Training-Based Antenna Selection for PER Minimization: A POMDP Approach

Sinchu Padmanabhan, Reuben George Stephen, Chandra R. Murthy, and Marceau Coupechoux

***Abstract*—This paper considers the problem of receive antenna selection (AS) in a multiple antenna communication system having a single radio frequency (RF) chain. The AS decisions are based on noisy channel estimates obtained using known pilot symbols embedded in the data packets. The goal here is to minimize the average packet error rate (PER), by exploiting the known temporal correlation of the channel. As the underlying channels are only partially observed using the pilot symbols, the problem of AS for PER minimization is cast into a partially observable Markov decision process (POMDP) framework. Under mild assumptions, the optimality of a myopic policy is established for the 2-state channel case. Also, two heuristic AS schemes are proposed based on a weighted combination of the estimated channel states on the different antennas. These schemes utilize the continuous-valued received pilot symbols to make the AS decisions, and are shown to offer performance comparable to the POMDP approach, which requires one to quantize the channel and observations to a finite set of states. The performance improvement offered by the POMDP solution and the proposed heuristic solutions relative to existing AS training-based approaches is illustrated using Monte Carlo simulations.**

***Index Terms*—Antenna selection, POMDP, myopic policy, finite state Markov chain**

I. Introduction

Antenna selection (AS) is a popular technique for reducing the hardware complexity and cost of a multiple input multiple output system [2]–[6]. In AS, since only a subset of the available antennas is used for transmission/reception, only a small number of the relatively more expensive radio frequency (RF) chains need to be deployed. AS is supported by wireless standards such as IEEE 802.11n [7] and 3GPP Long Term Evolution (LTE) [8]. AS can be employed at the transmitter as well as at the receiver. The focus of this work is on AS for a multiple-antenna receiver with a single RF chain, with the goal of exploiting the knowledge of the temporal correlation in the wireless channel to perform the optimal tradeoff between exploring for new antennas and exploiting the best antenna based on current knowledge.

There has been an enormous amount of research in the area of antenna selection for MIMO systems in the recent years; we refer the interested reader to [2] and [3] for excellent tutorial surveys of the area. Some of the early work assumed perfect channel state information (CSI) at the receiver [9]–[14]. In practice, the channel state is typically estimated using a small number of pilot symbols embedded in the packet, which leads to imperfect knowledge of CSI at the receiver. The impact of imperfect CSI on the performance of AS is studied in [15] and [16], and it is shown that, surprisingly, the diversity order achievable with perfect CSI is still preserved. Other studies related to AS include AS with transmit beamforming [17], AS with analog power estimators [18], and AS with spatial correlation between antennas [19], [20]. Another approach that has been explored in the literature is the use of reinforcement learning techniques (see, e.g., [21]–[23]). Here, the goal is to minimize the regret compared to a policy that always chooses the statistically best antenna. These are applicable when the channel statistics are not known and the policy must be determined solely from the past AS decisions and resulting outcomes. In this work, we focus on receive antenna selection in the spatially uncorrelated channel case, but we will also briefly indicate how the approach easily allows one to incorporate the effect of spatial correlation between antennas.

Typically, in order to perform AS, the receiver first requests for an AS training phase, following which the transmitter sends out $L \geq 1$ sets of $N \geq 1$ known training symbols to the receiver [24], where N is the number of receive antennas. The time duration between consecutive pilots is $T_p \triangleq \eta T_s$, where T_s is the pilot symbol duration and $\eta \geq 2$ [25].[1] Thus, the total AS training duration is $\eta N L T_s$. The AS training phase is repeated whenever the channel estimates get outdated, imposing a non-trivial overhead on the AS based system. In [26], the authors consider the receive AS based on noisy and outdated channel estimates obtained from the AS training phase. They propose scheme for weighting the channel gain estimates that minimizes the symbol error probability (SEP). The channel state estimates obtained during the AS training phase are used for both AS and data decoding purposes.

In many practical systems, there are additional pilots in data phase also, viz. the demodulation reference signals (DM RS), which may be utilized for data decoding [8]. In [25], Saleh et al. take this into account in data decoding, and propose an algorithm for AS, that maximizes the post-processing SNR. A channel prediction method based on Slepian basis expansion, utilizing the CSI from the training phase, is proposed for AS. After selection, estimation of the channel on the selected antenna is done using the DM RS available in the data phase, again based on Slepian basis expansion. These estimates are

S. Padmanabhan is with the Naval Physical and Oceanographic Laboratory, Kochi, India. She was a student at the EE department at the Indian Institute of Science (IISc), Bangalore, India, during the course of this work. Email: sinchup@gmail.com. R. G. Stephen is a Ph. D. student at the Dept. of ECE, NUS, Singapore. Email: reubenstephen@gmail.com. C. R. Murthy is with the Dept. of ECE, IISc, Bangalore, India. Email: cmurthy@ece.iisc.ernet.in. M. Coupechoux is with Telecom ParisTech and CNRS LTCI, Paris, France. Email: marceau.coupechoux@telecom-paristech.fr.

This work has appeared in part in [1].

[1]The pilot symbols are usually embedded in a training packet with physical layer header [26], and are hence spaced several symbols apart.

used for data decoding. However, the CSI obtained from the data phase is not used for making future AS decisions. In [27], AS is formulated in a decision theoretic framework with the aim of maximizing the throughput. A training based selection is assumed, with each frame consisting of an AS training phase, a data packet and an error check information. Information obtained from the error check observation is used in the future AS decisions. However, the channel is assumed to remain constant for the entire frame duration, which may not hold when the channel is fast-varying. Also, the structure of the optimal policy is not analyzed.

In the context of the above, it is pertinent to consider the use of DM RS for AS also, thereby alleviating the dependence of the AS process on the lengthy training phase. As the channel is correlated in time, and since each packet reveals new information about the channel state only on the selected antenna, the AS decisions affect both the immediate packet reception and *future* packet receptions. The AS can thus be viewed as a form of control, as it determines the accuracy of the CSI available at the receiver on the different antennas. Hence, we formulate the AS problem in a decision theoretic framework, where the problem is to sequentially choose an antenna to receive the current packet, based on the history of past actions and observations, with the goal of maximizing a notion of expected long term reward.

The fundamental trade off involved here is as follows. Sticking to a given antenna for as long as its channel *seemingly good* is not optimal in general, since we lose track of the channel on the other antennas, some of which might be in a better state. On the other hand, frequent switching between antennas would result in not fully utilizing the ones that are in good channel states. An optimal policy is the one which balances between the two and achieves the maximum expected long term reward. Now, at any given time, the true states of antennas are not fully revealed to the receiver, i.e., the states are partially observable through the DM RS. Since the action taken by the receiver controls the observability, the problem is cast as a partially observable Markov decision process (POMDP) [28]–[30]. The goal is to obtain an optimal policy for minimizing the average packet error rate (PER).

There are two kinds of POMDP formulations: the finite horizon POMDP and the infinite horizon POMDP. In the former approach, the goal is to minimize the average PER over a fixed (and typically, small) number of packets that are to be received. The infinite horizon POMDP assumes that the data stream is very long, and is therefore convenient for optimizing a long-term reward. In both cases, the CSI is estimated on the selected antenna upon reception of each packet, and AS based on the optimal POMDP solution strikes the right balance between exploration (to find better antennas) and exploitation (of the best antenna in hand). Our POMDP formulation in this paper is valid both the finite and infinite horizon cases. The SARSOP algorithm from the Approximate POMDP Planning Toolkit [31] used to solve the infinite horizon POMDP in this paper can also be used to solve a finite horizon POMDP scenario, by adjusting the stopping criterion. We have, however, opted for the infinite horizon model in the sequel for the following reasons: 1) the long data stream is realistic if we compare a typical file size to the number of bits carried by an individual packet; 2) it is needed to know the time horizon to solve a finite horizon POMDP, an information that is not available at the physical layer in practice; 3) the finite horizon assumption results in non stationary optimal policies (i.e. policies that depend on the time index), which are more difficult to implement in practice. On the other hand, the optimal policy for an infinite horizon problem is known to be stationary [30].

The contributions of our work are as follows. We cast the problem as a POMDP, which allows us to leverage a host of existing approaches to find an optimal AS scheme. Moreover, our proposed approach obviates the need for an expensive AS training phase at the start of each data packet, unlike most of the past work on AS. On the other hand, our method can also exploit an AS training phase, when present. Hence, it can be employed in systems designed based on previous AS approaches as well. For the case when the number of states per antenna is two, with perfect CSI on the selected antenna and positively correlated[2] channels, we show that optimal policy for the AS problem is myopic in nature. An optimal policy, in general, maximizes the *long term* reward, while a myopic policy is designed to maximize only the immediate reward, ignoring the future rewards. In our set-up, a myopic policy selects the antenna by only considering the probability of correctly receiving the current packet. The myopic policy is simpler to compute as well as to implement, compared to a general POMDP solution. We evaluate the average PER performance of different AS policies via Monte Carlo simulations. The results show that, even with imperfect CSI on all antennas and for $N > 2$, the myopic policy offers performance comparable to the POMDP solution. Inspired by this result on the nature of the optimal policy, we propose two heuristic schemes for AS and evaluate their performance. The performance comparison of these schemes, which are based on continuous-valued channel gain, with that of the finite state Markov chain (FSMC)-based POMDP solution gives further insights into the nature of the POMDP solution.

We also compare our results with the weighting scheme proposed in [26], which is based on AS training. Another scheme which picks the antenna with the highest channel gain in the AS training phase for receiving the subsequent packets is also evaluated. We show the proposed scheme outperforms both these schemes. The results highlight the advantage of utilizing the DM RS information for data decoding as well as for AS purposes, in terms optimizing the PER performance. For example, in the case of PER vs. SNR, the PER of the existing schemes exhibits an error floor, whereas, the PER of the proposed scheme decreases monotonically with SNR.

The rest of the paper is organized as follows. In Sec. II, we describe the system model. We develop the POMDP formulation of the AS problem in Sec. III. We establish the optimality of the myopic policy under certain conditions in Sec. IV. We present a discussion about the relative merits of different policies, in terms of performance and computational

[2]A 2-state channel is said to be positively correlated if the state transition probabilities of the channel are such that the transition to the same state has a higher probability than that to the other state.

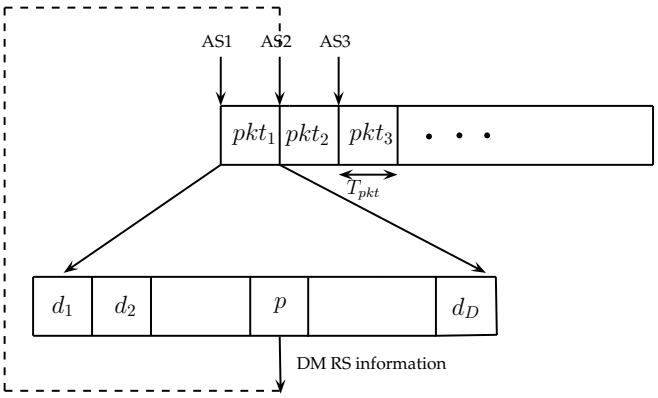


Fig. 1. AS with CSI feedback from packet reception.

complexity, in Sec. V. Monte Carlo simulation results are presented in Sec. VI, and concluding remarks are offered in Sec. VII. Proofs of some of the lemmas and other claims are provided in the appendices.

