Pilot Allocation and Receive Antenna Selection: A Markov Decision Theoretic Approach

Reuben G. Stephen, Chandra R. Murthy, Marceau Coupechoux

Pilot Allocation and Receive Antenna Selection: A Markov Decision Theoretic Approach

Reuben George Stephen*, Chandra R. Murthy†, and Marceau Coupechoux‡

***Abstract*—This paper considers antenna selection (AS) for packet reception at a receiver equipped with multiple antenna elements but only a single radio frequency chain. The receiver makes its AS decisions based on noisy channel estimates obtained from the training symbols (pilots). The time-correlation of the wireless channel and the results of the link-layer error checks upon data packet reception provide additional information that can be exploited for AS. This information can also be used to optimally distribute pilots among the antenna elements, so that packet loss due to selection errors is minimized. The task of the receiver, then, is to sequentially select (a) the pilot symbol allocation for channel estimation on each of the receive antennas and (b) the antenna to be used for data packet reception. The goal is to maximize the expected throughput, based on the history of allocation and selection decisions, and the corresponding noisy channel estimates and error check observations. This joint problem of pilot allocation and AS is solved as a partially observed Markov decision problem (POMDP) and the solutions yield the optimal policies that maximize the long-term expected throughput. The performance of the POMDP solution is compared with several other schemes for a 2-state Markov channel model, and it is illustrated that it outperforms the others.**

***Index Terms*—Antenna selection, pilot allocation, POMDP.**

I. Introduction

Antenna selection (AS) [1], [2] is a popular technique for reducing the hardware costs at the transmitter and/or receiver of a multiple antenna wireless link. The idea is to use a limited number of radio frequency (RF) chains while adaptively switching to subsets of a larger number of available antenna elements. AS maintains the same diversity order as a system that uses all the available antenna elements, and only a small loss in data rate is suffered when the receiver uses the best possible subset [2]. AS can be employed at the transmitter, receiver or both ends; this work focuses on receive AS.

Several algorithms for AS that assume perfect channel state information (CSI) at the receiver have been proposed earlier [3, and references therein]. However, in practice, it is necessary to estimate CSI, using, for example, a pilot-based training scheme. Imperfect CSI can lead to both inaccurate AS and erroneous decoding of data, increasing the symbol error probability (SEP) [4]. Quite surprisingly, it has been shown that transmit and receive AS can achieve full diversity order, even in the presence of channel estimation errors [5]. However, most of the past work on AS with imperfect CSI suffers from three drawbacks. First, it assumes that the receiver equally divides the pilots among the available antenna elements during the training phase [4], [6]. However, when the channel is slowly-varying, such an equal allocation is not optimal, as past estimates of the channel and the time-correlation information can be used to re-allot pilots among the antennas in subsequent training periods. Second, link-layer error checks on the received packets provide additional information on the channel, and this is typically not exploited in the literature. Third, a quasi-static block-fading channel is usually assumed [1], [7], which precludes the receiver from fully exploiting the temporal channel correlation. This work seeks to overcome all these three drawbacks and fully exploit the information available at the receiver in deciding the optimal pilot allocation for channel estimation and AS for data packet reception.

The system model in this work consists of a transmitter with a single antenna and a receiver with N antenna elements. The receiver has a single RF chain, so it needs to decide on the antenna with which it should receive data from the transmitter. The transmitter sends data in frames, with each frame having L pilot or training symbols, followed by a data packet. The receiver then has the following trade-off. On the one hand it could allot many pilots out of the available L to one particular antenna, getting an accurate estimate of the channel on that antenna. However, this would mean losing track of possibly better channels on other antennas. Alternatively, fewer pilots can be allotted to each of the antennas, tracking all of their channels. But now the receiver will have poorer quality estimates of the channels on a larger number of antennas, leading to errors in subsequent AS decisions, and packet loss. As the receiver can vary the accuracy with which to estimate the channels at the antennas, and select the one to be used for packet reception, it can control the (partial) observability of the system. These controls must be applied so as to maximize some notion of long-term reward. Hence, the joint problem of pilot allotment and antenna selection at the receiver in each frame is modeled in this work as a Partially Observable Markov Decision Process (POMDP) [8]–[10] with the objective of maximizing the long-term packet success rate. The contributions of this work are as follows.

- For the first time in the literature, the general problem of joint pilot allocation and AS in a time-correlated

This work was supported in part by research grant no. 4900-IT-B funded by the Indo French Centre for the Promotion of Advanced Research.

