Enhancing Blind Interference Alignment with Reinforcement Learning

Simon Begashaw, Danh H. Nguyen, and Kapil R. Dandekar
Drexel University, Philadelphia, PA. Email: {sgb42, dnguyen, dandekar}@drexel.edu

***Abstract*—Blind interference alignment (IA) is a signaling scheme that suppresses interference in multi-user systems, without the knowledge of channel state information at the transmitter (CSIT). The key to performing IA without CSIT is the use of reconfigurable antennas (RA) that are capable of dynamically switching among a fixed number of radiation patterns to introduce artificial fluctuations in the channel. The radiation patterns used to realize blind IA have significant impacts on the overall performance of the system. Hence, an intelligent antenna pattern selection strategy is a crucial component of any practical RA-based blind IA implementation. In this work, we propose two reinforcement learning algorithms for selecting the optimal antenna configuration for blind IA. Furthermore, we evaluate the performance of these antenna mode selection techniques using over the air measurements on our software defined radio implementation of blind IA using a Reconfigurable Alford Loop Antenna that is capable of generating multiple radiation patterns. We quantify the performance of the algorithms in terms of received signal to interference and noise ratio (SINR) and show that our learning-based mode selection strategies are capable of choosing the highest performing mode 90% of the time and attain over 2 dB gain in SINR over other selection approaches.**

I. Introduction

Due to the increasing size and density of modern wireless networks, interference has become a crucial problem that limits the capacity of multi-user systems. First proposed in [1], interference alignment (IA) has emerged as a promising technique for mitigating multi-user interference and achieving significant increase in capacity over traditional orthogonal schemes. Although the capacity benefits of IA are substantial, the assumption of accurate, and sometimes global, channel state information at the transmitter (CSIT) fails in practice due to feedback delay and large overhead requirements [2].

To overcome these challenges, blind interference alignment, which does not require CSIT, was first proposed in [3] and later expanded upon in [4]. The blind IA scheme proposed in this paper exploits the staggered block fading nature of the wireless channel for each link to perform alignment. In [5], [6], staggered antenna state switching with reconfigurable antennas (RA) was proposed as a way to artificially create temporal correlations, enabling blind interference alignment. Since the RA-based blind IA scheme was first proposed, a number of papers have come out to validate and expand upon the scheme. However, most of these papers have been theoretical or simulation based and there are only a few experimental evaluations [7]–[10] of blind IA schemes in the literature. The first practical implementation of blind IA, described in [7], compares the achieved throughput and BER performance to TDMA for a two user X channel. Another experimental evaluation that compares a blind IA scheme against Linear Zero Forcing Beamforming (LZBF) is presented in [8]. Both of these studies simulate the behavior of RA using two spatially separated conventional antennas rather than actually employing RA in their experiments. In [9], the performance of blind IA using ESPAR antennas is investigated and the authors show improved performance in terms of ergodic sum rate and BER. This work also relies on simulation of the antenna and not measurements obtained using the ESPAR antenna. Furthermore, the authors do not address how the beams are selected in their system. In our previous work [10], we provided the first experimental evaluation of a RA system for blind IA.

Since the RA-based blind IA scheme was first proposed in [3], [5], the focus has been on using RA to generate channel fluctuations, with no attention paid to the specific antenna modes selected by the RA and how those modes may impact performance. Consequently, all of the existing literature in blind IA has followed the same model and not considered the antenna modes that are selected but only focused on the staggered switching pattern. While this model is acceptable in simulation-based studies or implementations of blind IA where multiple "dumb" antennas are used to model a RA, it is an inadequate model for practical implementations of blind IA with RA. In RA-based blind IA implementations, the specific antenna configurations used to create the artificial channel fluctuations have to be intelligently selected because they can impact the performance of the system or determine if interference alignment can be achieved at all. In this work, we consider the problem of how to select modes to enhance the performance of practical implementations of blind IA. The main contributions of this paper are as follows:

- two reinforcement learning techniques for selecting antenna modes to enhance the performance of RA-based blind IA systems.
- performance evaluation of the antenna mode selection strategies using over-the-air (OTA) measurements under different channel conditions.

The rest of this paper is organized as follows. In the next section, we present the system model and RA-based blind IA scheme. Section III provides a discussion of the proposed antenna selection techniques, while Section IV describes our implementation. The performance evaluation and discussion of results is presented in Section V and concluding remarks are provided in Section VI.

978-1-5090-1328-9/16/$31.00 ©2016 IEEE

Time Slot	Tx Ant 1	Tx Ant 2	Rx1 Ant State	Rx2 Ant State	Rx1 Signal	Rx2 Signal
1	$x_1 = u_1^1 + u_1^2$	$x_2 = u_2^1 + u_2^2$	s_i^1	s_i^2	$h_1^1(s_i^1)x_1^T + h_2^1(s_i^1)x_2^T$	$h_1^2(s_i^2)x_1^T + h_2^2(s_i^2)x_2^T$
2	$x_1 = u_1^1$	$x_2 = u_2^1$	s_j^1	s_i^2	$h_1^1(s_j^1)x_1^T + h_2^1(s_j^1)x_2^T$	$h_1^2(s_i^2)x_1^T + h_2^2(s_i^2)x_2^T$
3	$u_1 = u_1^2$	$x_2 = u_2^2$	s_i^1	s_j^2	$h_1^1(s_i^1)x_1^T + h_2^1(s_i^1)x_2^T$	$h_1^2(s_j^2)x_1^T + h_2^2(s_j^2)x_2^T$