II. System Model

We consider system where a single antenna transmitter is communicating with a receiver equipped with N antenna elements (AE) and one RF chain. The communication occurs in the form of data packets of duration T_{pkt}. Each packet has D data symbols, denoted by d_j, $j = 1, 2, \ldots, D$ and a DM RS, denoted by p. The goal at the receiver is to select the best out of the N AEs to receive each packet, for minimizing the PER. The channels from the transmit antenna to the receive antennas are modeled as frequency flat, Rayleigh faded and independent across the AEs. The time evolution of the channel follows the Jakes' spectrum [32], [33], with the receiver having the knowledge of the doppler frequency. For simplicity, we assume that the channel remains constant for the duration of a packet. Thus, in this model, the system operates in discrete time steps of duration T_{pkt}. A solid state switch achieves the connection between the selected AE and the RF chain, which has switching speeds on the order of a few hundreds of nanoseconds [25]. Hence, the switching delays are negligible.

The sequence of operations involved in the AS process is depicted in Figure 1. ASi denotes the AS decision for selecting the AE for the i^{th} packet, pkt_i. At the beginning of each packet, the channels make a state transition. The AE selection is based on the CSI available up to and including the previous packet. The DM RS embedded in the packet yields new information on the channel state of the AE that receives the packet. This information is used to decode the packet as well as to update the CSI of the selected AE. With the additional CSI gained in the current packet and the history of decisions and observations, a new selection decision is made for the next packet, and the process continues.

The modeling of the state process is crucial to the solvability of the resulting POMDP. In particular, since the channel state is continuous-valued, a direct formulation of the AS problem in a POMDP framework leads to a continuous state POMDP. On the other hand, quantizing the channel to a finite number of states and using a state transition probability matrix derived from the continuous-valued channel dynamics leads to a discrete state POMDP. There are three main approaches to solving continuous POMDPs: Perseus [34], Monte Carlo POMDP [35] and Monte Carlo value iteration [36]. References [34] and [35] assume that belief value functions are Gaussian or a mixture of Gaussian functions, an assumption which is not supported in our case. Reference [36] uses a particle based representation of the belief but assumes discrete observations, which is also not valid in our case. In contrast, an algorithm like SARSOP is known to solve discrete state POMDPs with up to $1,00,000$ states in a reasonable time, which is sufficient for our purposes. Thus, we propose to pose the problem as a discrete state POMDP. Accordingly, we model the Rayleigh faded time correlated channels as a finite state Markov chain (FSMC) [37], [38], to partition the received SNR on the AEs. Finite State Markov Channel (FSMC) is a popular model for a fading channel, and is known to be accurate for packet-level studies. In this work, we use the popular FSMC model proposed by Zhang and Kassam [39] to partition the instantaneous signal-to-noise ratios (SNRs) on the receive AEs. Let $\mathcal{G} = \{1, 2, \ldots, \kappa\}$ denote the state space of the FSMC channel for a given normalized Doppler frequency $f_m T_{\text{pkt}}$, where f_m is the maximum Doppler frequency. Let $\{\gamma_1, \gamma_2, \ldots, \gamma_{\kappa+1}\}$ denote the SNR thresholds corresponding to the states in $\mathcal{G}$, determined following the procedure in [39]. For a Rayleigh fading channel following the Jakes' spectrum for time variation, the state transition probability matrix of the FSMC as a function of the normalized Doppler frequency is known [39].

We emphasize that, in this work, the instantaneous SNR is discretized into a finite number of states only for the purpose of defining the state space, obtaining the corresponding state transition probabilities, and solving the POMDP. Our formulation can be directly applied to other channel models like Rician or Nakagami fading. The FSMC modeling of these channels are discussed in [40] and [41]. We also note that the formulation directly extends to frequency selective channels also, by using the so-called exponential effective SNR mapping (EESM) metric to convert the frequency-selective wideband channel first into a continuous-valued scalar channel [42], applying one of the above techniques to discretize the channel into a finite set of states. Once the state transition probabilities and the packet error rates for the different states are obtained, the framework developed in this paper can be used to find an optimal AS scheme.

For a POMDP, the statistical information of the system at the time step t, given the entire history of actions and observations, can be captured in a *belief vector* given by $\mathbf{b}(t) = \{b_S(t)\}_{S \in \mathcal{S}}$, where $\mathcal{S}$ is the state space and $b_S(t)$ is the conditional probability, given the history, that the system is in state S at time t. The dynamic behavior of the belief vector is thus a discrete-time continuous-state Markov process [43].

A *policy* for a POMDP is a prescription of an action corresponding to the current belief vector. Each policy has an expected long term reward associated with it. The optimal policy is one which has the maximum expected long term reward. Once the components of the POMDP are defined, a POMDP solver can be used to find the optimal policy for

the problem. At time t, let i denote the AE selected by the policy, and let $h_i(t)$ be the complex valued channel gain of the selected AE. Then, the instantaneous SNR at the receiver is given by $\gamma^{(i)} = |h_i(t)|^2\gamma_0$, where γ_0 is the average per-symbol SNR. If $\gamma_j \leq \gamma^{(i)} < \gamma_{j+1}$, then the AE is said to be in *state* j. The received DM RS on the selected AE, dropping the time index, is given by $y = h_i p + n$, where p is the known pilot symbol and n is the additive white Gaussian noise with variance σ_n^2. The maximum likelihood (ML) estimate of the channel gain is $\hat{h}_i = \frac{p^*}{|p|^2}y = h_i + e$, where e is the noise term, given by $\frac{p^*}{|p|^2}n$. The estimated channel gain $\hat{h}_i$ is used to decode the packet, and also as additional information for selecting the AE for receiving the next packet. The latter is accomplished by updating the belief vector. The optimal policy then maps the updated belief vector to the index of AE to be selected for receiving the next packet. In the next section, we develop the POMDP formulation of the AS problem for minimizing the average PER.

III. POMDP Formulation

The POMDP formulation of the AS problem consists of the following components.

1) State Space: The state space of the system is represented as $\mathcal{S} \triangleq \{1, 2, \ldots, \kappa\}^N$. The i^{th} state is given by the tuple $S_i \in \mathcal{S}$, whose entries specify the channel states on each of the N antennas. When the system makes a transition from state S_i to state S_j, each channel has a corresponding transition associated with it. Since the channels are assumed to be independent, the transition probability $\Pr(S_j|S_i)$ is given by the product of the state transition probabilities associated with each channel.

2) Action Space: The action space is given by $\mathcal{A} \triangleq \{1, 2, \ldots, N\}$ where the i^{th} action corresponds to selecting the i^{th} antenna for packet reception.

3) Observation Space: The observation on selecting an antenna is the received signal corresponding to the DM RS in the packet, which provides information on the channel state of that antenna. Since the CSI from the pilot is continuous-valued, we need to discretize it into states using the thresholds given by the FSMC model. Then, the observation space is $\mathcal{O} = \{1, 2, \ldots, \kappa\}$. Let the observation be $o \in \mathcal{O}$, when the state of the system is S and the action taken is a. Let $S(a)$ denote the state of the selected AE when the system state is S. Then the observation probability $O(S, a, o)$ is the probability of observing state o on the selected AE, given its true state, $S(a)$. The derivation of this probability is given in [1]. It varies with the pilot SNR, and in the case of perfect CSI on the selected AE, $O(S, a, o) = 1$ if $o = S(a)$, the true state of the selected AE, and $O(S, a, o) = 0$ otherwise.

4) Belief Vector: At each time step t, the belief vector $\mathbf{b}(t)$ captures the statistical information of the system. We start with an initial belief vector, $\mathbf{b}(1)$ and update it at each state transition and with each observation. In a training based AS, we can utilize the information from the training phase to obtain an initial belief state. When there is no AS training phase, we can initialize the belief states as the stationary probability of the Markov channel, which is a usual practice when no prior probabilities are available. The AS training phase helps in getting a good estimate of the initial belief state, which speeds up the convergence of the policy. This is beneficial when the channel is slowly varying. On the other hand, for fast varying channels, the initial estimate from the AS training phase is less important. In this case, more frequent DM RS pilots are required for tracking the time-varying channel.

5) Reward: Since we are interested in PER minimization, we define our reward as unity when the packet is correctly received, and zero otherwise. Thus, maximizing the long term reward is equivalent to minimizing the expected average PER.

The expected immediate reward associated with the action $a \in \mathcal{A}$ when the system state is S is given by

$$\varrho(a, S) = \sum_{j=1}^{\kappa} \Pr(o = j|S(a))P_{\text{cor}}(o = j, S(a)), \qquad (1)$$

where $S(a)$ denotes the true state of the selected antenna and o denotes the observation on the selected antenna. $P_{\text{cor}}(o, S(a))$ gives the probability of correctly receiving the packet when the true state is $S(a)$ and the observed state is o. It should be noted that in our case the reward depends on both the true state and the observed state unlike a standard POMDP formulation. The DM RS observation affects the reward as it is used for decoding the data packet. A closed form expression for $P_{\text{cor}}(o, S(a))$, when both observation o and $S(a)$ are discretized values, is analytically intractable. This is because the performance depends on the decoding algorithm used for packet reception, which makes it difficult to come up with a general, closed form expression for the PER under channel mismatch. Moreover, our focus in this paper is on showing how a decision theoretic formulation can be applied to solving the problem of receive antenna selection, rather than on analyzing the PER under channel estimation errors. Hence, $P_{\text{cor}}(o, S(a))$ is calculated experimentally via simulations. The parameters for this simulation will be explained in Section VI. The probability of correctly receiving the packet is calculated for all pairs of true and observation states.

The expected immediate reward can now be expressed as a function of the belief state, $\mathbf{b}$, as follows:

$$R(a, \mathbf{b}) = \sum_{S \in \mathcal{S}} b_S \varrho(a, S), \qquad (2)$$

where b_S is the component of the belief vector $\mathbf{b}$ corresponding to the state S.

6) Objective and the Optimal Policy: The objective is to minimize the expected average PER, over an infinite horizon. The averaging is done in a discounted sense, i.e., the future rewards are discounted by a factor β. A policy is a mapping from the set of all belief vectors to the action space, i.e., a policy has an action corresponding to a given belief vector. There is a reward associated with executing a policy. Let $J_\beta^\pi(\mathbf{b})$ denote the expected total discounted reward associated with a policy π starting from time step $t = 1$ and belief vector $\mathbf{b}$, with the discount factor being β. The optimal policy solves

the following optimization problem

$$\max_{\pi} J_{\beta}^{\pi}(\mathbf{b}) = \max_{\pi} \mathbb{E}\left[\sum_{t=1}^{\infty} \beta^{t-1} R(\pi(\mathbf{b}(t)), \mathbf{b}(t)) | \mathbf{b}(1) = \mathbf{b}\right], \tag{3}$$

where $0 \leq \beta < 1$, and $R(\pi(\mathbf{b}(t)), \mathbf{b}(t))$ is the reward collected under belief state $\mathbf{b}(t)$ when the AE $\pi(\mathbf{b}(t))$ is selected for packet reception.