*R. G. Stephen is with the Center for Development of Telematics, Bangalore, India. He was with the Dept. of ECE, IISc, Bangalore, during the course of this work. †C. R. Murthy is with the Dept. of ECE, IISc. ‡M. Coupechoux is with Telecom ParisTech and CNRS LTCI, Paris, France. He was a visiting faculty at the Dept. of ECE, IISc, during the course of this work. Emails: reubenstephen@gmail.com, cmurthy@ece.iisc.ernet.in and marceau.coupechoux@telecom-paristech.fr.

channel is solved in a decision-theoretic framework to obtain an optimal policy that maximizes the throughput. A challenge in the formulation is being able to deal with two different kinds of actions, viz., the pilot allocation and AS decisions, and two types of observations in the training and data phases, as elaborated in Section III.

- Insights are provided on the nature of the policies to be followed. For example, with $N = 2$ and a 2-state Markov channel model, it is found, somewhat surprisingly, that when the channel is fast-varying, the POMDP solution allots all the pilots to the same antenna, and selects the antenna that is most likely to be in a good state.
- With a 2-state channel and $N = 2$, it is found that employing the POMDP solution can lead to savings of $4-8$ dB in the pilot SNR, to achieve the same throughput as a scheme that allots pilots equally among antennas in all frames and selects antennas without using past information about the channels. For this case, it is also found that the myopic policy [11] is nearly optimal over a wide range of channel parameters and pilot SNR values.

The advantage of posing the problem as a POMDP is that it admits the use of a gamut of computationally efficient methods [12, and references therein] for solving it. The solution can be computed offline, and once it is obtained, implementing the optimal policy for pilot allotment and AS is simple. One has to update the belief vector for the system state based on the observations in every slot using Bayes' rule, and then employ the optimal action corresponding to the updated belief vector, possibly by using a look-up table. The solutions presented in this work can lead to a significant reduction in the pilot SNR or the number of pilot symbols required to obtain a given performance, or an improvement in the average data rate in practical AS based systems.

II. System Model

Consider a wireless system with a single transmit antenna, N receive antenna elements and a single RF chain at the receiver. Time is divided into frames of fixed duration T_f. Each frame has a training period T_t and a data transmission period T_d. In the training period, L pilot symbols, each of duration T_s, are received and used to estimate the channel gains at the N antenna elements. This is followed by a data packet transmission, at the end of which the receiver performs an error check and knows whether the packet was received correctly or not. Figure 1 shows the frame structure.

Let $h_i[k]$ denote the frequency-flat channel between the transmitter and the i^{th} receive antenna at the start of frame k. $h_i[k]$ is assumed to be constant for the duration T_f of each frame k, but correlated across frames. This holds if the coherence time T_c of the channel satisfies $T_c \gg T_f$. Consider a particular frame in which $\ell_i \in \{0, 1, \ldots, L\}$ pilots are used to estimate the channel[1] h_i at antenna i, where $\sum_{i=1}^{N} \ell_i = L$. The time overhead of switching between antennas is assumed to be negligible compared to the duration of the training

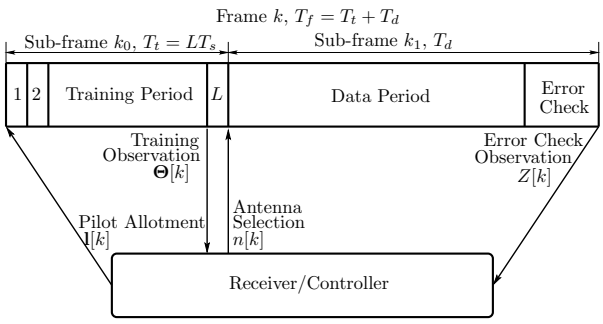

Fig. 1. Sequence of operations in frame k.

phase [2], and is hence ignored. It is common in AS literature to assume that pilots within the same training period can be received on different antenna elements [4], [13]. If $\mathbf{y}_i = [y_1 \quad \cdots \quad y_{\ell_i}]^H \in \mathbb{C}^{\ell_i}$ is the vector of received training symbols on the i^{th} antenna and $\mathbf{p}_i = \sqrt{\frac{E_p}{L}}\ [1 \quad \cdots \quad 1]^T$ is the ℓ_i-length vector of pilots with energy E_p/L each, one can write,

$$\mathbf{y}_i = h_i \mathbf{p}_i + \mathbf{w}_i, \quad i = 1, \ldots, N, \tag{1}$$

when[2] $\ell_i > 0$, where $\mathbf{w}_i \in \mathbb{C}^{\ell_i}$ is the additive white Gaussian noise vector, with $\mathbf{w}_i \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_{\ell_i})$.

