa, and X. Fang, “Coping with a smart jammer in wireless networks: A stackelberg game approach,” *IEEE Trans. Wireless Commun*, vol. 12, no. 8, pp. 4038–4047, Jul. 2013.

[9] 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. Vehicular Technology*, vol. 67, no. 4, pp. 3377–3389, Apr. 2018.

[10] Y. Li, L. Xiao, H. Dai, and H. V. Poor, “Game theoretic study of protecting mimo transmissions against smart attacks,” in *Proc. IEEE Int'l Conf. Commun. (ICC)*, pp. 1–6, Paris, France, May 2017.

[11] Sinno, J. Pan, and Q. Yang, “A survey on transfer learning,” *IEEE Trans. Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[12] M. Bowling and M. Veloso, “Rational and convergent learning in stochastic games,” in *Proc. Int’l Joint Conf. Artificial Intell*, pp. 1021–1026, Seattle, WA, Aug. 2001.

[13] R. Cai, D. Liu, Q. Chen, and X. Peng, “Optimal SINR-based scheduling in mmWave WPANs with power control and rate adaption,” in *Proc. IEEE Vehicular Technology Conference*, pp. 1–6, Glasgow, UK, May 2015.

[14] A. Garnaev and W. Trappe, “The eavesdropping and jamming dilemma in multi-channel communications,” in *Proc. IEEE Int’l Conf. Commun. (ICC)*, pp. 2160–2164, Budapest, Hungary, Jun. 2013.

[15] T. S. Rappaport, R. W. Heath Jr, R. C. Daniels, and J. N. Murdock, *Millimeter wave wireless communications*. USA: Prentice-Hall, Sep. 2014.