We have thus formulated the AS problem as a POMDP. There are several tools available for solving POMDPs [44]–[46]. However, solving the POMDP can quickly become computationally burdensome, as the number of states of the system under consideration becomes large. On the other hand, using a smaller number of states compromises on the accuracy of the FSMC model of the underlying continuous-valued channel. A usual practice, in this scenario, is to explore the effectiveness of a simpler but possibly suboptimal policy for AS. A myopic policy is one such policy. In the next section, we show that under the mild assumption of positively correlated channels and perfect CSI on the selected AE, for the 2-states-per-antenna model, the myopic policy is indeed optimal for the AS POMDP problem. Note that, although the 2-state channel model might appear overly simplistic, it retains the essence of the time-variations of the wireless channel, and therefore provides useful intuitions on how to design near-optimal policies for channel models with number of states greater than two. As will be demonstrated through simulations, the myopic policy continues to remain nearly optimal even when the channel is modeled with more than 2 states.

IV. Optimality of the Myopic Policy

A myopic policy is one which maximizes the immediate reward alone, rather than the long term reward. It is oblivious to the impact of current action on future rewards. In this section, for a 2-states channel with perfect CSI on the selected AE and positively correlated channels, we show that the optimal policy is myopic. A pictorial representation of the 2-states channel model with state transition probabilities, is given in Fig. 2. A positively correlated channel is one where the transition probabilities satisfy $p_{11} \geq p_{01}$. This means that in the next time step, the channel state has a higher probability to remain in the present state rather than to switch to the other state. The FSMC model yields a positively correlated channel for normalized Doppler frequencies as high as 0.2. Hence, the channels are positively correlated for all practical purposes. We present the proof of the optimality of the myopic policy for the finite horizon case. However, it can be extended to the infinite horizon case using standard techniques [47]. Unfortunately, the extension of this result to channels with more than 2 states or to the case with imperfect channel estimation does not seem to be straightforward.

The two states per antenna model allows us to simplify the formulation as follows. We redefine the belief vector at time t as $\mathbf{\Omega}(t) \triangleq [\omega_1(t), \omega_2(t), \ldots, \omega_N(t)]$, where $\omega_i(t) \triangleq \Pr(s_i(t) = 1 | \text{Past actions and observations})$, i.e., the conditional probability that the channel i is in the good state (denoted by $s_i(t) = 1$) at time step t, given all past actions and observations. Note that $\mathbf{\Omega}(t)$ differs from $\mathbf{b}(t)$, since, in the former, the belief is on each antenna, whereas in the latter, the belief is on the joint state of the N antennas. Let $a(t)$ denote the antenna selected at time t. Once an AE $a(t)$ is selected, its true channel state $s_{a(t)}$ is revealed by the DM RS. With the new observation on the selected antenna, using Bayes' rule, we update the belief vector as follows:

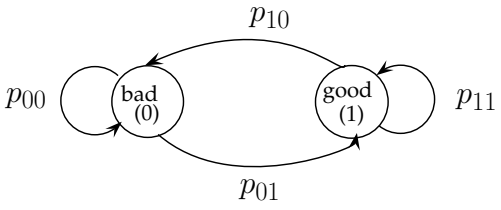


Fig. 2. 2-state model of the channel.

$$\omega_i(t+1) = \begin{cases} p_{11} & \text{if } a(t) = i, s_{a(t)} = 1, \\ p_{01} & \text{if } a(t) = i, s_{a(t)} = 0, \\ \tau(\omega_i(t)) & \text{if } a(t) \neq i, \end{cases} \tag{4}$$

where $\tau(\omega_i(t)) = \omega_i(t)p_{11} + (1 - \omega_i(t))p_{01}$ is the one-step belief update when antenna i is not selected.

We seek to maximize the total expected discounted reward over a horizon of T. That is, we wish to solve

$$\pi^* = \arg\max_{\pi} \mathbb{E}_{\pi}\left[\sum_{t=1}^{T} \beta^{t-1} R(\pi_t(\mathbf{\Omega}(t)), \mathbf{\Omega}(t)) \mid \mathbf{\Omega}(1)\right]. \tag{5}$$

Any admissible policy can be written as $\pi = [\pi_1, \pi_2, \ldots, \pi_T]$, where π_t maps $\mathbf{\Omega}(\mathbf{t})$ to an action $a(t)$; $t = 1, 2, \ldots, T$. Here, policies are indexed by t since the optimal policy for a finite horizon problem is, in general, non-stationary.

We define the *value function* $V_t(\mathbf{\Omega}(t))$ of the optimal policy at time t as

$$V_T(\mathbf{\Omega}) = \max_{a=1,\ldots,N} \mathbb{E}[R(a, \mathbf{\Omega})] \tag{6}$$

$$\begin{aligned} V_t(\mathbf{\Omega}) &= \max_{a=1,\ldots,N} \mathbb{E}\left[R(a, \mathbf{\Omega}) + \beta V_{t+1}(\mathcal{T}(\mathbf{\Omega}))\right] \qquad (7) \\ &= \max_{a=1,\ldots,N} \mathbb{E}[R(a, \mathbf{\Omega})] \\ &\quad + \beta\omega_a(t+1) V_{t+1}(\mathcal{T}(\mathbf{\Omega}|S(a) = 1)) \\ &\quad + \beta(1 - \omega_a(t+1)) V_{t+1}(\mathcal{T}(\mathbf{\Omega}|S(a) = 0)) \qquad (8) \end{aligned}$$

which is the expected sum reward gained, starting in belief vector $\mathbf{\Omega}(t)$, from time t to T. Here, $\mathcal{T}(\cdot)$ is the one-step update operator of the belief vector, defined as in (4). Also, notational simplicity, we have dropped the time index in $\mathbf{\Omega}$.

Let $P_c(s)$ denote the probability of correctly receiving a packet when the channel state is $s \in \{0, 1\}$. Then, the expected immediate reward collected is given by

$$R(a, \mathbf{\Omega}) = \omega_a P_c(1) + (1 - \omega_a) P_c(0) \triangleq f(\omega_a). \tag{9}$$

Since state 1 corresponds to a higher channel gain than state 0, the associated probability of correctly receiving a packet is higher for state 1, and it is reasonable to assume $P_c(1) \geq P_c(0)$. Hence, $f(\omega_a)$ increases linearly with ω_a. A myopic policy chooses that action which maximizes $f(\omega_a)$.

Due to the linearity of $f(\omega_a)$, this is equivalent to choosing the antenna with the highest belief state. Now, we define a pseudo value function $W_t(\mathbf{\Omega})$, $t = 1, 2, \ldots, T$, as follows [48]. We let $W_T(\mathbf{\Omega}) = f(\omega_N)$. For $t < T$, we let

$$W_t(\mathbf{\Omega}) = f(\omega_N) + \beta\left[\omega_N W_{t+1}(\tau(\omega_1), \ldots, \tau(\omega_{N-1}), p_{11}) + (1-\omega_N) W_{t+1}(p_{01}, \tau(\omega_1), \ldots, \tau(\omega_{N-1}))\right]. \quad (10)$$

A few observations on the properties of this pseudo value function are listed below.

1) $W_t(\mathbf{\Omega}(t))$ corresponds to the expected total discounted reward of a policy which chooses, at time t, the AE corresponding to the last entry in $\mathbf{\Omega}(t)$. If a 1 is observed on the selected AE (that is, the channel is observed to be in the good state), then it is selected for receiving the subsequent packets until a 0 is observed on it. As long as a 1 is observed, the next belief vector $\mathbf{\Omega}(t+1)$ remains ordered such that the belief state of the selected AE is the last entry of the vector, i.e., it is the channel to be selected for receiving the next packet also. If the observation is 0 (that is, the channel is observed to be in a bad state), then the AE is moved to be the first entry of the vector, $\mathbf{\Omega}(t+1)$. Correspondingly, it becomes the last one to get selected. The ordering of the unobserved AEs are retained. This can be easily verified by noting the linearity of $\tau(\cdot)$ and the assumption $p_{11} \geq p_{01}$.
2) When the elements in $\mathbf{\Omega}(t)$ are ordered such that $\omega_1 \leq \omega_2 \leq \ldots \leq \omega_N$, $W_t(\mathbf{\Omega}(t))$ is the expected total discounted reward obtained by following the myopic policy from time t to T. This is because, at any time from t to T, the entries in the vector $\mathbf{\Omega}(t)$ remain sorted in increasing order due to the monotonicity of $\tau(\cdot)$. Since the AE corresponding to the last entry is always selected, which has the highest belief state, the policy implemented by selecting the antenna corresponding to the last entry in $\mathbf{\Omega}$ turns out to be the myopic policy.
3) It can be shown that the following *decomposability* property holds for all $l \in \{1, 2, \ldots, N\}$. The proof is by induction, and is relegated to Appendix A.

$$\begin{aligned} &W_t(\omega_1, \ldots, \omega_l, \ldots, \omega_N) \\ &\quad = \omega_l W_t(\omega_1, \ldots, 1, \ldots, \omega_N) \\ &\quad + (1-\omega_l) W_t(\omega_1, \ldots, 0, \ldots, \omega_N). \end{aligned} \quad (11)$$

We can further extend the above result to show that

$$\begin{aligned} &W_t(\omega_1, \ldots, y, x, \ldots, \omega_n) - W_t(\omega_1, \ldots, x, y, \ldots, \omega_n) \\ &\quad = (x-y)[W_t(\omega_1, \ldots, 0, 1, \ldots, \omega_n) \\ &\qquad - W_t(\omega_1, \ldots, 1, 0, \ldots, \omega_n)]. \end{aligned} \quad (12)$$

We will use the above result in the proof of Lemma 1 in the sequel. A necessary and sufficient condition for the optimality of the myopic policy is given in Lemma 2 of [47]. It says, to show the optimality of the myopic policy at time t, given its optimality at $t+1, \ldots, T$, it suffices to show that

$$W_t(\omega_1, \ldots, \omega_{i-1}, \omega_{i+1}, \ldots, \omega_N, \omega_i) \leq W_t(\omega_1, \ldots, \omega_N). \quad (13)$$

for all $\omega_1 \leq \ldots \leq \omega_i \leq \ldots \leq \omega_N$.

The condition given above essentially requires that selecting the AE corresponding to the last entry in the vector, $\mathbf{\Omega}(t)$, followed by myopic selection be better than selecting any other AE followed by myopic selection. With the assumption of optimality of myopic policy from time $t+1$ onwards, this condition ensures that the myopic policy is optimal for time t also. In order to prove that the above condition holds for the AS POMDP problem, we first prove the following. The pseudo value function, $W_t(\mathbf{\Omega}(t))$, does decrease in value, if we switch the order of two neighbouring AEs i and $i+1$ so as to make $\omega_{i+1} \geq \omega_i$. This is established in Lemma 1. Let $\Delta \triangleq f(1) - f(0) = P_c(1) - P_c(0)$. Since $P_c(1) \geq P_c(0)$, we have $\Delta \geq 0$.