In the sequel, the channels at the N antenna elements are modeled by N independent Gilbert-Elliot channels [14]. This is a popular model to characterize packet-level performance measures of correlated Rayleigh fading channels [15], [16]. The extension to a finite number of states is straightforward [17], but handling continuous-valued channels is much harder, and is out of the scope of this work. The channels are assumed to be mutually independent, which holds true if the receive antennas are placed sufficiently apart. This is not a necessary assumption for the POMDP formulation in this work, but is used as it simplifies the derivation of the observation probabilities in the training and observation phases as given in the Appendix. The receiver observes the state of the channel at antenna i with some error that depends on ℓ_i. The precise relation depends on the channel estimation method used. For the sake of concreteness, a particular observation model for the 2-state channel is described in the Appendix along with the derivations of the corresponding probabilities in the training and data phase. However, this is not restrictive; other observation models can be used with the POMDP framework constructed in the sequel. Successful packet reception is assumed to depend only on the true channel state of the selected antenna, rather than the receiver's estimate of the channel. This leads to a tractable relation between the AS decisions and packet success probabilities, which is required to design the optimal policies. As an example, in LTE systems, there are separate sounding reference signals for channel quality estimation, and demodulation reference signals for channel estimation during coherent demodulation [18].

[1]The frame index k in $h_i[k]$ is dropped here for convenience.

[2]When $\ell_i = 0$, no symbol is received on the i^{th} antenna.

III. POMDP Formulation

Let $\mathbf{S}[k] = [S_1[k] \quad \cdots \quad S_N[k]]^T$ denote the state vector of the channels at the N antenna elements in frame k, with $S_i[k] \in \{0,1\}$, $i = 1,\ldots,N$. Here, $S_i[k] \triangleq 0$ if the channel at antenna i is in the bad state and $S_i[k] \triangleq 1$ if it is in the good state. Let k_0 and k_1 represent the training and data sub-frames, respectively. At the start of frame k, the receiver decides on the value of $\ell_i[k]$ to be used to estimate the channels at antennas $i = 1,\ldots,N$ in the training period. The actual channel state vector transits to $\mathbf{S}[k]$ according to the transition probabilities of the underlying Markov chains. Observations $\Theta_i[k]$ that depend on $S_i[k]$ as well as $\ell_i[k]$ are made on each antenna i, and the receiver determines the antenna $n \in \{1,\ldots,N\}$ to be used to receive the data packet. $\Theta_i[k]$ are obtained from (6) using a MAP detector as described in the Appendix. It is assumed that the receiver knows the channel statistics, and hence it can determine the probability $\mathbb{P}_{\ell_i}\{S_i = s|\Theta_i = \theta_i\}$ that the true channel state is s, given the observations $\theta_i \in \{0,1\}$. At the end of the packet reception, $Z[k] \in \{0,1\}$ is observed, which indicates whether the packet was received in error (0), or there was no error (1).

The sequential decision-making process described above and illustrated in Figure 1 is now formalized as a POMDP. Here two actions, viz., pilot allocation and AS, are to be taken at different points in a single frame, and two different observations are made in the training and data phase, while the channel state remains the same. In the classical POMDP framework, actions belong to a single set at all decision points. Hence, the pilot allocation and AS decisions are combined to form a single composite action, to be taken at the start of both the training and the data phase. Also, only one observation can be obtained for a single state-action combination. This mandates a distinction between the state of the system in the training and the data phase, and hence the state vector is expanded with an additional variable $m \in \{0,1\}$ that represents the two different decision points in a single frame. This is necessary to bring the joint pilot allocation and AS problem to a standard form, where it can be solved using existing POMDP solving algorithms.

