A POMDP Solution to Antenna Selection for PER Minimization

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

***Abstract*—In this work, the problem of receive antenna selection (AS) is considered, in a multiple antenna communication system having a single radio frequency (RF) chain at the receiver. The AS is performed on a per-packet basis, and AS decisions are based on noisy estimates of the channel gains obtained using pilot symbols embedded in the data packet for coherent demodulation, along with the receiver's knowledge of the time correlation of the channel. The problem is posed as a partially observable Markov decision process (POMDP) with the goal of minimizing the average packet error rate (PER). The performance of a myopic policy is compared with that of the POMDP solution, and it is shown that the former is optimal under certain conditions. As the POMDP approach requires the channel gains to be quantized to a finite set of states, we also propose two heuristic AS schemes that use the continuous-valued received pilot symbols to make AS decisions, and thereby offer comparable or better performance than the POMDP approach. Unlike previous work, the schemes proposed here for AS do not require a lengthy AS training phase to precede each data packet. 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.**

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

I. Introduction

Antenna selection (AS) is a technique that is widely explored in both academia and the industry for reducing the hardware complexity and cost of a multiple input multiple output (MIMO) system [1]–[5]. With AS, only a subset of the available antenna elements (AEs) is selected for transmission/reception, while maintaining the diversity order of the full-complexity system. Some of the early research on AS assumed perfect channel state information (CSI) at the receiver [6]–[9]. In practice, the CSI is imperfect due to errors in its estimation, which is typically done using a small number of pilot symbols embedded in the packet. Surprisingly, the diversity order achievable with perfect CSI is preserved even with imperfect CSI [10], [11]. The focus of this work is on AS for a multiple-antenna receiver with a single RF chain at the receiver, where the selection is done on a per-packet basis.

In training based AS, the receiver requests for an AS training phase, and the transmitter sends out L sets of N known training symbols to the receiver [12]. Here, $L \geq 1$ is an integer and 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 $\eta \geq 2$[1] [14]. Thus, the total AS training duration is $\eta N L T_s$, which is repeated whenever the channel estimates get outdated, imposing a non-trivial overhead on the AS based system. When the CSI is noisy and outdated, Kristem and Mehta [13] derive an optimal scheme for weighting the channel gain estimates obtained in the AS training phase, that minimizes the symbol error probability (SEP). However, in their work, both AS and subsequent data decoding are performed using the same channel state estimates obtained during the AS training phase. In many practical systems, there are additional pilots in the data phase also, viz. demodulation reference signals (DM RS) which are used for data decoding [5]. Saleh et al. [14] take these DM RS into account, and propose an algorithm for AS, that maximizes the post-processing SNR. However, they do not use the CSI obtained from the data phase for making future AS decisions. Stephen et al. [15] formulate AS in a decision theoretic framework, with the aim of maximizing the throughput, again assuming an AS training phase. Also, the channel is assumed to remain constant over an entire frame which consists of the AS training phase and a data packet, which may not hold when the channel is fast-varying.

The studies mentioned above depend on frequent repetition of the training phase for AS. This motivates the use of DM RS for AS also, which would eliminate the dependence of the AS process on lengthy training phases. Also, in a packet based selection scheme, each data packet reveals some information on the selected antenna via the DM RS, which can be used in making *future* selection decisions. As the channel is correlated in time, each selection not only affects the immediate packet reception, but future receptions as well. Specifically, since the channel states on the antennas are only partially observable through the DM RS, and on only one antenna at a time, we cast the problem in a partially observable Markov decision process (POMDP) [16], [17] framework, with the goal of minimizing the average packet error rate (PER).

The contributions of our work are as follows. We propose an AS scheme which eliminates the need for an expensive AS training phase at the start of each data packet. This is unlike most of the past work on AS, which assumes the use of an AS training phase between successive packets. We note that our method can also exploit an AS training phase, when present.

We formulate the sequential AS problem as an infinite

Sinchu P. is with the Naval Physical and Oceanographic Laboratory, Kochi, India. She was a student at the EE dept. at the Indian Institute of Science (IISc), Bangalore, India, during the course of this work. Email: sinchup@gmail.com. R. G. Stephen is with the Dept. of ECE, National Univ. of Singapore. Email: reubenstephen@gmail.com. C. R. Murthy is with the Dept. of ECE, IISc, Bangalore. 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 was supported in part by a research grant funded by the Indo-French Centre for the Promotion of Advanced Research.