Lemma 1. *For $\omega_1 \leq \omega_2 \leq \ldots \leq \omega_N$, the following inequalities hold for all $t = 1, 2, \ldots, T$, when $p_{11} \geq p_{01}$:*

$$1) \quad \Delta + W_t(\omega_2, \ldots, \omega_N, \omega_1) \geq W_t(\omega_1, \ldots, \omega_N) \quad (14)$$

$$2) \quad W_t(\omega_1, \ldots, \omega_l, y, x, \ldots, \omega_N) \geq W_t(\omega_1, \ldots, \omega_l, x, y, \ldots, \omega_N) \quad (15)$$

where $x \geq y$, $0 \leq l \leq N-2$, and $l = 0$ implies $W_t(y, x, \omega_3, \ldots, \omega_N) \geq W_t(x, y, \omega_3, \ldots, \omega_N)$.

We prove this lemma using a sample path argument in Appendix B. Next, we state and prove the theorem on the optimality of myopic policy.

Theorem 1. *The myopic policy is optimal for the problem stated in (6), for $t = 1, 2, \ldots, T$, and $\forall\, \mathbf{\Omega} = [\omega_1, \ldots, \omega_N] \in [0,1]^N$ under the assumption that $p_{11} \geq p_{01}$.*

Proof: The proof is by induction. At time $t = T$, the greedy policy is obviously optimal. Assuming it is optimal for times $t+1, t+2, \ldots, T$, to show the optimality at time t, by Lemma 2 of [47], it suffices to show

$$W_t(\omega_1, \ldots, \omega_{i-1}, \omega_{i+1}, \ldots, \omega_N, \omega_i) \leq W_t(\omega_1, \ldots, \omega_N). \quad (16)$$

By applying (15) repeatedly to the above equation

$$\begin{aligned} &W_t(\omega_1, \ldots, \omega_{i-1}, \omega_{i+1}, \ldots, \omega_N, \omega_i) \\ &\quad \leq W_t(\omega_1, \ldots, \omega_{i-1}, \omega_{i+1}, \ldots, \omega_i, \omega_N) \\ &\quad \leq W_t(\omega_1, \ldots, \omega_{i-1}, \omega_{i+1}, \ldots, \omega_i, \omega_{N-1}, \omega_N) \\ &\quad \ldots \\ &\quad \leq W_t(\omega_1, \ldots, \omega_N). \end{aligned} \quad (17)$$

This completes the proof of Theorem 1. □

The myopic policy explained above has an interesting structure to it. It sticks with an antenna if a 1 is observed on it, otherwise discards that antenna and picks the one with the highest probability to be in state 1. The optimality of such a policy is intuitive. There are only two states and if an antenna in state 1, it is most probable to stay in state 1 in the next time step due to the assumption $p_{11} \geq p_{01}$. On the other hand, if the antenna is in state 0, it has the lowest probability to be in state 1 in the next time step. Hence, by following the myopic policy, the antenna with the highest probability to be in state 1 is selected for receiving the next packet. However, in a general set up with more number of states per antenna, the optimal policy is not straight-forward.

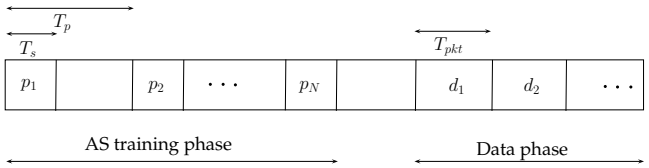


Fig. 3. As training phase followed by the data phase.

V. Discussion

We now briefly discuss the relative merits and demerits of the proposed POMDP approach compared to the existing approaches from the literature. We also present two new heuristic schemes for AS inspired by the existing approaches and the proposed POMDP based approach. Finally, we present a computational complexity analysis of the various schemes.

A. Existing Schemes

The existing schemes for AS are generally based on the use of an AS training phase, shown pictorially in Fig. 3. The AS training phase comprises of N pilot symbols $p_1, p_2, \ldots, p_N$. The pilot symbols are followed by several data packets; we let d_i denote the i^{th} data packet. The receiver obtains estimates of the channel gains of each AE from the pilot symbols in the training phase. These estimates are used in selecting AEs in the data phase. It is assumed that the receiver requests the transmitter for a training phase, when the resulting PER is below some *acceptable* level [24]. In subsection VI-B, we compare the performance of the POMDP solution with two approaches from the literature: the weighting scheme proposed in [26] and the `Max picking` scheme. In both these schemes, the channel gain estimates obtained from the training phase are used for AS as well as for data decoding, i.e., the additional CSI obtained from the DM RS is not exploited. In the `Max picking` scheme, the channel gain estimates of the AEs from the AS training phase are compared, and the AE with the highest estimate is selected for receiving the packets in the data phase. The scheme proposed in [26]is to weigh the estimates from the AS training phase, and select the AE with the highest weighted estimate to receive a symbol. The weights are chosen to minimize the symbol error probability at the receiver, accounting for the fact that the channel estimates on different antennas are outdated by different amounts of time. In order to compare the PER performance of this scheme with that of the POMDP solution, a per-packet selection is done in this paper. Both the above mentioned schemes depend solely on the estimates from the training phase for selecting AEs. We compare their performance with that of our scheme which utilizes the DM RS information in AS decision making. We will show that, by taking the information obtained through the DM RS into account, we are able to significantly improve the AS performance. To facilitate comparison, we consider a given packet index, e.g., the 10^{th} packet after the AS training phase, and illustrate the effect of utilizing the DM RS, as opposed to using only the AS training phase, on the PER performance.

B. Heuristic schemes

In this subsection, we present two heuristic schemes for AS, based on the `Max picking` and the weighting schemes, as follows. In the `Max picking` scheme, once an AE is used to receive a packet, we update the CSI of this AE with the channel gain estimate obtained from the DM RS in the packet. We compare this new estimate with the outdated estimates of other AEs, while selecting AE for receiving the next packet. Similarly, in the modified weighting scheme, the CSI of the selected AE is updated using the DM RS information. For selecting AE for the next packet, the weighting is done on the updated estimate for the selected AE and the outdated estimates of the rest of the AEs. The weight calculation takes into account the delay in the estimates, in the same manner as was done in the original scheme. In addition to this, the channel gain estimate from the DM RS is used to decode the data in the packet, for both the modified schemes.

Recall that the POMDP formulation requires the continuous-valued channel gains to be discretized into a finite number of states for the purpose of solving the POMDP and determining the optimal policy. Due to this, the information available from the received pilot symbols is not fully utilized by the receiver, especially when the number of states per antenna is small.[3] This will be illustrated in in the simulation results in Section VI, where we plot the PER versus the normalized Doppler frequency. However, the heuristic schemes proposed above based on the intuition gleaned from the POMDP utilize the DM RS information not only for data decoding but also for AS, and recover the performance loss of the POMDP solution. In subsection VI-C, we will show that the two heuristic schemes presented above outperform the existing weighting and max picking schemes described in the previous subsection.

C. Computational Complexity

In this subsection, we discuss the computational complexity of the different AS schemes discussed above. We focus on the complexity in using the policy suggested by the POMDP planning, and not on the computational cost of solving the POMDP. This is because the POMDP can be solved offline, and only needs to be repeated when the channel statistics change, i.e., only very rarely, compared to the AS process.

Let κ denote the number of states per antenna and let N denote the number of antenna elements, as before. The total number of channel states is then given by $K = \kappa^N$. We make the following remarks about the complexity:

- SARSOP policy: The output of the SARSOP tool is a set of vectors, denoted by Λ, which represents a piecewise-linear approximation of the optimal value function. Each vector is associated with an action. At each time step, the inner product between these vectors and the current belief vector is computed. The complexity of this computation is $|\Lambda|\mathcal{O}(K)$. The optimal action is the one corresponding to the vector in Λ is most correlated with the current belief

[3] A way to overcome the loss of optimality in quantizing the channel into discrete states is to increase the number of states on each antenna; however, this drives up the complexity of finding the optimal solution. Also, a limitation of the FSMC model in [39] is that it restricts the state transitions to happen only between adjacent channel states, and this affects the optimality of the policy vis-a-vis the behavior of the channel when the number of states becomes large.

vector. The complexity of finding the maximum among $|\Lambda|$ values is $\mathcal{O}(|\Lambda|)$. The last step of the algorithm is to update the belief vector by multiplying it with the state transition probability matrix. This incurs a complexity of $\mathcal{O}(K^2)$.

- Myopic policy: In this case, at each time step, the immediate reward corresponding to each action is calculated. This involves computing the inner product between current belief vector and the expected immediate reward $\varrho(a, S)$, which has a complexity of $N\mathcal{O}(K)$. Then, we find the maximum among the N inner product values, the complexity being $\mathcal{O}(N)$. The final step in the procedure is to update the belief vector, similar to the SARSOP policy, which has a complexity of $\mathcal{O}(K^2)$.
- Weighting scheme: This scheme involves multiplying the channel gain estimates from the AS training phase with the weight vector, incurring a complexity of $\mathcal{O}(N)$. Note that, as the weight vector computation can be performed offline, its complexity is not included in this analysis. Finally, we choose maximum from these N weighted gains. The complexity for this is $\mathcal{O}(N)$.
- `Max picking:` This schemes simply picks the maximum among the N channel gains obtained from the AS training phase. This has a complexity of $\mathcal{O}(N)$. All the packets are received using the same antenna.
- Modified schemes: In the modified schemes, the channel gain estimate of the selected antenna is updated using the DM RS information. Its complexity is similar to the corresponding schemes discussed above.

From the above discussion, we see that the POMDP based solution has a higher complexity than the other schemes. The complexity increases with the number of quantized states. Between SARSOP policy and myopic policy, the former has higher complexity as $|\Lambda|$ is typically much greater than K for a "good" policy obtained from the SARSOP tool. Due to this, the computational complexity of the SARSOP policy may not scale well for large dimensional AS problems. On the other hand, the myopic policy offers a good compromise between complexity and performance, and would be the candidate of choice as the number of antennas gets large, for example, in massive MIMO systems.

VI. Simulation Results

In this section, we present Monte Carlo simulation results to support the analysis and discussion in the previous sections.

A. Simulation Setup

In order to evaluate the performance of the proposed schemes and compare with existing training-based AS schemes, we simulate an initial training phase, followed by transmission of several data packets, as shown in Fig. 3. We have $T_p = T_{\text{pkt}}$, and we assume that the channel stays constant for the duration T_{pkt}, for all the simulations. In case of the POMDP, the information from the training phase is used to compute the initial belief vector. Each packet consists of ten data symbols and one DM RS. The data symbols are drawn uniformly at random from an 8-PSK constellation, scaled by the signal power. We assume that there are $N = 4$ AEs, and that the noise at the receiver is AWGN. We fix the normalized Doppler frequency is fixed at $f_m T_{\text{pkt}} = 0.02$. The pilot symbols are transmitted at the same power as the data symbols.