Within a frame k, $m = 0$ denotes the start of the training period and $m = 1$, the start of the data packet reception period. Since the channels are constant over a frame, transitions are naturally restricted so that

$$\mathbb{P}\{\mathbf{S}[k_1] = \tilde{\mathbf{s}}_1|\mathbf{S}[k_0] = \mathbf{s}_0\} = \begin{cases} 0, & \tilde{\mathbf{s}} \neq \mathbf{s}, \\ 1, & \tilde{\mathbf{s}} = \mathbf{s}, \end{cases} \tag{2}$$

where $\tilde{\mathbf{s}}, \mathbf{s} \in \{0,1\}^N$, $\mathbf{s}_1 \triangleq [\mathbf{s} \quad 1]^T$ and $\mathbf{s}_0 \triangleq [\mathbf{s} \quad 0]^T$. Here $\mathbf{s} = [s_1 \quad \cdots \quad s_N]^T$, denotes the channel state vector without the decision point indication m. Subscripts 0 and 1 are used on $\mathbf{s}$ to indicate the state of the system in the training phase and the data phase, respectively, and

$$k_m + 1 \triangleq (k+m)_{m'}, \tag{3}$$

with $m' \triangleq 1 - m$, $m \in \{0,1\}$. That is, the POMDP slots are the subframes indexed as $1_0, 1_1, 2_0, 2_1$, and so on. The components of the POMDP are formally described next.

1) State Space: The state space of the system is defined as $\mathcal{S} \triangleq \{0,1\}^{N+1}$. The transition probabilities are denoted by $\mathbb{P}\{\tilde{\mathbf{s}}_{\tilde{m}}|\mathbf{s}_m\}$ where $\mathbb{P}\{\tilde{\mathbf{s}}_1|\mathbf{s}_0\}$ is given by (2) and $\mathbb{P}\{\tilde{\mathbf{s}}_0|\mathbf{s}_1\}$ is the transition probability from state $\mathbf{s}_1$ to $\tilde{\mathbf{s}}_0$, calculated from the transition probability matrices of the Markov chains governing the evolution of the channel states.

2) Action Space: The action in a frame has two parts:

- A pilot allocation vector $\mathbf{l} = [\ell_i]_{i=1}^N \in \mathcal{L}$, where $\mathcal{L} \triangleq \left\{\mathbf{l} : \ell_i \in \{0,\ldots,L\}, \sum_{i=1}^N \ell_i = L\right\}$, $|\mathcal{L}| = \binom{N+L-1}{L}$.
- An antenna selection decision $n \in \mathcal{C} \triangleq \{1,\ldots,N\}$.

The receiver takes the composite action $A \triangleq \{\mathbf{l}, n\} \in \mathcal{A}$, where $\mathcal{A} \triangleq \mathcal{L} \times \mathcal{C}$, and $|\mathcal{A}| = \binom{N+L-1}{L}N$, at the start of every decision period $k_m = 1_0, 1_1, 2_0$, and so on. However, for points k_0, only the pilot allocation $\mathbf{l}$ affects the observation, and for k_1, only the selection decision n is of relevance.

3) Observation Space: The observation also has two parts:

- The vector of channel state observations at the antennas, $\mathbf{\Theta}[k_0] = [\Theta_i[k_0]]_{i=1}^N$, whose reliability depends on $\ell_i[k_0]$.
- The packet error indication $Z[k_1] \in \{0,1\}$ obtained at the end of each frame, which depends on the channel state of the antenna selected.

In general, on taking action $A \in \mathcal{A}$, at each k_m, the receiver observes $z[k_m] \in \Omega_m$. For points k_0, $z[k_0] \triangleq \mathbf{\Theta}[k_0] \in \Omega_0$, with $\Omega_0 \triangleq \{0,1\}^N$, and for points k_1, $z[k_1] \triangleq Z[k_1] \in \Omega_1 \triangleq \{0,1\}$. The combined observation set is thus $\Omega \triangleq \Omega_0 \cup \Omega_1$ with $|\Omega| = 2^N + 2$ for the 2-state channels considered here. The probabilities of observing $z \in \Omega_m$ satisfy $\mathbb{P}_A\{z \in \Omega_1|\mathbf{s}_0\} = \mathbb{P}_A\{z \in \Omega_0|\mathbf{s}_1\} = 0$.

4) Reward: The reward is defined as the number of bits or symbols that can be delivered if the packet is received successfully. Given the action $A[k_m] = \{\mathbf{l}[k_m], n[k_m]\}$, and the system state vector $\mathbf{S}[k_m] = \mathbf{s}_m$, the expected immediate reward for the decision period k_m is given by: $R(\mathbf{s}_m, A[k_m]) = m\mathbb{P}_A\{Z[k_m] = 1|\mathbf{s}_m\} \cdot B$. In the sequel, $B = 1$ is assumed without loss of generality. Thus, $R(\mathbf{s}_0, A[k_0]) = 0 \ \forall k$, as the receiver does not collect any immediate reward in the training phase, reward being counted only for packets received successfully. However, the choice of vector $\mathbf{l}[k_0]$ indirectly affects the selection decision at k_1, and hence, the *future* reward. The expected discounted total reward of the POMDP over an infinite horizon represents the expected total number of bits that can be delivered, after applying a discounting factor for future rewards.