[1]The pilot symbols are typically embedded in a training packet with a physical layer header [13] and hence are spaced several symbols apart.

horizon POMDP. The optimal policy for an infinite horizon problem is stationary [17]. Although finding an optimal policy for a general POMDP is PSPACE hard [18], for the case when the number of states per antenna is two, under the assumption that the pilot symbols reveal the channel state perfectly and the channel is positively correlated[2], we show that the optimal policy is myopic in nature[3]. We evaluate the average PER performance of different AS policies via Monte Carlo simulations, and show that, even when the number of channel states is greater than two, and with finite pilot SNRs, the myopic policy performs very close to the optimal solution.

We compare our results with the optimal weighting scheme of Kristem and Mehta [13], which is based on AS training. Another scheme that picks the antenna with the highest channel gain in the AS training phase for receiving the subsequent packets is also evaluated. We show that the proposed scheme outperforms both these schemes.

In addition to the POMDP solution, we propose two heuristic schemes for AS and evaluate their performance. The performance comparison of these schemes, which are based on continuous-valued channel gains, with that of the finite state Markov chain (FSMC)-based POMDP solution, gives further insights into the nature of the POMDP solution.

II. System Model

We consider a communication system with a single antenna transmitter and a receiver with N antenna elements (AE), but one RF chain. The channels from the transmitter to the AEs are frequency flat, Rayleigh faded and i.i.d. across the AEs. AS is done on a per-packet basis. The goal is to select the *best* AE out of the N, so that the average PER is minimized. Each channel is time-correlated from packet to packet, with the receiver having the knowledge of the correlation coefficient, but is assumed to remain constant for a packet 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 [14]. Hence, switching delays are negligible.

The system operates in discrete time steps of duration T_{pkt}. Each packet has D data symbols and a DM RS denoted by p. At the beginning of each time step, the channels make a state transition. The AE which is selected based on the information available up to and including the previous time step, is used to receive the current packet. The DM RS embedded in the packet yields information on the channel state of the AE that receives the packet. This information is used to decode the data symbols in the packet, as well as to update the CSI of the selected AE. With this additional information gained in the current time step and the history of decisions and observations, a new selection decision is made for the next time step, and the process continues.

The POMDP formulation requires the states of the system to be discrete-valued. Hence, we model the Rayleigh faded time correlated channels as finite state Markov chains (FSMCs), which is known to be accurate for packet-level studies [19], [20]. We use the method of Zhang and Kassam [21] to partition the instantaneous signal-to-noise ratios (SNRs) on the receive AEs. The thresholds on the channel gains that determine the states depend on the normalized Doppler frequency. We emphasize that the instantaneous SNR is discretized into a finite number of states only for the purpose of defining the state space and solving the POMDP. The underlying channel remains continuous-valued in the implementation and evaluation of the policy prescribed by the solution of the POMDP.

In a POMDP, due to the partial observability, the state of the system is not fully revealed to the receiver. However, it has been shown [22] that 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 system starts with an initial belief vector, $\mathbf{b}(1)$, that is updated at each state transition and with each observation. In training based AS, $\mathbf{b}(1)$ is obtained from the training phase. When there is no AS training phase, $\mathbf{b}(1)$ can be set to the stationary probabilities of the Markov chains. A policy for a POMDP maps the current belief vector into an action, namely, the selection of an AE. Each policy has an expected long term reward associated with it, and the optimal policy is the one that maximizes this reward. Once the elements of the POMDP are defined, one of the many available POMDP solving algorithms can be used to find an optimal policy.