The channel is Rayleigh faded with time correlation dictated by the Jakes' spectrum, and is independent across AEs (see a later subsection for a simulation result that incorporates spatial correlation in the channel). We generate a large number $(10,000)$ of time-correlated Rayleigh fading channel traces generated using the algorithm given in [49]. For each trace, we start with an AS training phase followed by data transmission via packets. We perform receive AS using the different algorithms, and collect the statistics of packet success/failure of each packet index separately, and average it across the different channel traces. This way, we arrive at the PER corresponding to each packet index. We plot the performance for the 10^{th} packet. We repeat this process for different SNR values, different number of receive antennas, different number of states in the FSMC, and different AS algorithms.

The POMDP problem is formulated as explained in Section III. We consider POMDP models where the number of states per channel is 2, 4 and 8. For each POMDP problem we find two solutions: as given by the Approximate POMDP Planning Toolkit [31] and the myopic policy. We plot the performance of POMDP in two cases: when there is perfect CSI on the selected antenna, and when the CSI is estimated on all antennas.

B. Comparison with Existing Schemes

In this subsection, we compare the performance of the POMDP solution with the AS training based on the weighting scheme proposed in [26]. Another scheme that is evaluated is `Max picking`, which picks the antenna with the highest estimated channel gain in the AS training phase. We also evaluate the PER in case of a single AE (`No AS`). `Perfect CSI` denotes the PER curve with a genie-aided receiver that has perfect CSI on all antennas. Except for the POMDP, all the schemes deal with continuous-valued channel gains. We present comparisons of the PER performance of the different schemes as a function of the data SNR, normalized Doppler frequency and packet index.

1) Variation of PER with SNR: The performance of the optimal policy as obtained from SARSOP tool and myopic policy for the 2-states channel model is plotted in Fig. 4 and those for the 4-states model is plotted in Fig. 5. For the 2-states model with perfect CSI on the selected antenna, we have shown the optimality of myopic policy. Hence, in the 2-state case, the myopic policy marginally outperforms the SARSOP policy, as expected. In the 4-state cases, the results show that the myopic policy performs very close to the SARSOP policy, which is again expected, since the SARSOP tool yields a near-optimal policy, and the optimality of the myopic policy is not valid in this case. However, in all cases, the performance difference between the myopic policy and the SARSOP policy is marginal, indicating that the computationally simpler myopic performance achieves near-optimal performance. The dramatic

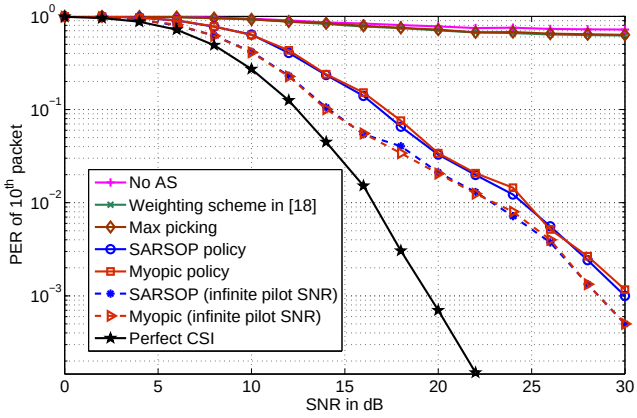


Fig. 4. PER vs. SNR for the 10^{th} packet for 2-states channel model with $f_m T_{\text{pkt}} = 0.02$.

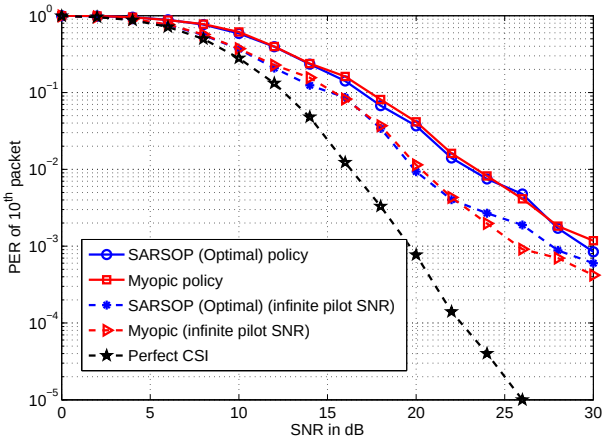


Fig. 5. PER vs. SNR for the 10^{th} packet for the 4-states channel model with $f_m T_{\text{pkt}} = 0.02$.

improvement in performance of our proposed scheme is due to its effectiveness in utilizing the DM RS for data decoding as well as for AS decisions. The other schemes are wholly dependent on the channel estimates from AS training phase for AS as well as for the data decoding. These estimates start getting outdated as the data packets are received, while the proposed scheme updates the channel estimates from the DM RS pilots available in each packet. In Fig. 5, the other schemes' performance are not shown; they are unchanged because they do not discretize the channel state. However, the performance with `Perfect CSI` is plotted for comparison.

2) Variation of PER with Normalized Doppler Frequency: Here, we plot the PER of different schemes as a function of normalized Doppler frequency ($f_m T_{\text{pkt}}$). In Fig. 6, we plot the performance of 2-states channel model along with the other schemes. At lower normalized Doppler frequencies, the optimal weighting scheme and the `Max picking` scheme outperform the POMDP solution. This is because these schemes have the advantage of comparing among the continuous-valued channel gains whereas the POMDP solution has to deal with belief vector of a finite state channel model. However, as the $f_m T_{\text{pkt}}$ increases, the channel varies considerably with time, and the other schemes fail to track the variation. Hence, they perform worse than the POMDP solution. In Fig. 7, where there are 4 states per channel, the POMDP solution performs better than the 2-states model. The `No AS` scheme has only one AE, and hence, performs the worst, even at lower normalized Doppler frequencies.

3) Variation of PER with packet index: Here, we plot the variation in the PER as a function of the packet index for the different schemes. The SNR is fixed at 20 dB and $f_m T_{\text{pkt}}$ at 0.02. Figs. 8 and 9 give the performance of the 2-states and 4-states channel model, respectively. Starting from the first packet itself, the POMDP solution performs better than the other solutions. This is because the POMDP solution uses the DM RS to track the channel state of the selected AE, which is not incorporated by the other schemes. As one gets farther away from the AS training phase, the performance gets progressively worse. However, the degradation in performance is far lower for the POMDP solution compared to the other

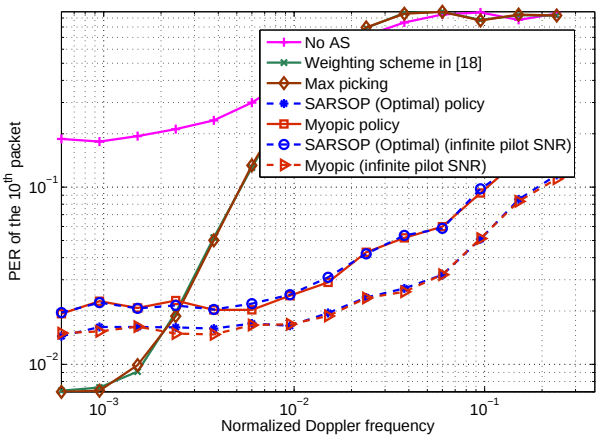


Fig. 6. PER vs. normalized Doppler frequency ($f_m T_{\text{pkt}}$) for the 10^{th} packet for 2-states channel model with data SNR = 20 dB.

schemes. The other schemes necessitate the transmission of a fresh AS training phase after, say, the 10^{th} packet, since the PER becomes close to 1.

C. Comparison with Heuristic Methods

We now consider the performance of the heuristic schemes for AS proposed in Section V. In Fig. 10, we plot the PER as a function of normalized Doppler frequency for the heuristic schemes, and compare it with the POMDP solution. For the sake of simplicity, in case of the POMDP, the performance of the myopic policy is plotted. The overall performance of both the modified schemes is superior to that of their counterparts in the literature, as they utilize information from the DM RS in addition to the initial AS phase for finding the best AE. In fact, the modified `Max picking` (labeled `Mod. max picking`) outperforms the POMDP solutions (with the 2-states and 4-states channel models) at the lower Doppler frequencies. This is because the POMDP discretizes the channel gains. However, as we increase the number of states to model the channel in the POMDP, the performance is improved. The POMDP solution for 8-states channel model

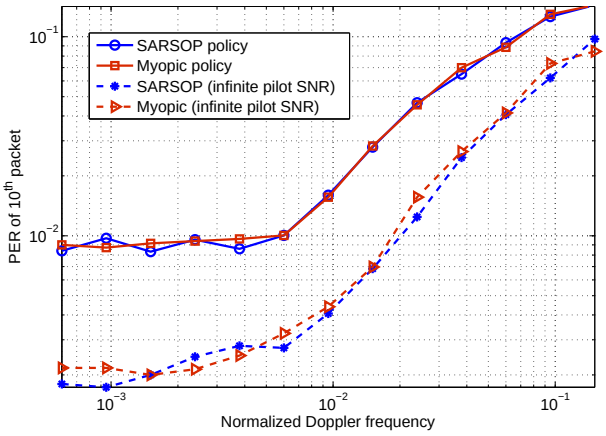


Fig. 7. PER vs. normalized Doppler frequency ($f_m T_{\text{pkt}}$) for the 10^{th} packet for 4-states channel model with data SNR = 20 dB.

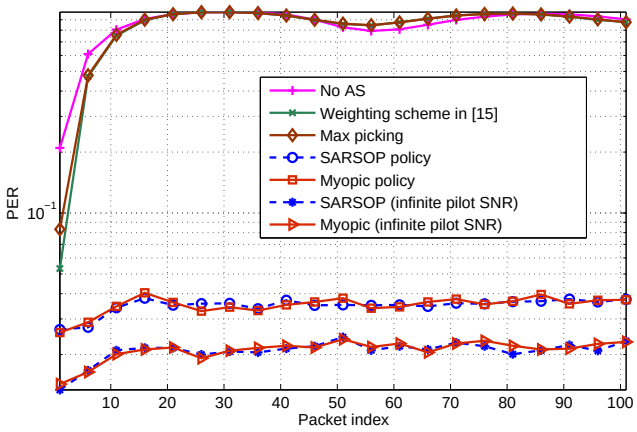


Fig. 8. PER vs. packet index for 2-states channel model with data SNR = 20 dB and $f_m T_{\text{pkt}} = 0.02$.

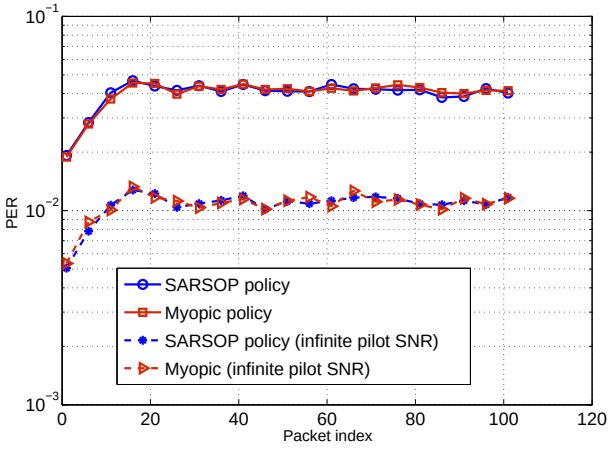


Fig. 9. PER vs. packet index for 4-states channel model with data SNR = 20 dB and $f_m T_{\text{pkt}} = 0.02$.