TABLE I: Blind IA for two user 2×1 MISO-BC

II. Background

A. Signal Model

Consider a system model for a $N = 2$ user multiple-input-single-output broadcast channel (MISO-BC) scenario where the base station is equipped with M traditional antennas, and the N users are each equipped with a single RA that can switch among S antenna states. Unlike most other blind IA research that assumes that $S = M$, we assume $S > M$ because it is more representative of modern RAs and motivates the need for antenna state selection. Let $\mathbf{h}^n(s) \in \mathbb{C}^{2\times 1}$ denote the $1 \times M$ channel vector associated with the s-th state of user n's RA. In developing the system model, the blind IA literature typically assumed that the channel vectors are generic [3], [5], drawn from a continuous distribution, so that any M of them are linearly independent almost surely. In reality, the number of linearly independent channel vectors depends on factors such as the amount of scattering and reflection in the multipath environment and the radiation patterns of the antennas.

With an RA-based blind IA scheme, the receivers switch between their antenna configurations in a predetermined pattern. At time t, the antenna state selected by receiver n is represented by $s^n(t)$ and the corresponding channel for the user is denoted $\mathbf{h}^n(s^n(t))$. Under this model, if the signal vector $\mathbf{x}(t) \in \mathbb{C}^{M\times 1}$ is sent from the transmitter, the received signal at user n is

$$y^n(t) = \mathbf{h}^n(s^n(t))\mathbf{x}(t) + z^n(t) \tag{1}$$

where $z^n(t)$ represents additive white Gaussian noise with zero mean and unit variance. The channel input is subject to an average power constraint $\mathbb{E}\left[||x||^2\right] \leq P$. We assume that the transmitter does not have knowledge of the channel coefficients or the antenna modes selected by the receiver. However, we do assume that the antenna switching pattern is known to the transmitter, since they are predetermined by design.

B. Blind IA With Reconfigurable Antennas

In this section, we review the RA-based blind IA scheme first proposed in [5] and experimentally validated recently in [10]. The objective in blind IA is to construct signals intended for K different users, such that at each receiver, the signals intended for that user remain distinct while the interference (the signals intended for the remaining users) is aligned to an orthogonal dimension. The key to achieving this goal with RA is the design of the antenna switching pattern, the corresponding beamforming strategy at the transmitter, and the intelligent selection of antenna modes at the receiver. The antenna switching pattern and the symbol extension period over which this switching occurs is commonly referred to as a supersymbol structure in the blind IA literature. Using the $N = M = 2$ MISO-BC scenario, we now show the design of the supersymbol structure, and the transmit beamforming strategy. The antenna mode selection strategies will be presented in Section III. For the two user 2×1 MISO-BC, the goal is to achieve two degrees of freedom (DoF) for each user over three symbol extensions. To achieve this goal, the transmitter sends two independent signal streams, each carrying one DoF to each user over a supersymbol. The receivers use a staggered antenna switching pattern in receiving each of the symbols in the supersymbol. We assume that the coherence times of the channels are long enough so that the channels stay constant across a supersymbol. This assumption was verified through extensive experimental measurements in [7], [10]. For one supersymbol, Table I shows the transmitted signal vectors, the selected Rx antenna states and the received signals for both users.

To obtain an interference free signal, receiver 1 can use the interference received in the third slot and subtract it from the first slot as shown below:

$$\begin{bmatrix} y^1(1) - y^1(3) \\ y^1(2) \end{bmatrix} = \begin{bmatrix} h_1^1(s_i^1) & h_2^1(s_i^1) \\ h_1^1(s_j^1) & h_2^1(s_j^1) \end{bmatrix} \begin{bmatrix} u_1^1 \\ u_2^1 \end{bmatrix} + \begin{bmatrix} z^1(1) - z^1(3) \\ z^1(2) \end{bmatrix} \tag{2}$$

where $h_m^n(s_k)$ represents the coefficient associated with the channel from the m-th antenna of the transmitter to receiver n when the antenna state s_k of the RA is selected. Based on our earlier assumption about the channel vectors being linearly independent, user 1 is able to access a full rank channel matrix and therefore can resolve the symbols intended for it and achieve 2 DoF as shown in (2). By symmetry, user 2 can follow a similar procedure and cancel out its interference received in the second slot to also achieve 2 DoF, so that a total of 4 DoF are achieved over 3 symbol extensions.