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 from earlier results [21]. At time t, depending on the current belief vector, the optimal policy gives the AE index to be selected for receiving the next packet. Let $h(t)$ be the complex valued channel gain of the selected AE. Let γ_0 denote the average per-symbol SNR. Then, the instantaneous SNR is given by $\gamma = |h(t)|^2 \gamma_0$. If $\gamma_j \leq \gamma < \gamma_{j+1}$, 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

$$y = hp + n, \tag{1}$$

where p is the known pilot symbol and n is the additive white Gaussian noise (AWGN) with zero mean and variance σ_n^2. The maximum likelihood (ML) estimate of the channel gain is

$$\hat{h} = \frac{p^*}{|p|^2} y = h + e, \tag{2}$$

where $e \triangleq \frac{p^*}{|p|^2} n$. When perfect channel state information is available on the selected antenna, $\hat{h} = h$. The estimate $\hat{h}$ is used to decode the packet, as well as to update the belief vector. The optimal policy maps this new belief vector to the index of the AE to be selected in the next time step. The next section develops the POMDP formulation of the AS problem for minimizing the average PER.

[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. This is a very reasonable assumption in practice.

[3] A myopic policy is one that ignores the effect of the current action on future rewards, and hence is simpler to implement than a general optimal policy.

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. Since the channels are assumed to be independent, the transition probability $P(S_j|S_i)$ 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} AE 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 CSI for that antenna. Since this CSI is continuous-valued, we need to discretize it into states using the thresholds of the FSMC model. Then, the observation space is $\mathcal{O} = \{1, 2, \ldots, \kappa\}$. Let $S(a)$ denote the state of the selected antenna a when the system state is S. Then $O(S, a, o)$ is the probability of observing state $o \in \mathcal{O}$ on the selected antenna, given its true state, $S(a)$. This probability is derived in Appendix A. It varies with the pilot SNR, and in the case of perfect knowledge of the channel state on the selected antenna, $O(S, a, o) = 1$ if $o = S(a)$, and $O(S, a, o) = 0$ otherwise.

4) Reward: As 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

$$\varrho(a, S) = \sum_{j=1}^{\kappa} P(o = j|S(a))P_{\text{cor}}(o = j, S(a)). \tag{3}$$

$P_{\text{cor}}(o, S(a))$ gives the probability of correctly receiving the packet when the true state is $S(a)$ and the observed state is o. Its value is assumed to be known here; it can be easily calculated, for example, via simulations, for different true states and observation states. It should be noted that, here, 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. The expected immediate reward can be expressed as a function of the belief state, $\mathbf{b}$, as follows:

$$R(a, \mathbf{b}) = \sum_{S \in \mathcal{S}} b_S \varrho(a, S), \tag{4}$$

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

5) Objective and the Optimal Policy: The objective is to minimize the expected PER, over an infinite horizon. The averaging is done in a discounted sense, i.e., the future rewards are discounted by a factor $\beta \in [0, 1)$. Let $J^{\pi}_{\beta}(\mathbf{b})$ denote the expected total discounted reward associated with a policy π starting from time step $t = 1$ and belief vector $\mathbf{b}$. Then, the *optimal* policy solves the following optimization problem

$$\max_{\pi} J^{\pi}_{\beta}(\mathbf{b}) = \max_{\pi} \mathbb{E}\left[\sum_{t=1}^{\infty} \beta^{t-1} R(\pi(\mathbf{b}(t)), \mathbf{b}(t))|\mathbf{b}(1) = \mathbf{b}\right], \tag{5}$$

where $R(\pi(\mathbf{b}(t)), \mathbf{b}(t))$ is the reward collected under belief state $\mathbf{b}(t)$ when the AE $\pi(\mathbf{b}(t))$ is selected

This completes the POMDP formulation of the AS problem. There are several tools available for solving POMDPs [23]–[25]. However, solving the POMDP is a non-trivial problem, especially when the number of states of the system under consideration is large. A usual practice is to explore the effectiveness of a simpler policy for the problem, which may not be optimal, such as the myopic policy. In the sequel, we state that the myopic policy is indeed optimal for the AS POMDP problem, under the assumptions of positively correlated channels and full observability of the channel state on the selected AE, for a 2-states-per-antenna model. Full observability on the selected AE is equivalent to a scenario where the DM RS has infinite SNR, and, hence, the channel state of the selected AE is fully revealed.