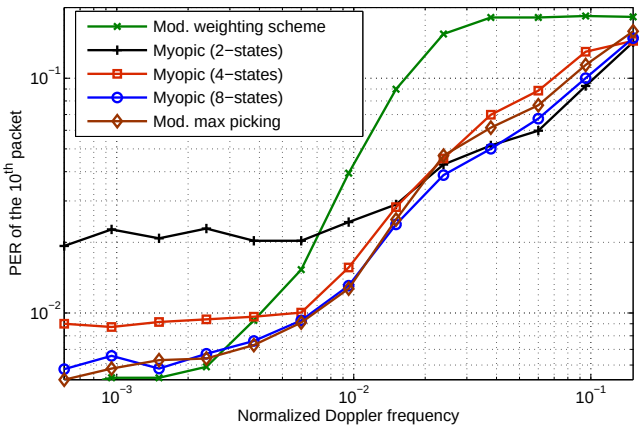


Fig. 10. PER vs. normalized Doppler frequency for the heuristic methods for the 10^{th} packet with data SNR = 20 dB.

performs as well as `Mod. max picking`. A difference between the POMDP approach and the heuristic schemes is that the POMDP approach accounts for the evolution of the belief states of the non-selected antennas also, while in the heuristic scheme, only the selected antenna's channel gain is updated. Note that, the relative performance of the different schemes depends on a variety of factors including the pilot SNR, data SNR, doppler frequency, number of receive antennas, etc. In the typical settings considered in this simulation result, we see that the performance of the modified weighting scheme (`Mod. weighting scheme`) degrades faster than that of the other schemes. The reason is that, the original weighting scheme [26] assumes that the channel estimates from the AS training phase are used for both data decoding and AS. However, in the modified scheme, we obtain a new estimate of the channel on the selected AE from the DM RS, which can be used for data decoding. Also, the weighting scheme in [26] is no longer optimal since the additional CSI is not exploited. The improved performance of the modified schemes compared to the existing weighting schemes underlines the importance of the DM RS in helping the receiver decide which antenna to use in selecting the next packet, in terms of maximizing the long-term reward. In Fig. 10, for lower values of the normalized Doppler frequency, the 2-states channel model gives worse performance than the others. This is expected, since, with more number of states, we can track a slow-varying channel more accurately. However, as the channel varies faster, the performance of 2-states model starts to outperform others. This is partly an artifact of the FSMC model, which allows state transitions between adjacent states only. Due to that restriction, a channel model with fewer number of states is better suited to represent a fast-varying channel.

D. Channel correlation

The system model considered so far assumed spatially uncorrelated channels. It is well-known that correlation among the channels degrades the PER performance of AS based MIMO systems. Hence, it is of interest to study how the proposed scheme performs in the presence of spatial correlation.

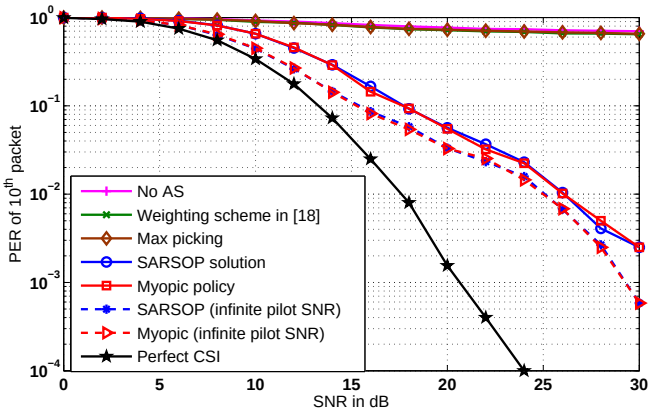


Fig. 11. PER vs. SNR for the 10^{th} packet for 2-states channel model with $f_m T_{\text{pkt}} = 0.02$ and spatially correlated channels with correlation coefficient $= 0.6$ between adjacent antennas.

The system model remains the same as explained in Sec. II, except that the N channels are now correlated. We assume that the correlation matrix has entries $1, \rho, \rho^2, \ldots, \rho^{N-1}$ on the first row, i.e., the correlation coefficient between adjacent antennas is ρ. The POMDP formulation also remains the same as before, except that the state transition probability $\Pr(S_j|S_i)$ needs to account for the spatial correlation between the antennas. Other than using the modified state transition probability matrix, the POMDP solution remains identical to the one presented above. In this experiment, we evaluate the PER performance of different schemes under spatially correlated channels with $\rho = 0.6$. For the POMDP formulation, we have considered the 2-state model. The result is shown in Fig. 11. We see that, compared to the uncorrelated case in Fig. 4, the performance of all schemes has degraded, but the relative performance of the different schemes remains more or less unchanged.

VII. Conclusion

In this work, we considered a wireless communication system with a receiver having single RF chain and N AEs. By formulating the problem as a POMDP, we were able to exploit the temporal correlation of the channel as well as the additional information available in the DM RS. We showed the optimality of the computationally simple myopic policy for the 2 state channel model under the assumption of perfect CSI on the selected antenna and positively correlated channels. Through simulations, we showed that the performance of the myopic policy is very close to that of the optimal policy obtained from the SARSOP POMDP solver, even for the 4-states channel model. We also proposed two heuristic policies that offer excellent performance and are simple to implement. The primary advantage of our proposed approach is that it obviates the need for frequent AS training phases. This reduction in training overhead can translate to improved spectral efficiency, or allow transmission at lower power to improve the energy efficiency and reduce interference to other systems. Future work can include the analysis of optimality of the myopic policy for a general κ-state channel model, and the design of joint transmit-receive AS schemes based on Markov decision theory.

Appendix A
Proof of Decomposability of the Pseudo Value Function

We are interested in proving the following for all $t = 1, 2, \ldots, T$ and all $l \in \{1, 2, \ldots, N\}$.

$$\begin{aligned} &W_t(\omega_1, \ldots, \omega_l, \ldots, \omega_N) \\ &= \omega_l W_t(\omega_1, \ldots, 1, \ldots, \omega_N) \\ &\quad + (1-\omega_l) W_t(\omega_1, \ldots, 0, \ldots, \omega_N). \end{aligned} \tag{18}$$

The proof proceeds by induction. The result obviously holds for $t = T$. Assuming it holds for $t+1, \ldots, T$ also, (18) can be proved as follows. Consider the case when $l \neq N$. We can expand the LHS as given below:

$$\begin{aligned} W_t(\boldsymbol{\Omega}(t)) = f(\omega_N) &+ \beta\omega_N W_{t+1}(\tau(\omega_1), \ldots, \tau(\omega_l), \ldots, p_{11}) \\ &+ \beta(1-\omega_N) W_{t+1}(p_{01}, \ldots, \tau(\omega_l), \ldots, \tau(\omega_{N-1})). \end{aligned} \tag{19}$$

Now, expanding the first term in the RHS of (18) and applying the induction hypothesis we have

$$\begin{aligned} &\omega_l W_t(\omega_1, \ldots, 1, \ldots, \omega_N) \\ &= \omega_l \left[f(\omega_N) + \beta\omega_N W_{t+1}(\tau(\omega_1), \ldots, p_{11}, \ldots, p_{11}) \right. \\ &\quad \left. + \beta(1-\omega_N) W_{t+1}(p_{01}, \ldots, p_{11}, \ldots, \tau(\omega_{N-1})) \right] \\ &= \omega_l f(\omega_N) + \beta\omega_l p_{11} \left[\omega_N W_{t+1}(\tau(\omega_1), \ldots, 1, \ldots, p_{11}) \right. \\ &\quad \left. + (1-\omega_N) W_{t+1}(p_{01}, \ldots, 1, \ldots, \tau(\omega_{N-1})) \right] \\ &\quad + \beta\omega_l(1-p_{11}) \left[\omega_N W_{t+1}(\tau(\omega_1), \ldots, 0, \ldots, p_{11}) \right. \\ &\quad \left. + (1-\omega_N) W_{t+1}(p_{01}, \ldots, 0, \ldots, \tau(\omega_{N-1})) \right]. \end{aligned} \tag{20}$$

Similarly, we expand the second term in the RHS of (18), to get

$$\begin{aligned} &(1-\omega_l) W_t(\omega_1, \ldots, 0, \ldots, \omega_N) \\ &= (1-\omega_l) \left[f(\omega_N) + \beta\omega_N W_{t+1}(\tau(\omega_1), \ldots, p_{01}, \ldots, p_{11}) \right. \\ &\quad \left. + \beta(1-\omega_N) W_{t+1}(p_{01}, \ldots, p_{01}, \ldots, \tau(\omega_{N-1})) \right] \\ &\quad = (1-\omega_l) f(\omega_N) \\ &\quad + \beta(1-\omega_l) p_{01} \left[\omega_N W_{t+1}(\tau(\omega_1), \ldots, 1, \ldots, p_{11}) \right. \\ &\quad \left. + (1-\omega_N) W_{t+1}(p_{01}, \ldots, 1, \ldots, \tau(\omega_{N-1})) \right] \\ &\quad + \beta(1-\omega_l)(1-p_{01}) \left[\omega_N W_{t+1}(\tau(\omega_1), \ldots, 0, \ldots, p_{11}) \right. \\ &\quad \left. + (1-\omega_N) W_{t+1}(p_{11}, \ldots, 0, \ldots, \tau(\omega_{N-1})) \right]. \end{aligned} \tag{21}$$

Combining (20) and (21) and noting the fact that $\tau(\omega_l) = p_{11}\omega_l + (1-\omega_l)p_{01}$, we get

$$\begin{aligned} &\omega_l W_t(\omega_1, \ldots, 1, \ldots, \omega_N) + (1-\omega_l) W_t(\omega_1, \ldots, 0, \ldots, \omega_N) \\ &= f(\omega_N) + \beta\tau(\omega_l) \left[\omega_N W_{t+1}(\tau(\omega_1), \ldots, 1, \ldots, p_{11}) \right. \\ &\quad \left. + (1-\omega_N) W_{t+1}(p_{01}, \ldots, 1, \ldots, \tau(\omega_{N-1})) \right] \\ &\quad + \beta(1-\tau(\omega_l)) \left[\omega_N W_{t+1}(\tau(\omega_1), \ldots, 0, \ldots, p_{11}) \right. \\ &\quad \left. + (1-\omega_N) W_{t+1}(p_{01}, \ldots, 0, \ldots, \tau(\omega_{N-1})) \right] \\ &= f(\omega_N) + \beta \left[\omega_N W_{t+1}(\tau(\omega_1), \ldots, \tau(\omega_l), \ldots, p_{11}) \right. \\ &\quad \left. + (1-\omega_N) W_{t+1}(p_{01}, \ldots, \tau(\omega_l), \ldots, \tau(\omega_{N-1})) \right] \\ &\quad = W_t(\omega_1, \ldots, \omega_l, \ldots, \omega_N). \end{aligned} \tag{22}$$

This proves (18) for the case when $l \neq N$. Now, consider the case, when $l = N$, and expand the LHS of (18), to get

$$W_t(\mathbf{\Omega(t)}) = f(\omega_N) + \beta\omega_N W_{t+1}(\tau(\omega_1), \ldots, p_{11}) + \beta(1-\omega_N)W_{t+1}(p_{01}, \ldots, \tau(\omega_{N-1})). \quad (23)$$

Expanding the RHS of (18), when $l = N$, gives the following terms:

$$\omega_N W_t(\omega_1, \ldots, \omega_{N-1}, 1) = \omega_N \left[f(1) + \beta W_{t+1}(\tau(\omega_1), \ldots, \tau(\omega_{N-1}), p_{11})\right], \quad (24)$$

$$(1-\omega_N)W_t(\omega_1, \ldots, \omega_{N-1}, 0) = (1-\omega_N)\left[f(0) + \beta W_{t+1}(p_{01}, \tau(\omega_1), \ldots, \tau(\omega_{N-1}))\right]. \quad (25)$$

Noting that $f(\omega_N) = \omega_N f(1) + (1-\omega_N)f(0)$, it is straight forward to verify that combining (24) and (25) gives the RHS of (23).