III. Antenna Mode Selection

Recent advances in antenna technologies have increased the offerings of compact smart antennas with large numbers of available radiation patterns [11]. To realize blind IA using modern RAs, the receiver needs to select suitable modes, a subset of antenna states out of all possible combinations of the available radiation patterns, that exhibit high performance. For the $K = M = 2$ MISO BC scenario described in the previous section, each user needs to independently select antenna state pair (s_i, s_j), out of $\binom{S}{2}$ combinations. The sheer size of the possible antenna mode search space for blind IA, coupled with potential performance degradation caused by imperfect CSIR, renders exhaustive search prohibitive and motivates the need for more tractable solutions. The performance of any antenna mode is governed by the wireless channel. Changes in the channel due

to the mobility of devices or surrounding objects is certain to affect the relative performance of the radiation patterns used for blind IA. In addition, as the number of network nodes increases, the nature of co-channel interference is less predictable and more varied, rendering previously favorable modes suboptimal over time. It is, therefore, essential that any practical antenna mode selection approach for blind IA can quickly identify favorable modes with minimal training overhead and rapidly adapt if other modes show superior performance.

We present a sequential learning framework to achieve adaptive selection of high-performing Rx antenna modes for blind IA with minimal training overhead. We pose the Rx antenna mode selection process for blind IA as a multi-arm bandit (MAB) problem and propose two well-known approaches to solving it based on reinforcement learning: *Upper Confidence Bound (UCB)* [12] and *adaptive pursuit* (AP) [13]. For a more extensive analysis of the same learning framework for antenna state selection under different scenarios, we refer the reader to prior studies by Gulati et al. [14] (using UCB for MIMO transmission) and Nguyen et al. [15] (using AP for directional cognitive networking).

A. Multi-Armed Bandit Approach

In the classic multi-armed bandit formulation [16], the blind IA receiver operates in an environment with incomplete CSIR for all possible state pairs. For each supersymbol in a sequence of trials, the receiver is repeatedly faced with $\binom{S}{2} = K$ arms or choices $\{a_i\}, i = 1, \cdots, K$, each representing a possible antenna state pair to be used for blind IA. At each supersymbol time index t, the receiver selects to play an arm and receives a stochastic reward $R(t)$, which in this work we identify as the post-processing signal to interference and noise ratio ($\mathrm{SINR_p}$) (see Sec. IV-C). The receiver's goal is to maximize the sum of collected rewards at the end of T rounds, $\sum_{t=1}^{T} R(t)$. Note that the nature of the random reward for each antenna state pair is unknown to the blind IA receiver *a priori*.

Besides the cumulative reward, a MAB selection policy can also be evaluated in terms of *regret* [17], defined as the policy's expected loss in reward compared to the best possible outcome. Formally, the regret of a policy after l supersymbols is defined as

$$\theta(l) = \mu^* \cdot l - \sum_{i=1}^{K} \mu_i \mathbb{E}\left[n_i(l)\right] \tag{3}$$

where μ_i is the mean reward for state pair (arm) a_i; $\mu^* = \max_{1 \le i \le K} \mu_i$ is the optimal mean reward; and $n_i(l)$ is the number of times arm a_i has been played up to supersymbol slot l. $\mathbb{E}[\cdot]$ is the expectation operator.

The Upper Confidence Bound (UCB) selection policy, first proposed by Auer et al. [12] and adapted for antenna state selection in [14], has been shown to achieve the optimal regret growth rate, which is logarithmically bounded over time [17]. This selection policy has multiple variants, and we have chosen to implement two of them, UCB1 and UCB1-Tuned [12] for practical implementation purposes in blind IA. Both of these policies use *deterministic* arm selection rules to bound the regret growth rate. Specifically, the UCB1 policy selects, for the current time step, the arm a_i that maximizes the quantity

$$\bar{\mu}_i(l) + \sqrt{\frac{2 \ln l}{n_i(l)}} \tag{4}$$

where $\bar{\mu}_i$ is the sample mean of all observed rewards for arm a_i up to supersymbol index l.

The UCB-1 Tuned policy adapts UCB-1 for practical implementations by replacing the upper confidence bound (second term in Eq. 4) with a different bound to account for the variance of the reward distributions. Under UCB-1 Tuned, the maximization quantity becomes

$$\bar{\mu}_i(l) + \sqrt{\frac{\ln l}{n_i(l)} \min\left\{\frac{1}{4}, V_i(l)\right\}} \tag{5}$$

where V_i(l) denotes the observed variance of the reward samples $\mu_{ik}(l)$ for arm i after l supersymbols:

$$V_i(l) = \left(\frac{1}{n_i(l)} \sum_{k=1}^{n_i(l)} \mu_{ik}^2(l)\right) - \bar{\mu}_i^2(l) + \sqrt{\frac{2 \ln l}{n_i(l)}}$$

While UCB-1 Tuned has been shown to work well in practice, no mathematical proof for its regret bound exists in the literature. Nevertheless, we include it in this work for a complete coverage of the practical UCB policies.

B. Adaptive Pursuit Approach

Prior work in applying MAB techniques to optimize antenna state selection [14], [18] often overlook the issue of non-stationary environments, wherein the reward distributions change their properties over time. For the non-stationary bandit problem, pursuit methods have been shown to be well-suited to track environmental changes while remaining susceptible to fine-tuning to improve performance [16].