Let $P_c(s)$ denote the probability of correctly receiving a packet when the channel state is $s \in \{0, 1\}$. Here, state 1 corresponds to a higher channel gain than state 0, and hence, $P_c(1) \geq P_c(0)$. For a positively correlated 2-states channel, the transition probabilities satisfy $p_{11} \geq p_{01}$, where p_{ij} denotes the transition probability from state i to state j. 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 state the theorem 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 [26].

Since there are only two states per antenna, the belief vector can be written as $\mathbf{\Omega}(t) \triangleq [\omega_1(t), \omega_2(t), \ldots, \omega_N(t)]$, where $\omega_i(t)$ is the conditional probability that the channel i is in the good state at time step t given all past actions and observations. $\mathbf{\Omega}(t)$ 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, and s_i the state of the i^{th} antenna at time t. The belief vector can be updated according to Bayes' rule, as follows:

$$\omega_i(t+1) = \begin{cases} p_{11} & \text{if } a(t) = i, s_{a(t)} = 1, \\ p_{01} & \text{if } a(t) = i, s_{a(t)} = 0, \\ \tau(\omega_i(t)) & \text{if } a(t) \neq i, \end{cases} \tag{6}$$

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. The objective is to maximize the total expected discounted reward over a horizon of T. The optimal policy is then,

$$\pi^* = \arg\max_{\pi} \mathbb{E}_{\pi}\left[\sum_{t=1}^{T} \beta^{t-1} R(\pi_t(\mathbf{\Omega}(t)), \mathbf{\Omega}(t)) \mid \mathbf{\Omega}(1)\right]. \tag{7}$$

Any admissible policy, $\pi = [\pi_1, \pi_2, \ldots, \pi_T]$, is a vector, where π_t maps $\mathbf{\Omega}(t)$ to an action $a(t), t = 1, 2, \ldots, T$. The time index t is necessary here, as the optimal policy for a finite horizon problem is, in general, non-stationary.

Let $V_t(\mathbf{\Omega}(t))$ denote the value function of the optimal policy at time t, which is the expected sum of reward gained, starting in belief vector $\mathbf{\Omega}(t)$, from time t to T. The dynamic programming formulation of the value function associated with

the optimal policy can be expressed as follows:[4]

$$V_T(\mathbf{\Omega}) = \max_{a=1,\ldots,N} \mathbb{E}[R(a,\mathbf{\Omega})] \quad (8)$$

$$V_t(\mathbf{\Omega}) = \max_{a=1,\ldots,N} \mathbb{E}\left[R(a,\mathbf{\Omega}) + \beta V_{t+1}(\mathcal{T}(\mathbf{\Omega}))\right] \quad (9)$$

where $\mathcal{T}(.)$ denotes the one-step update operator for the belief vector $\mathbf{\Omega}$, and is as defined in (6). The expected immediate reward collected is given by

$$R(a,\mathbf{\Omega}) = \omega_a P_c(1) + (1-\omega_a)P_c(0) \triangleq f(\omega_a). \quad (10)$$

As $P_c(1) \geq P_c(0)$, $f(\omega_a)$ increases linearly with ω_a. A myopic policy chooses that action which maximizes $f(\omega_a)$, and hence this is equivalent to choosing the antenna with the highest belief state. Now, we state the theorem on the optimality of the myopic policy.

Theorem 1. *The myopic policy is optimal for the problem stated in* (8) *and* (9)*, 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 proceeds by induction, and is based on Lemma 2 in Ahmad et al. [26]. Details are omitted here due to lack of space, and will be included in an extended version of this work.

IV. Discussion

In this work, we proposed an AS scheme that utilizes the DM RS information from the data phase and exploits the known time-correlation of the channel. We formulated the problem as a POMDP and stated the optimality of the myopic policy under certain conditions.

A. Existing Schemes

In Section V, we compare the performance of the proposed scheme, with that of two existing schemes which are based on AS training. 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 occasionally requests the transmitter for a training phase, for example, when the resulting PER is below some acceptable level [12].