Appendix B
Proof of Lemma 1

The two inequalities in the lemma will be proven together by induction. For time $t = T$, equation (14) becomes $\Delta + f(\omega_1) \geq f(\omega_N)$. This is true, since Δ is the maximum value $f(\omega_N) - f(\omega_1)$ can take. In (15), for time $t = T$ and when $l = N-2$, we have $f(x) \geq f(y)$ since $x \geq y$. When $l \leq N-3$, we have the equality. Assuming both (14) and (15) are true for time $t+1, t+2, \ldots, T$, let us first prove (14) holds for t. The second term in the LHS of (14) corresponds to the case when antenna 1 is selected and the RHS to that when antenna N is selected, at time t. A sample path argument similar to the one in [50] is adopted in our proof. We consider all the four realizations for AEs 1 and N and show that (14) holds in all cases.

1) Case 1.a: The states of AE 1 and N are 0 and 1, respectively. Let us denote the LHS and RHS of (14) under this realization as $L|_{[0,1]}$ and $R|_{[0,1]}$, respectively. We have

$$L|_{[0,1]} = \Delta + f(0) + \beta W_{t+1}(p_{01}, \tau(\omega_2), \ldots, \tau(\omega_{N-1}), p_{11}) \quad (26)$$

$$R|_{[0,1]} = f(1) + \beta W_{t+1}(p_{01}, \tau(\omega_2), \ldots, \tau(\omega_{N-1}), p_{11})). \quad (27)$$

The last summation term in both the above equations is evidently the same. Noting that in this particular realization $f(0) + \Delta = f(1)$, we have $L|_{[0,1]} = R|_{[0,1]}$.

2) Case 1.b: The states of AE 1 and N are both 0.

$$\begin{aligned} L|_{[0,0]} &= \Delta + f(0) + \beta W_{t+1}(p_{01}, \tau(\omega_2), \ldots, \tau(\omega_{N-1}), p_{01}) \\ R|_{[0,0]} &= f(0) + \beta W_{t+1}(p_{01}, p_{01}, \tau(\omega_2), \ldots, \tau(\omega_{N-1})) \\ &\leq f(0) + \beta[\Delta + W_{t+1}(p_{01}, \tau(\omega_2), \ldots, \tau(\omega_{N-1}), p_{01})] \\ &\leq f(0) + \Delta + \beta W_{t+1}(p_{01}, \tau(\omega_2), \ldots, \tau(\omega_{N-1}), p_{01}) \\ &= L|_{[0,0]} \end{aligned} \quad (28)$$

where the first inequality is due to the induction hypothesis of (14). The second inequality utilizes the fact that $\beta \leq 1$.

3) Case 1.c: The states of AE 1 and N are both 1.

$$\begin{aligned} L|_{[1,1]} &= \Delta + f(1) + \beta W_{t+1}(\tau(\omega_2), \ldots, \tau(\omega_{N-1}), p_{11}, p_{11}) \\ R|_{[1,1]} &= f(1) + \beta W_{t+1}(p_{11}, \tau(\omega_2), \ldots, \tau(\omega_{N-1}), p_{11}) \\ &\leq f(1) + \beta W_{t+1}(\tau(\omega_2), \ldots, \tau(\omega_{N-1}), p_{11}, p_{11}) \\ &= L|_{[1,1]} - \Delta \\ &\leq L|_{[1,1]} \end{aligned} \quad (29)$$

where the first inequality is due to the repeated application of the induction hypothesis of (15) and the last equality is due to the assumption that $\Delta \geq 0$.

4) Case 1.d: The states of AE 1 and N are 1 and 0 respectively.

$$\begin{aligned} L|_{[1,0]} &= \Delta + f(1) + \beta W_{t+1}(\tau(\omega_2), \ldots, \tau(\omega_{N-1}), p_{01}, p_{11}) \\ R|_{[1,0]} &= f(0) + \beta W_{t+1}(p_{01}, p_{11}, \tau(\omega_2), \ldots, \tau(\omega_{N-1})) \\ &\leq f(0) + \beta W_{t+1}(p_{01}, \tau(\omega_2), \ldots, \tau(\omega_{N-1}), p_{11}) \\ &\leq f(0) + \beta[\Delta + W_{t+1}(\tau(\omega_2), \ldots, \tau(\omega_{N-1}), p_{11}, p_{01})] \\ &\leq f(0) + \Delta + \beta W_{t+1}(\tau(\omega_2), \ldots, \tau(\omega_{N-1}), p_{01}, p_{11}) \\ &= L|_{[1,0]} - \Delta \\ &\leq L|_{[1,0]} \end{aligned} \quad (30)$$

where the first and third inequality uses the induction hypothesis of (15). The second inequality utilizes the induction hypothesis of (14).

Now we proceed to prove (15) for time t. We consider two cases as given below.

5) Case 2.a: $l \leq N-3$.

$$\begin{aligned} \text{LHS} &= f(\omega_N) + \beta\omega_N W_{t+1}(\tau(\omega_1), \ldots, \tau(y), \tau(x), \ldots, p_{11}) \\ &\quad + \beta(1-\omega_N)W_{t+1}(p_{01}, \tau(\omega_1), \ldots, \tau(y), \tau(x), \ldots, \tau(\omega_{N-1})) \\ &\geq f(\omega_N) + \beta\omega_N W_{t+1}(\tau(\omega_1), \ldots, \tau(x), \tau(y), \ldots, p_{11}) \\ &\quad + \beta(1-\omega_N)W_{t+1}(p_{01}, \tau(\omega_1), \ldots, \tau(x), \tau(y), \ldots, \tau(\omega_{N-1})) \\ &= \text{RHS} \end{aligned} \quad (31)$$

where the inequality is due to the induction hypothesis of (15).

6) Case 2.b: $l = N-2$. From (12), we have

$$W_t(\omega_1, \ldots, \omega_{N-2}, y, x) - W_t(\omega_1, \ldots, \omega_{N-2}, x, y) = (x-y)\left[W_t(\omega_1, \ldots, \omega_{N-2}, 0, 1) - W_t(\omega_1, \ldots, \omega_{N-2}, 1, 0)\right]. \quad (32)$$

Expanding the last term on the RHS,

$$\begin{aligned} &W_t(\omega_1, \ldots, \omega_{N-2}, 1, 0,) \\ &\quad = f(0) + \beta W_{t+1}(p_{01}, \tau(\omega_1), \ldots, \tau(\omega_{N-2}), p_{11}) \\ &\quad \leq f(0) + \beta(\Delta + W_{t+1}(\tau(\omega_1), \ldots, \tau(\omega_{N-2}), p_{11}, p_{01})) \\ &\quad \leq f(0) + \Delta + \beta W_{t+1}(\tau(\omega_1), \ldots, \tau(\omega_{N-2}), p_{11}, p_{01}) \\ &\quad \leq f(1) + \beta W_{t+1}(\tau(\omega_1), \ldots, \tau(\omega_{N-2}), p_{01}, p_{11}) \\ &\quad = W_t(\omega_1, \ldots, \omega_{N-2}, 0, 1). \end{aligned} \quad (33)$$

The first inequality is due to the induction hypothesis of (14) and the second inequality is due to $\beta \leq 1$. The last inequality is due to the induction hypothesis of (15) and also $f(0) + \Delta = f(1)$. Since $x \geq y$, (32) evaluates to a positive quantity. This completes the proof Lemma 1.

REFERENCES

[1] P. Sinchu, R. S. George, C. R. Murthy, and M. Coupechoux, "A POMDP solution to antenna selection for PER minimization," in *Proc. IEEE Globecom*, 2014.

[2] A. F. Molisch and M. Z. Win, "MIMO systems with antenna selection," *IEEE Microw. Mag.*, vol. 5, no. 1, pp. 46–56, 2004.

[3] S. Sanayei and A. Nosratinia, "Antenna selection in MIMO systems," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 68–73, 2004.

[4] S. Kashyap and N. Mehta, "Joint antenna selection and frequency-domain scheduling in OFDMA systems with imperfect estimates from dual pilot training scheme," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3473–3483, July 2013.

[5] J. Coon and M. Sandell, "Combined bulk and per-tone transmit antenna selection in OFDM systems," *IEEE Commun. Lett.*, vol. 14, no. 5, pp. 426–428, May 2010.

[6] G. Brante, I. Stupia, R. D. Souza, and L. Vandendorpe, "Outage probability and energy efficiency of cooperative MIMO with antenna selection," *IEEE Trans. Wireless Commun.*, vol. 12, no. 11, pp. 5896–5907, November 2013.

[7] "Draft amendment to wireless LAN media access control (MAC) and physical layer (PHY) specifications: enhancements for higher throughput," IEEE, Tech. Rep. P802.11n/D0.04, Mar 2006.

[8] "Technical specification group radio access network; evolved universal terrestrial radio access (E-UTRA); physical layer procedures," 3rd Generation Partnership Project (3GPP), Tech. Rep. 36.211, 2008.

[9] X. N. Zeng and A. Ghrayeb, "Performance bounds for space-time block codes with receive antenna selection," *IEEE Trans. Inf. Theory*, vol. 50, pp. 2130–2137, 2004.

[10] A. Dua, K. Medepalli, and A. J. Paulraj, "Receive antenna selection in MIMO systems using convex optimization," *IEEE Trans. Wireless Commun.*, vol. 5, no. 9, pp. 2353–2357, 2006.

[11] A. F. Molisch, M. Z. Win, Y. Choi, and J. H. Winters, "Capacity of MIMO systems with antenna selection," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1759–1772, 2005.

[12] M. Z. Win and J. H. Winter, "Virtual branch analysis of symbol error probability for hybrid selection/maximal-ratio combining in Rayleigh fading," *IEEE Trans. Commun.*, vol. 49, no. 11, pp. 1926–1934, 2001.