The adaptive pursuit (AP) strategy, originally proposed for learning automata [13] and adapted for antenna state selection in [15], is a *probabilistic* selection policy. This method identifies at each supersymbol time step t the suitable selection probability $P_i(t)$ for every antenna mode (arm) a_i to be used for interference alignment, with the objective to maximize the expected cumulative reward at the end of the run. The arms' selection probabilities are specified in an *operator probability vector*

$$\mathbf{P}(t) = \left[P_i(t)\right], \qquad i = 1, \cdots, K$$

where $0 \le P_i(t) \le 1$ and $\sum_{i=1}^{K} P_i(t) = 1$. For its operations, the adaptive pursuit algorithm maintains an *operator quality vector*

$$\mathbf{Q}(t) = \left[Q_i(t)\right], \qquad i = 1, \cdots, K$$

that keeps a running estimate of the reward $Q_i(t)$ for each arm. Whenever an antenna state pair a_i is selected, its current reward estimate is updated with the corresponding received reward from the environment $R(t)$, using a weighted averaging mechanism:

$$Q_i(t+1) = (1-\alpha) Q_i(t) + \alpha R(t)$$

where the adaptation rate α, $0 \leq \alpha \leq 1$, controls the memory of past reward estimates.

At each supersymbol time index t, the AP method *pursues* the arm a_{i^*} that currently has the maximum estimated reward $Q_{i^*}(t)$ by favoring its selection probability over others in the next supersymbol time step:

$$P_{i^*}(t+1) = P_{i^*}(t) + \beta\left[P_{max} - P_{i^*}(t)\right] \tag{6}$$

where parameter β determines the convergence rate toward the maximum exploitation percentage P_{max}. Meanwhile, the algorithm also maintains a minimum selection probability P_{min} for all other arms to enforce mandatory exploration and agility to environmental changes. The operation selection probabilities for all other arms are updated for the next time slot as follows:

$$P_i(t+1) = P_i(t) + \beta\left[P_{min} - P_i(t)\right], \ \forall i \neq i^* \tag{7}$$

Finally, to ensure that all selection probabilities add up to 1, the following constraint is enforced:

$$P_{max} = 1 - (K-1)P_{min}$$

The AP algorithm is a highly versatile solution that can adapt well to changing reward environments, which can be more suitable for dynamic wireless networks.

C. Pattern Correlation Approach

One additional approach to antenna mode selection is to use the spatial correlation between the different radiation patterns of the antenna. The conventional wisdom is that uncorrelated radiation patterns lead to uncorrelated channels in rich multi-path environments [19]. The pattern correlation coefficient (PCC) provides a measure of antenna diversity performance. The PCC ρ between radiation patterns corresponding to antenna states i and j is defined in [19] as:

$$\rho_{i,j} = \frac{\int_{4\pi} E_j(\Omega) d\Omega E_i^{\dagger}(\Omega)}{\sqrt{\int_{4\pi} |E_i(\Omega)|^2 d\Omega \int_{4\pi} |E_j(\Omega)|^2 d\Omega}} \tag{8}$$

where $E_i(\Omega)$ is the radiation pattern of the i^{th} state and $\dagger$ denotes a Hermitian transpose. After calculating the pattern correlation coefficients for each pair of antenna states, the selection strategy would be to choose antenna state pairs (s_i, s_j) corresponding to the minimum correlation coefficient $\rho_{i,j}$.

D. Periodic Exhaustive Search

For completeness in our evaluations of antenna state selection methods, we consider an impractical approach wherein the performance outcomes of all antenna states are periodically measured to guide selection. In the periodic exhaustive search (PES) scheme, our measurement procedure sweeps through all available state pairs to determine their performance during their training phase. Then, the optimal mode is selected for use until the next training round. The granularity of training during operations determines a PES scheme's agility to environmental changes. In our implementation of PES, we used a training interval T = 300, which means a round of training was carried out every 300 packets.

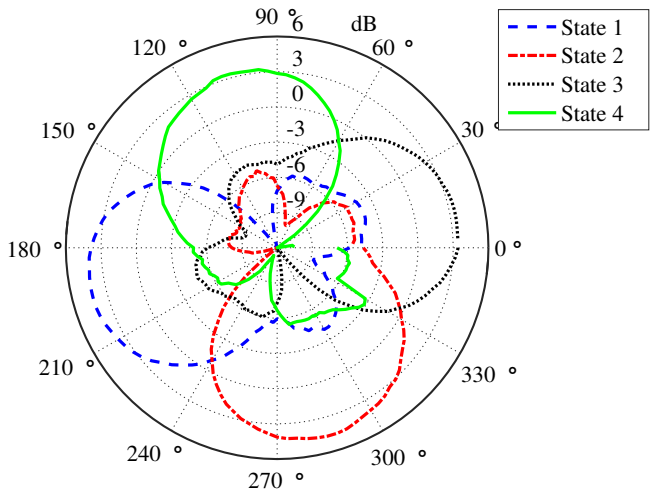

Fig. 1: Four directional radiation patterns of the RALA

	State 1	State 2	State 3	State 4
State 1	1	0.24	0.30	0.28
State 2	0.24	1	0.31	0.26
State 3	0.30	0.31	1	0.32
State 4	0.28	0.26	0.32	1