5) Belief Vector: With a Markovian evolution of the states, it is known that [9] the entire decision and observation history can be encapsulated in a belief vector $\mathbf{b}[k_m] \triangleq [b_{\mathbf{s}_m}[k_m]]_{\mathbf{s}_m \in \mathcal{S}}$. Here, $b_{\mathbf{s}_m}[k_m] \in [0,1]$ denotes the conditional probability, given the decision and observation history, that the state of the system in decision period k_m is $\mathbf{s}_m$, after taking some action at the start of k_m, and making an observation in k_m. Thus, $b_{\mathbf{s}_m}[k_m] \triangleq \mathbb{P}\big\{\mathbf{S}[k_m] = \mathbf{s}_m\big|\mathbf{b}[0], \{\mathbf{l}[\nu_\mu], n[\nu_\mu], \mathbf{\Theta}[\nu_\mu], Z[\nu_\mu]\}_{\nu_\mu=1_0}^{k_m}\big\}$, where $\mathbf{b}[0]$ is the initial belief vector, i.e., the a priori distribution on the system state just before the start of frame $k = 1$. If no

information on the initial state is available, this can be set to the stationary distribution of the underlying Markov chain.

6) Policy: A policy π specifies the action to be taken at each decision point, in order to meet some objective. The optimal policy for infinite horizon problems is a stationary mapping from the belief space to the action space [10], and hence the optimal policy at decision point k_m maps the belief vector $\mathbf{b}[k_m-1]$ to an action $A[k_m]=\{\mathbf{l}[k_m],n[k_m]\}\in\mathcal{A}$.

7) Objective: It is desired to design the optimal policy π^* that maximizes the expected total number of bits that can be received, i.e., the expected total discounted reward of the POMDP over an infinite horizon. Thus,

$$\pi^*=\arg\max_{\pi}\mathbb{E}_{\pi}\left\{\sum_{k_m=1_0,1_1,\ldots}\beta^q R\left(\mathbf{s}_m[k_m],A[k_m]\right)\big|\mathbf{b}[0]\right\}$$

where $\beta\in[0,1)$ is the discount factor [19], and the exponent $q\triangleq 2(k-1)+m,\ \forall k,m$.

IV. Solving the POMDP

The value function [10] of the POMDP, denoted by $V(\mathbf{b}[k_m])$, represents the *maximum* expected discounted reward that can be obtained, starting in the belief state $\mathbf{b}[k_m]$. According to the notation introduced in the preceding section, at the end of decision period k_m, the receiver takes action $A[k_m+1]=A\in\mathcal{A}$ and observes $z[k_m+1]=z\in\Omega_{m'}$, where $m'=1-m$, $m\in\{0,1\}$. Then, the reward that can be accumulated starting from point k_m+1 consists of two parts:

- immediate reward $R\left(\mathbf{s}'_{m'}[k_m+1],A\right)=m'\mathbb{1}_{\{z=1\}}\cdot 1$,
- maximum expected future reward $V(\mathbf{b}[k_m+1])$,

where k_m+1 is as defined in (3) and $\mathbf{s}'_{m'}$ is the new state in k_m+1 that the system transitions to, starting from $\mathbf{s}_m$ in k_m. Also, $\mathbf{b}[k_m+1]\triangleq\left[b_{\mathbf{s}'_{m'}}[k_m+1]\right]_{\mathbf{s}'_{m'}\in\mathcal{S}}=f(\mathbf{b}[k_m],A,z)$, represents the updated knowledge of the state of the system, after incorporating action $A[k_m+1]=A$ at the start of period k_m+1, and observation $z[k_m+1]=z$, obtained during period k_m+1. Averaging over all possible states $\mathbf{s}_m\in\mathcal{S}$ and observations $z\in\Omega_{m'}$, and then maximizing over all actions $A\in\mathcal{A}$, the optimality equations can be written as:

$$V\left(\mathbf{b}[k_0]\right)=\max_{A\in\mathcal{A}}\sum_{\mathbf{s}_0\in\mathcal{S}}b_{\mathbf{s}_0}[k_0]\sum_{z\in\Omega_1}\mathbb{P}_A\left\{z|\mathbf{b}[k_0]\right\}\cdot [z\cdot 1+\beta V(f(\mathbf{b}[k_0],A,z))],\quad\text{and}\quad(4)$$

$$V(\mathbf{b}[k_1])=\max_{A\in\mathcal{A}}\sum_{\mathbf{s}_1\in\mathcal{S}}b_{\mathbf{s}_1}[k_1]\sum_{\boldsymbol{\theta}\in\Omega_0}\beta\mathbb{P}_A\left\{\boldsymbol{\theta}|\mathbf{b}[k_1]\right\}\cdot V(f(\mathbf{b}[k_1],A,\boldsymbol{\theta})).\quad(5)$$