[13] B. H. Wang, H. T. Hui, and M. S. Leong, "Global and fast receiver antenna selection for MIMO systems," *IEEE Trans. Commun.*, vol. 58, no. 9, pp. 2505–2510, 2010.

[14] R. Vaze and H. Ganapathy, "Sub-modularity and antenna selection in mimo systems," *IEEE Commun. Lett.*, vol. 16, no. 9, pp. 1446–1449, Sep. 2012.

[15] T. Gucluoglu and E. Panayirci, "Performance of transmit and receive antenna selection in the presence of channel estimation errors," *IEEE Commun. Lett.*, vol. 12, no. 5, pp. 371–373, 2008.

[16] W. M. Gifford, M. Z. Win, and M. Chiani, "Antenna subset diversity with non-ideal channel estimation," *IEEE Wireless Commun. Mag.*, vol. 7, pp. 1527–1539, 2009.

[17] A. Coskun and O. Kucur, "Performance analysis of maximal-ratio transmission/receive antenna selection in nakagami- m fading channels with channel estimation errors and feedback delay," *IEEE Trans. Veh. Technol.*, vol. 61, no. 3, pp. 1099–1108, Mar. 2012.

[18] A. B. Narasimhamurthy and C. Tepedelenlioglu, "Antenna selection for MIMO-OFDM systems with channel estimation error," *IEEE Trans. Veh. Technol.*, vol. 58, no. 5, pp. 2269–2278, 2009.

[19] A. Molisch and X. Zhang, "Fft-based hybrid antenna selection schemes for spatially correlated mimo channels," *IEEE Commun. Lett.*, vol. 8, no. 1, pp. 36–38, Jan. 2004.

[20] B. H. Wang and H. T. Hui, "Investigation on the fft-based antenna selection for compact uniform circular arrays in correlated mimo channels," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 739–746, Feb. 2011.

[21] V. Krishnamurthy and R. Evans, "Hidden Markov model multiarm bandits: a methodology for beam scheduling in multitarget tracking," *IEEE Trans. Signal Process.*, vol. 49, no. 12, pp. 2893–2908, Dec. 2001.

[22] A. Mukherjee and A. Hottinen, "Learning algorithms for energy-efficient MIMO antenna subset selection: Multi-armed bandit framework," in *Proc. EUSIPCO*, Aug. 2012, pp. 659–663.

[23] N. Gulati and K. Dandekar, "Learning state selection for reconfigurable antennas: A multi-armed bandit approach," *IEEE Trans. Antennas Propag.*, vol. 62, no. 3, pp. 1027–1038, Mar. 2014.

[24] H. Zhang, A. F. Molisch, and J. Zhang, "Applying antenna selection in WLANs for achieving broadband multimedia communications," *IEEE Trans. Broadcast.*, vol. 52, no. 4, pp. 475–482, 2006.

[25] H. A. Saleh, A. F. Molisch, T. Zemen, S. D. Blostein, and N. B. Mehta, "Receive antenna selection for time-varying channels using discrete prolate spheroidal sequences," *IEEE Trans. Wireless Commun.*, vol. 11, no. 7, pp. 2616–2627, 2012.

[26] V. Kristem, N. B. Mehta, and A. F. Molisch, "Optimal receive antenna selection in time-varying fading channels with practical training constraints," *IEEE Trans. Commun.*, vol. 58, no. 7, pp. 2023–2034, 2010.

[27] R. G. Stephen, C. R. Murthy, and M. Coupechoux, "A Markov decision theoretic approach to pilot allocation and receive antenna selection," *IEEE Trans. Wireless Commun.*, vol. 12, no. 8, pp. 3813–3823, Aug. 2013.

[28] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artificial Intelligence*, vol. 101, no. 1-2, pp. 99–134, 1998.

[29] R. A. Howard, *Dynamic Programming and Markov Processes*. MIT Press, 1960.

[30] E. J. Sondik, "The optimal control of partially observable Markov processes over the infinite horizon: Discounted costs," *Operations Research*, pp. 282–304, 1978.

[31] [Online]. Available: http://bigbird.comp.nus.edu.sg/pmwiki/farm/appl/

[32] Y. Li and Y. Guan, "Modified jakes model for simulating multiple uncorrelated fading waveforms," in *Proc. IEEE ICC*, 2000, pp. 46–49.

[33] W. C. Jakes, *Microwave Mobile Communications*. Piscataway, NJ: IEEE Press, 1993.

[34] J. M. Porta, N. Vlassis, M. T. J. Spaan, and P. Poupart, "Point-based value iteration for continuous POMDPs," *J. Machine Learning Research*, vol. 7, pp. 2329–2367, 2006.

[35] S. Thrun, "Monte carlo POMDPs," in *NIPS*, vol. 12, 1999, pp. 1064–1070.

[36] H. Bai, D. Hsu, W. S. Lee, and V. A. Ngo, "Monte carlo value iteration for continuous-state POMDPs," in *Algorithmic foundations of robotics IX*. Springer, 2011, pp. 175–191.

[37] P. Sadeghi, R. A. Kennedy, P. B. Rapajic, and R. Shams, "Finite-state Markov modeling of fading channels," *IEEE Trans. Signal Process.*, vol. 25, no. 5, pp. 57–80, 2008.

[38] H. S. Wang and N. Moayeri, "Finite-state Markov channel-a useful model for radio communication channels," *IEEE Trans. Veh. Technol.*, vol. 44, no. 1, pp. 163–171, 1995.

[39] Q. Zhang and S. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 1688–1692, 1998.

[40] X. Li, C. Shen, A. Bo, and G. Zhu, "Finite-state markov modeling of fading channels: A field measurement in high-speed railways," in *Communications in China (ICCC), 2013 IEEE/CIC International Conference on*, Aug. 2013, pp. 577–582.

[41] C. D. Iskander and P. Mathiopoulos, "Fast simulation of diversity nakagami fading channels using finite-state markov models," *IEEE Trans. Broadcast.*, vol. 49, no. 3, pp. 269–277, Sep. 2003.

[42] H. Song, R. Kwan, and J. Zhang, "General results on SNR statistics involving EESM-based frequency selective feedbacks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1790–1798, May 2010. [Online]. Available: http://dx.doi.org/10.1109/TWC.2010.05.091431

[43] R. D. Smallwood and E. J. Sondik, "The optimal control of partially observable Markov processes over a finite horizon," *Operations Research*, pp. 1071–1088, 1973.

[44] A. Cassandra, "Exact and approximate algorithms for partially observable Markov decision processes," Ph.D. dissertation, 1998.

[45] J. Pineau, G. Gordon, and S. Thrun, "Point-based value iteration: An anytime algorithm for POMDPs," in *Proc. Int. Joint Conf. Artificial Intelligence*, vol. 18. Lawrence Erlbaum Associates Ltd., 2003, pp. 1025–1032.

[46] H. Kurniawati, D. Hsu, and W. S. Lee, "SARSOP: efficient point-based POMDP planning by approximating optimally reachable belief spaces," in *Proc. Robotics: Science and Systems*, 2008.

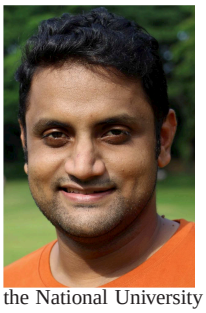
[47] S. H. A. Ahmad, M. Liu, T. Javidi, Q. Zhao, and T. Krishnamachari, "Optimality of myopic sensing in multichannel opportunistic access," *IEEE Trans. Inf. Theory*, vol. 50, pp. 2130–2137, 2004.

[48] K. Wang and L. Chen, "On optimality of myopic policy for restless multi-armed bandit problem: An axiomatic approach," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 300–309, Jan 2012.

[49] D. J. Young and N. C. Beaulieu, "The generation of correlated Rayleigh random variates by inverse discrete fourier transform," *IEEE Trans. Commun.*, vol. 48, pp. 1114–1127, 2000.

[50] S. Ahmad and M. Liu, "Multi-channel opportunistic access: A case of restless bandits with multiple plays," in *Proc. Allerton Conf. on Commun., Control and Comput.*, 2009, pp. 1361–1368.

Sinchu Padmanabhan received the B. Tech. degree in Electronics and Communication Engineering from the National Institute of Technology, Calicut, India, in 2003, and the M. E. degree in Signal Processing from the Dept. of Electrical Engineering, Indian Institute of Science, Bangalore, India, in 2013. She is a scientist in the Naval Physical and Oceanographic Laboratory, Kochi, India. She works in the area of underwater acoustic signal processing.



Reuben George Stephen received the B. Tech. degree in electronics and communication engineering from the Cochin University of Science and Technology, Kerala, India, in 2009 and the M. E. degree in telecommunications from the Dept. of ECE, Indian Institute of Science, Bangalore, India, in 2012. From Aug. 2012 to Dec. 2013, he was a Research Engineer in the Center for Development of Telematics (C-DoT), Bangalore, and involved in the development and testing of LTE femto base stations. Since Jan. 2014, he is working towards the Ph. D. degree in the National University of Singapore. His research interests are broadly in the areas of resource allocation and optimization in wireless systems.

Chandra R. Murthy (S'03–M'06 – SM'11) received the B. Tech. degree in Electrical Engineering from the Indian Institute of Technology, Madras in 1998, the M. S. and Ph. D. degrees in Electrical and Computer Engineering from Purdue University and the University of California, San Diego, in 2000 and 2006, respectively.

From 2000 to 2002, he worked as an engineer for Qualcomm Inc., where he worked on WCDMA baseband transceiver design and 802.11b baseband receivers. From Aug. 2006 to Aug. 2007, he worked as a staff engineer at Beceem Communications Inc. on advanced receiver architectures for the 802.16e Mobile WiMAX standard. In Sept. 2007, he joined the Department of Electrical Communication Engineering at the Indian Institute of Science, where he is currently working as an Associate Professor. His research interests are in the areas of Cognitive Radio, Energy Harvesting Wireless Sensors, MIMO systems with channel-state feedback, and sparse signal recovery techniques applied to wireless communications. He is currently serving as the Chair of the IEEE Signal Processing Society, Bangalore Chapter and as an associate editor for the IEEE Signal Processing Letters. He is an elected member of the IEEE SPCOM Technical Committee for the years 2014-16.

Marceau Coupechoux has been working as an Associate Professor at Telecom ParisTech since 2005. He obtained his Masters' degree from Telecom ParisTech in 1999 and from University of Stuttgart, Germany, in 2000, and his Ph.D. from Institut Eurecom, Sophia-Antipolis, France, in 2004. From 2000 to 2005, he was with Alcatel-Lucent (in Bell Labs former Research & Innovation and then in the Network Design department). He was a Visiting Scientist at the Indian Institute of Science, Bangalore, India, during 2011–2012. Currently, at the Computer and Network Science department of Telecom ParisTech, he is working on cellular networks, wireless networks, ad hoc networks, cognitive networks, focusing mainly on layer 2 protocols, scheduling, and resource management.