TABLE II: Pattern correlation coefficients between different states of RALA

IV. Implementation

A. Reconfigurable Alford Loop Antenna

The antenna employed in this work, the Reconfigurable Alford Loop Antenna (RALA) [20], is capable of generating both directional and omni-directional radiation patterns by switching between the radiating elements. The layout of the antenna consists of four pairs of 90 degree microstrip elements arranged symmetrically between the top and bottom layer of a standard FR-4 substrate. Each of the four pairs of branches are connected to the central feed port with PIN diodes. When all branches are connected to the feed port, the antenna exhibits an omni-directional radiation pattern with horizontal polarization. Alternatively, four directional radiation patterns with 90 degree spacing can be achieved by connecting just one pair of branches. Additional directional and bi-directional beams could be obtained by exciting different combinations of elements and, altogether, this antenna is capable of generating eleven different radiation patterns. In our study, we only focus on the four directional radiation patterns displayed in Fig. 1, giving each receiver $\binom{4}{2} = 6$ possible combinations to choose from. The pattern correlation coefficients for the four antenna states used in our study, depicted in Fig. 1, are calculated using (8) and listed in Table II.

B. RALA-Based Blind IA Implementation on WARP

We now provide a brief description of our implementation of the blind IA scheme using the RALA for the two user 2×1 MISO-BC case. A detailed explanation of this implementation

can be found in our prior work [10]. Our earlier blind IA implementation did not include an antenna state selection strategy. For the experiments carried out in that work, extensive channel measurements were performed over all the possible antenna states to determine which modes would be suitable for blind IA in each experiment configuration/location, whereas in this paper we consider the performance of practical antenna state selection techniques.

Our implementation was carried out using the WARPLab [21] framework and the WARP v3 [22] SDR platform. The system had an OFDM based physical layer with a bandwidth of 20 MHz using 64 subcarriers, with 48 subcarriers used for payload. For symbol-level alignment, antenna switching needed to occur in real-time at the OFDM symbol level. To enable WARPLab to carry out low-latency operations, such as switching antenna states in the middle of packet reception, we augmented WARPLab's sample buffer system with custom FPGA signal processing. In our blind IA implementation, the transmission scheme follows the procedures shown in Table I. By default, both user 1 and user 2 have the omni-directional state of their antenna selected to facilitate packet detection. Upon detecting a packet, user 1 will select directional states (s_i^1, s_j^1, s_i^1) to receive the 3 OFDM payload symbols that constitute a supersymbol. User 2, on the other hand, receives the first 2 symbols in antenna state s_i^2 and switches to state s_j^2 for the third slot. The radiation patterns s_i and s_j are selected independently for each user by the algorithms described in Section III. The receivers then performs OFDM demodulation, interference cancellation, and aligned symbol detection.

C. Reward Metrics

The goal of antenna mode selection in blind IA is to improve the overall system performance and the reward metric used for learning the optimal mode should be an accurate measure of the overall system performance. Additionally, this reward metric should be obtained easily, without significant computation and processing delay, to allow real-time implementation in wireless networks. The authors in [14] have identified post-processing signal to interference and noise ratio (SINR_p) as a quality metric that satisfies the aforementioned criteria. SINR_p is an approximation of SINR using the inverse of the average error vector magnitude squared. In our OFDM-based implementation where two signal vectors are sent to each receiver, the instantaneous reward R_i is the average SINR_p over all subcarriers and signal vectors and is calculated as:

$$R_i = \sum_{i=1}^{2} \sum_{f=1}^{F} \frac{1}{\mathbb{E}[|u_f[i] - \hat{u}_f[i]|^2]} \tag{9}$$

where $u_f[i]$ and $\hat{u}_f[i]$ represent the received and idealized symbols of the i-th stream at subcarrier index f.

V. Results and Analysis

To evaluate the performance of the antenna mode selection techniques, we collected OTA measurements using our experimental setup cycling through all the antenna modes and ran the algorithms offline. There are two significant advantages of using this offline approach. First, since we have the data for all antenna modes, we know the optimal modes at a given time and the corresponding rewards, which allows us to establish an upper bound on the mode selection performance, compute the percentage of time the optimal mode was selected and accurately calculate regret. This approach also allows us to fairly compare the performance of the various algorithms over the same channels, which would not be possible with any other approach since the wireless channel is time-varying. Throughout this section, we use the term "optimal" to refer to the antenna mode with the highest instantaneous average reward (SINR_p). We study the performance of the antenna mode selection strategies under two experimental scenarios:

Scenario 1: The optimal antenna mode stays constant for the runtime of the experiment.

Scenario 2: The optimal mode changes (i.e. another antenna mode becomes optimal) halfway through experiment. This scenario was achieved by physically rotating (horizontally) the Rx antenna by 90°, while maintaining the same Tx-Rx distance.

For the two scenarios, the various antenna mode selection techniques are evaluated in terms of three performance metrics. The first evaluation metric is percent of time (measured in number of packets) that the mode with the maximum SINR_p or within an error margin (0.5 dB) of the best SINR_p is selected over the last 200 received packets. Secondly, the reward performance of the selection techniques is evaluated through the empirical cumulative distribution function (CDF) of the SINR_p that each selection scheme achieves. Finally, we evaluate the mode selection strategies with respect to regret, which is defined in (3) and quantifies the accrued cost of selecting a non-optimal mode over time.