In subsection V-A, the performance of the POMDP solution is compared with that of the weighting scheme [13] 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. `Max picking` compares the channel gain estimates of the AEs from the AS training phase and selects the AE with the highest estimate for receiving the packets in the data phase. In the weighting scheme [13], the estimates from the AS training phase are linearly weighted, and the AE with the highest weighted estimate is selected to receive a symbol. In order to compare the PER performance of this scheme with that of the POMDP solution, a per-packet selection is done. 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 that uses the DM RS information for AS. The objective is to show that taking into account the information obtained through the DM RS, improves the AS performance significantly. Hence, we consider a fixed packet, the 10^{th} packet, for simulation purposes, and show the effect of using the DM RS, on the PER performance, as opposed to using only the AS training phase.

[4]For the ease of presentation, we drop the time index for $\mathbf{\Omega}$.

B. Heuristic schemes

The POMDP formulation requires the continuous-valued channel gains estimated from the pilot symbols to be discretized into a finite number of states for the purpose of solving the POMDP and determining the optimal policy. Hence, the receiver is not using all the information available from the received pilots[5] This becomes evident in Figure 2, where the other two schemes, explained above, outperform the POMDP solution when the channel varies very slowly. This motivates one to come up with an approach that does not require the channel gains to be *quantized*, but nevertheless utilizes the DM RS information for data decoding as well as for AS.

We present two heuristic schemes, by modifying the `Max picking` and the weighting schemes so as to include the information from the DM RS. The modification that we propose for these schemes is 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 estimate obtained from the DM RS in the packet. This new estimate is compared with the outdated estimates of other AEs, while selecting the AE for receiving the next packet. Similarly, in the modified weighting scheme, for selecting the 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 way 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. In subsection V-B, we present the simulation results for these modified schemes and compare their performance with that of the POMDP solution.

V. Simulation Results

In this section, we present Monte Carlo simulation results to validate the claims in the previous sections.

A. Comparison with Existing Schemes

In order to compare our results with the existing AS training based schemes, we assume there is an initial training phase, which is followed by several data packets. For 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 from an 8-PSK constellation. We assume that there are $N = 4$ AEs and $\kappa = 4$ states per channel. The POMDP problem is formulated as in Section III. We find two solutions for the POMDP problem: one, as given by a POMDP solver tool, the Approximate POMDP Planning Toolkit [27], and the

[5]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 [21] is that it restricts the state transitions to happen only between adjacent channel states, and this affects the optimality of the policy due to the inaccuracy of the model, when the number of states becomes large.

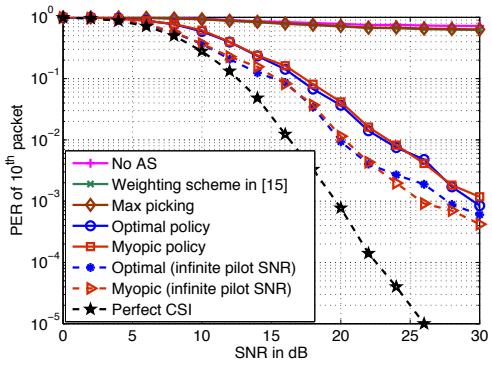

Fig. 1. PER vs SNR for the 10^{th} packet with $\kappa = 4$ and $f_m T_{\text{pkt}} = 0.02$.

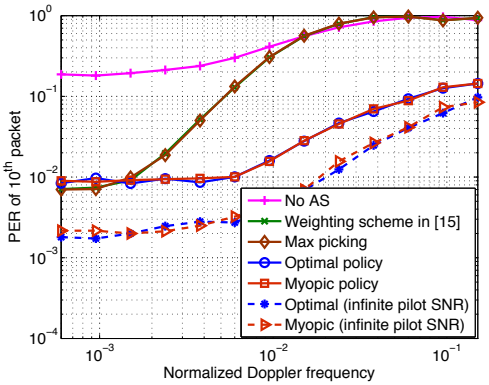

Fig. 2. PER vs normalized Doppler frequency ($f_m T_{\text{pkt}}$) for the 10^{th} packet for 4-states channel model with data SNR = 20 dB.