A. Experimental Scenario 1

Optimal Mode Selection Percentage: It is interesting to note that PCC-based selection performs poorly in both scenarios, due to the fact that the antenna pattern correlations do not capture the effects of the channel. Random selection also has poor performance, which validates our assertion that the intelligent

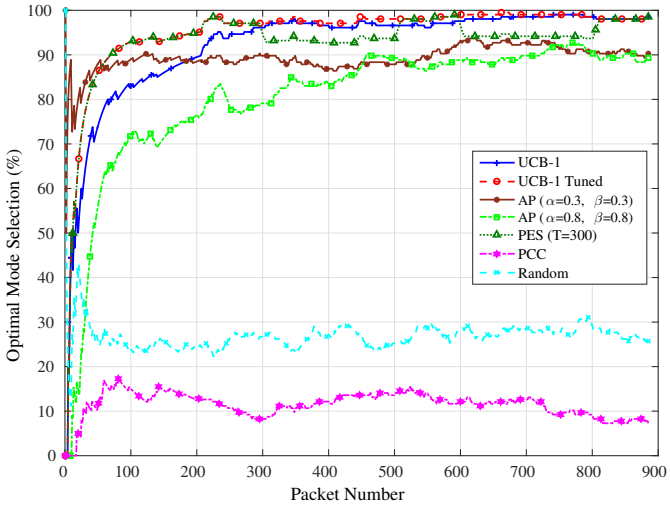

Fig. 2: Optimal Mode Selection Percentage for Scenario 1

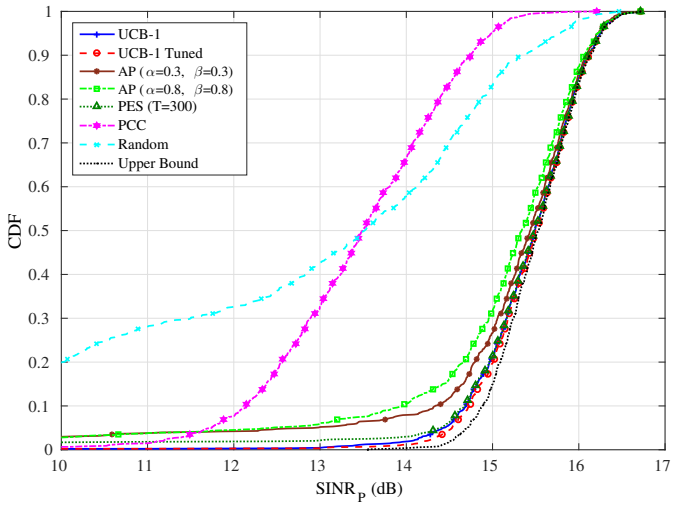

Fig. 3: CDF of $SINR_p$ (Reward) for Scenario 1

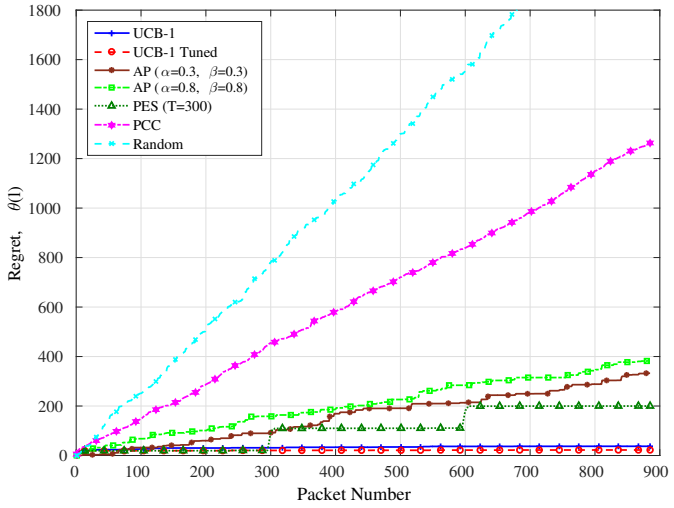

Fig. 4: Regret Growth Over Time for Scenario 1

antenna state selection strategies are needed to enhance blind IA performance. The learning-based selection techniques (UCB-1, UCB-1 Tuned, AP and PES) all show a growth in the percentage of time the best mode is selected as seen in Fig. 2. The best performance is achieved by UCB-1 Tuned and PES in this scenario, as they both show an optimal mode selection rate greater than 95% after only 100 packets. UCB-1 is able to attain the same performance after about 225 packets. Because the AP algorithm implementations have a maximum exploitation probability $P_{max} = 0.9$, they only select the best mode with a rate of 90%. Note that when the learning and adaptation parameter are set to lower values ($\alpha = \beta = 0.3$), the algorithm selects the best mode with approximately 90% probability after about 100 packets. When $\alpha = \beta = 0.8$, the algorithm spends more time exploring other antenna modes than in the previous case, and therefore does not get close to 90% optimal mode selection probability until 450 packets.

Reward: The reward distributions, shown in Fig. 3, match the results observed in the earlier section. The UCB policies and PES exhibit reward distributions very close to the upper bound. The AP algorithms attain a reward performance within 0.5 dB of the upper bound for most of the distribution. Both random and PCC-based selection achieve much worse performance in reward, approximately 2-3 dB degradation in $SINR_p$, compared to the learning-based schemes.

Regret: The superior performance of UCB policies, when the optimal mode is constant, is most evident in their regret performance shown in Fig. 4. It is observed that the regret growth is logarithmically bounded over time for the two UCB policies. While PES had very similar performance to UCB in terms of optimal mode selection percentage and reward distribution, it does not match the regret growth of UCB.

B. Experimental Scenario 2

Optimal Mode Selection Percentage: In the second experiment, the optimal antenna mode changes and a different mode becomes optimal at approximately 1000 packets. Initially, the UCB policies show, in Fig. 5, improved performance in optimal mode selection percentage over the other techniques. Once the best mode changes, UCB policies do not adapt and continue to select the formerly optimal mode. The AP algorithms, in contrast, adapt to the change in optimal mode, selecting the new optimal mode with 90% probability. When the adaptation and learning parameters are set to 0.8, the AP algorithm selects the new optimal mode with 90% probability within 200 packets of the change taking place. With $\alpha = \beta = 0.3$, it takes the algorithm an additional 400 packets to select the new optimal mode with 90% probability. Because the change in optimal mode happens in the middle of the training period for PES, it takes significantly longer to adapt.

Reward: Fig. 6 shows that the AP algorithms once again display the best performance when the best mode is not constant. While the AP algorithms are within 1 dB of the upper bound for 90% of the distribution, the UCB algorithms suffer a $SINR_p$ degradation exceeding 2 dB for over 40% of their distribution.

Regret: The regret performances, in Fig. 7, agree with the previous two results for this experimental scenario. UCB regret

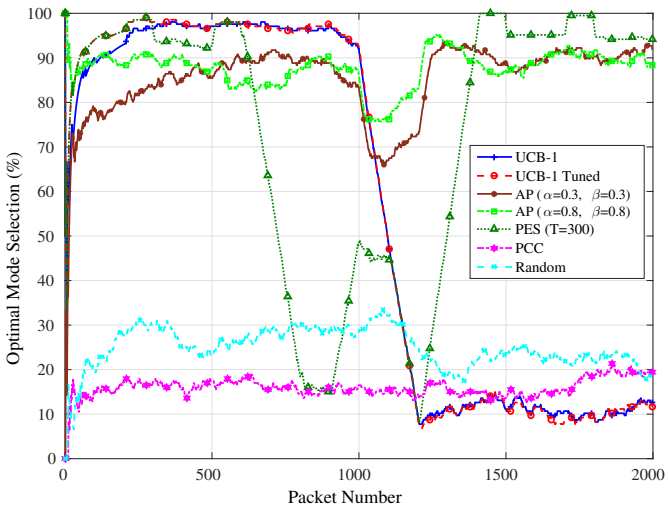

Fig. 5: Optimal Mode Selection Percentage for Scenario 2

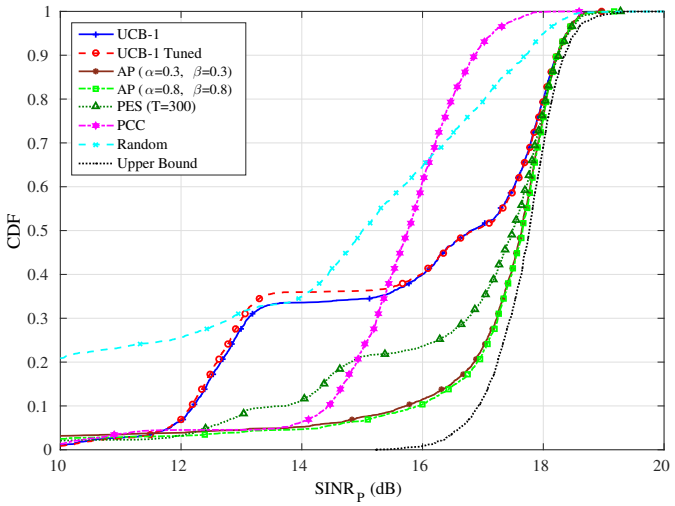

Fig. 6: CDF of $SINR_p$ (Reward) for Scenario 2

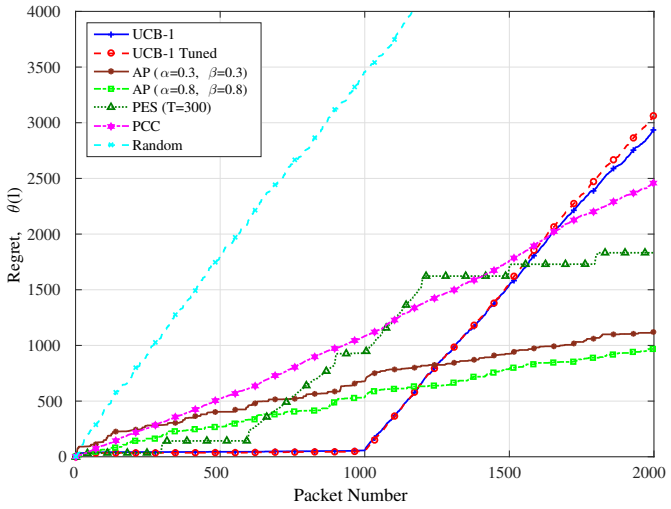

Fig. 7: Regret Growth Over Time for Scenario 2

grows with a logarithmic rate for the first 1000 packets but begins to grow with a steep linear slope once the optimal mode changes. The regret growth for the AP algorithm with higher α and β parameters shows a smaller slope than the implementation with lower values, showing that these parameters should be carefully chosen.