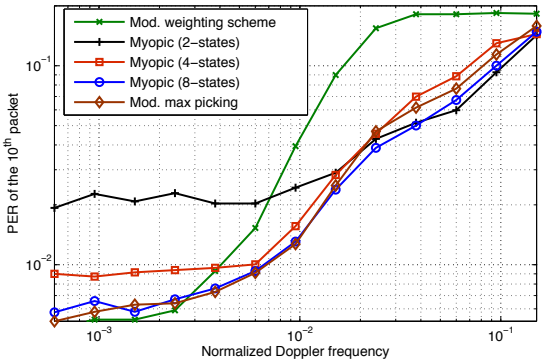

Fig. 3. PER vs normalized Doppler frequency for the heuristic methods for the 10^{th} packet with data SNR = 20 dB.

other, the myopic policy. The performance of the POMDP solution is plotted in two cases: when under full observability of the channel state on the selected AE, and when the pilot SNR is finite and is equal to the data SNR. We compare the performance of the POMDP solution with the weighting scheme [13], and `Max picking`, which picks the antenna with the highest estimated channel gain in the AS training phase, as explained in Section IV. `No AS` plots the PER in case of a single AE, and `Perfect CSI` plots the PER of a genie-aided receiver that has perfect CSI on all antennas. Except for the POMDP, all the schemes deal with continuous-valued channel gains.

In the next subsection, we present comparisons of the PER performance of the different schemes as a function of the data SNR and normalized Doppler frequency. Unless otherwise mentioned, we plot the performance for the 10^{th} packet, with the normalized Doppler frequency, $f_m T_{\text{pkt}} = 0.02$ and AWGN at the receiver. The pilot symbols are assumed to have the same SNR as that of the data symbols. The channels are Rayleigh faded with time correlation dictated by the Jakes' spectrum [28], and are independent across AEs.

1) Variation of PER with SNR: Figure 1 shows that there is a dramatic improvement in performance for the POMDP schemes as compared to the others. This is mainly due to their effective use of the DM RS for data decoding as well as for AS decisions. Also, the myopic policy performs very close to the optimal policy, suggesting that it may be optimal even for channels with $\kappa > 2$ and with finite pilot SNR.

2) Variation of PER with Normalized Doppler Frequency: Figure 2 shows that at lower normalized Doppler frequencies, the weighting scheme and `Max picking` outperform the POMDP solution, as these schemes have the advantage of comparing among the continuous-valued channel gains, whereas the POMDP solution has to deal with the belief vector of a finite state channel model. However, as $f_m T_{\text{pkt}}$ increases, other schemes fail to track the fast time-varying channel, and therefore perform much worse than the POMDP solution. The `No AS` scheme has only one AE, and hence, performs the worst, even at lower normalized Doppler frequencies.

B. Comparison with Heuristic Methods

In this section, we present the performance of the heuristic solutions for AS that were proposed in Section IV. Figure 3, plots the PER as a function of normalized Doppler frequency for the heuristic schemes. For the sake of simplicity, in case of the POMDP, only the performance of the myopic policy is shown. It can be seen that the overall performance of both the modified schemes is better than that of the original schemes. In fact, the modified `Max picking` (`Mod. max picking`) scheme performs better than the POMDP solution with the 2-states and 4-states channel model, at relatively 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. For the 8-states channel model, the POMDP solution performs as well as `Mod. max picking`. The performance of modified weighting scheme (`Mod. weighting scheme`) degrades faster than that of other schemes. The reason is that in the original scheme [13], weighted channel estimates are used for both data decoding and AS. In the modified scheme, we use the weighted estimate for AS and the new estimate from the DM RS for data decoding, which outperforms the use of the outdated weighted estimate. In Figure 3, as the channel varies faster, the 2-states model outperforms the others. This is due to an limitation in the model, which allows state transitions only between adjacent states. Due to this, a channel model with fewer number of states is more suited to representing a fast varying channel.

VI. Conclusion

In this work, we considered a wireless communication system with the receiver having a single RF chain and N AEs. We proposed an AS method that effectively utilizes the DM RS in the data phase and exploits the knowledge of the

time correlation of the channel. We formulated the problem as a POMDP, and found the optimal policy using a POMDP solver tool. We stated the optimality of the computationally simple myopic policy for the 2-state channel model under the assumption of perfect channel observability on the selected AE and positively correlated channels. Inspired by this result, we explored the effectiveness of the myopic policy for the finite pilot SNR case and for the 4-states channel model. Through simulations, we showed that the performance of the myopic policy is very close to that of the optimal policy obtained from the POMDP solver. An advantage of our approach is that the need for frequent lengthy AS training phases is eliminated. Thus, the proposed scheme can provide a higher throughput than schemes which depend on the AS training phase.