VI. Conclusion

Despite all the attention that blind IA has attracted, most of the literature has focused on theoretical and simulated-based studies. Both the theoretical and experimental studies ignore the impact that the specific radiation patterns of the RA can have on the performance of the blind IA system. In this paper, we proposed two different reinforcement learning approaches to optimal antenna mode selection for blind IA. Using experimental measurements and suitable metrics, we evaluated the performance of a antenna mode selection techniques under two different experimental scenarios. While the UCB-1 and UCB-1 Tuned policy showed superior performances when the optimal antenna mode stays constant, the AP algorithm is able to adjust to changes and select the optimal antenna mode with high probability. This ability to adapt to changes in the wireless channel makes AP more suitable for practical blind IA implementation in wireless networks.

ACKNOWLEDGMENTS

The research presented was based upon work supported by the National Science Foundation Grant No. CNS-1422964.

References

[1] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3425–3441, 2008.

[2] O. El Ayach, S. W. Peters, and R. W. Heath Jr., "The Practical Challenges of Interference Alignment," *IEEE Wireless Communications*, no. February, pp. 35–42, 2013.

[3] S. A. Jafar, "Exploiting Channel Correlations - Simple Interference Alignment Schemes with No CSIT," in *IEEE GLOBECOM*, Dec. 2010, pp. 1–5.

[4] ——, "Blind Interference Alignment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 3, pp. 216–227, Jun. 2012.

[5] T. Gou, C. Wang, and S. A. Jafar, "Aiming perfectly in the dark-blind interference alignment through staggered antenna switching," *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2734–2744, 2011.

[6] C. Wang, T. Gou, and S. A. Jafar, "Interference alignment through staggered antenna switching for MIMO BC with no CSIT," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, no. 1, Nov. 2010, pp. 2081–2085.

[7] K. Miller, A. Sanne, K. Srinivasan, and S. Vishwanath, "Enabling real-time interference alignment," in *Proc. ACM MobiHoc*, 2012, p. 55.

[8] M. M. Cespedes, M. S. Fernandez, and A. G. Armada, "Experimental Evaluation of Blind Interference Alignment," in *IEEE 81st Vehicular Technology Conference*, May 2015, pp. 1–5.

[9] R. Qian and M. Sellathurai, "Performance of the blind interference alignment using ESPAR antennas," in *IEEE ICC*, Jun. 2013, pp. 4885–4889.

[10] S. Begashaw, J. Chacko, N. Gulati, D. H. Nguyen, N. Kandasamy, and K. R. Dandekar, "Experimental evaluation of a reconfigurable antenna system for blind interference alignment," in *2016 IEEE WAMICON*, 2016, pp. 1–6.

[11] H.-N. Dai, K.-W. Ng, M. Li, and M.-Y. Wu, "An Overview of Using Directional Anatennas in Wireless Networks," *International Journal of Communication Systems*, vol. 26, no. 4, pp. 412–448, 2013.

[12] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine Learning*, vol. 47, no. 2-3, pp. 235–256, 2002.

[13] D. Thierens, "An adaptive pursuit strategy for allocating operator probabilities," in *Proc. of ACM GECCO '05*, 2005, pp. 385–386.

[14] N. Gulati and K. R. Dandekar, "Learning state selection for reconfigurable antennas: A multi-armed bandit approach," *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 3, pp. 1027–1038, 2014.

[15] D. H. Nguyen, A. Paatelma, H. Saarnisaari, N. Kandasamy, and K. R. Dandekar, "Enhancing Indoor Spatial Reuse through Adaptive Beamsteering," http://wireless.ece.drexel.edu/pdfs/tech_report_2_indoor_adaptive_beamsteering_using_pursuit_methods.pdf, Drexel University, Wireless System Laboratory, Tech. Rep., 2016.

[16] R. S. Sutton and A. G. Barto, *Reinforcement Learning*. Cambridge, MA: MIT Press, 1998.

[17] T. L. Lai and H. Robbins, "Asymptotically Efficient Adaptive Allocation Rules," *Advances in Applied Mathematics*, vol. 6, pp. 4–22, 1985.

[18] A. Mukherjee and A. Hottinen, "Learning Algorithms for Energy-Efficient MIMO Antenna Subset Selection: Multi-Armed Bandit Framework," in *Proc. of EUSIPCO*, 2012, pp. 659–663.

[19] R. Vaughan and J. Andersen, "Antenna diversity in mobile communications," *IEEE Transactions on Vehicular Technology*, vol. 36, no. 4, pp. 149–172, nov 1987.

[20] D. Patron and K. R. Dandekar, "Planar reconfigurable antenna with integrated switching control circuitry," in *The 8th European Conference on Antennas and Propagation*, April 2014, pp. 2737–2740.

[21] "WARPLab".[Online]. Available:http://warpproject.org/trac/wiki/WARPLab.

[22] "WARP Project". [Online]. Available:http://warpproject.org.