APPENDIX A
OBSERVATION PROBABILITY FOR THE FINITE SNR CASE

Multiplying y in (1) by $\frac{p^*}{|p|\sigma_n}$ gives $y' = \sqrt{\gamma} + w'$. where γ is the instantaneous SNR as defined in Section II, and $w' \sim \mathcal{CN}(0,1)$. An unbiased estimator of γ is $\hat{\gamma} = |y'|^2 - 1$.

The probability that the estimated SNR is in state $n \in \mathcal{G}$, given the true state $m \in \mathcal{G}$, is given by,

$$\begin{aligned} &P(\gamma_n < \hat{\gamma} < \gamma_{n+1} | \gamma_m < \gamma < \gamma_{m+1}) \\ &= P(\alpha_n < \tilde{\gamma} < \alpha_{n+1} | \gamma_m < \gamma < \gamma_{m+1}) \\ &= \int_{\gamma_m}^{\gamma_{m+1}} \frac{Q_1(\sqrt{\gamma}, \sqrt{\alpha_n}) - Q_1(\sqrt{\gamma}, \sqrt{\alpha_{n+1}})}{e^{-\frac{\gamma_m}{\gamma_0}} - e^{-\frac{\gamma_{m+1}}{\gamma_0}}} \times \frac{e^{-\frac{\gamma}{\gamma_0}}}{\gamma_0} \mathrm{d}\gamma \end{aligned} \tag{11}$$

where $\tilde{\gamma} = |y'|^2$, $\alpha_n = \gamma_n + 1$, and $Q_1(a,b)$ is the Marcum Q-function [29]. The integral in (11) can be evaluated in closed form [15, Appendix B] to get the observation probability as

$$P(\gamma_n < \hat{\gamma} < \gamma_{n+1} | \gamma_m < \gamma < \gamma_{m+1}) = \frac{F(\alpha_n) - F(\alpha_{n+1})}{e^{-\frac{\gamma_m}{\gamma_0}} - e^{-\frac{\gamma_{m+1}}{\gamma_0}}} \tag{12}$$

for all $n, m \in \mathcal{G}$, where

$$\begin{aligned} F(x) = {} & e^{-\frac{\gamma_m}{\gamma_0}} Q_1\left(\sqrt{\gamma_m}, \sqrt{x}\right) - e^{-\frac{\gamma_{m+1}}{\gamma_0}} Q_1\left(\sqrt{\gamma_{m+1}}, \sqrt{x}\right) \\ & + e^{-\frac{x}{\gamma_0(1+\frac{2}{\gamma_0})}} \left[Q_1\left(\sqrt{\gamma_{m+1}(1+\frac{2}{\gamma_0})}, \sqrt{\frac{x}{1+\frac{2}{\gamma_0}}}\right) \right. \\ & \left. - Q_1\left(\sqrt{\gamma_m(1+\frac{2}{\gamma_0})}, \sqrt{\frac{x}{1+\frac{2}{\gamma_0}}}\right) \right]. \end{aligned} \tag{13}$$

REFERENCES

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

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

[3] A. Gorokhov, D. Gore, and A. Paulraj, "Receive antenna selection for MIMO flat-fading channels: theory and algorithms," *IEEE Trans. Ind. Informat*., vol. 49, no. 10, pp. 2687–2696, 2003.

[4] "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.

[5] "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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] R. Stephen, C. 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, 2013.

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

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

[18] C. H. Papadimitriou and J. N. Tsitsiklis, "The complexity of optimal queueing network control," *Math. Oper. Res*., vol. 24, pp. 293–305, 1999.

[19] 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.

[20] 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.

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

[22] 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.

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

[24] 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.

[25] 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.

[26] 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.

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

[28] 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.

[29] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series and Products*, 5th ed. San Diego, CA: Academoc, 1994.