Here, $\forall z\in\Omega_{m'}$, and $\forall A\in\mathcal{A}$,

$$\mathbb{P}_A\left\{z|\mathbf{b}[k_m]\right\}=\sum_{\mathbf{s}'_{m'}\in\mathcal{S}}\mathbb{P}_A\left\{z|\mathbf{s}'_{m'}\right\}\sum_{\mathbf{s}_m\in\mathcal{S}}b_{\mathbf{s}_m}[k_m]\mathbb{P}\left\{\mathbf{s}'_{m'}|\mathbf{s}_m\right\}.$$

Note that two equations are needed to represent the value function updates in the training and data phases and these need to be simultaneously satisfied, unlike traditional POMDP value updates, where only one equation is required. The first term in (4) corresponds to the expected immediate reward in the data reception period k_1 and the second term $V(f(\mathbf{b}[k_0],A,z))$ is the value obtained starting from point k_1, scaled by discount factor β. On the other hand, (5) has only one term as there is no immediate reward accrued during the training phase (k_0), and only the value $V(f(\mathbf{b}[k_1],A,\boldsymbol{\theta}))$ is averaged over the conditional probability mass function (pmf) of training observation $\boldsymbol{\theta}$, $\mathbb{P}_A\left\{\boldsymbol{\theta}|\mathbf{b}[k_1]\right\}$. The term $\mathbb{P}_A\left\{\boldsymbol{\theta}|\mathbf{s}'_{m'}\right\}=\mathbb{P}_A\left\{\boldsymbol{\Theta}[k_m+1]=\boldsymbol{\theta}|\mathbf{S}[k_m+1]=\mathbf{s}'_{m'}\right\}$ is the conditional pmf of the channel state observation vector, given the landing state $\mathbf{S}[k_m+1]=\mathbf{s}'_{m'}$ and action $A[k_m+1]=A$, while $\mathbb{P}_A\left\{z|\mathbf{s}'_{m'}\right\}=\mathbb{P}_A\left\{Z[k_m+1]=z|\mathbf{S}[k_m+1]=\mathbf{s}'_{m'}\right\}$ is the pmf of the packet error indication. Since the channels are independent, given $\mathbf{l}$, the observations Θ_i depend only on the corresponding states S_i, and hence $\mathbb{P}_A\left\{\boldsymbol{\Theta}[k_0]=\boldsymbol{\theta}|\mathbf{s}_0\right\}=\prod_{i=1}^{N}\mathbb{P}_{\ell_i}\left\{\Theta_i[k_0]=\theta_i|s_{0,i}\right\}$. Similarly, $\mathbb{P}_A\left\{Z[k_1]=z|\mathbf{s}_1\right\}=\mathbb{P}_n\left\{Z[k_1]=z|s_{1,n}\right\}$, where $n[k_1]=n$ is the antenna selected in subframe k_1. The updated belief vector, $\mathbf{b}[k_m+1]$, is obtained by applying Bayes' rule, as

$$b_{\mathbf{s}'_{m'}}[k_m+1]=\mathbb{P}\left\{\mathbf{S}[k_m+1]=\mathbf{s}'_{m'}|\mathbf{b}[k_m],A,z\right\}=\frac{\mathbb{P}_A\left\{z|\mathbf{s}'_{m'}\right\}\sum_{\mathbf{s}_m\in\mathcal{S}}b_{\mathbf{s}_m}[k_m]\mathbb{P}\left\{\mathbf{s}'_{m'}|\mathbf{s}_m\right\}}{\sum_{\mathbf{s}'_{m'}\in\mathcal{S}}\mathbb{P}_A\left\{z|\mathbf{s}'_{m'}\right\}\sum_{\mathbf{s}_m\in\mathcal{S}}b_{\mathbf{s}_m}[k_m]\mathbb{P}\left\{\mathbf{s}'_{m'}|\mathbf{s}_m\right\}}$$

Since the channels at the antennas are mutually independent, $\mathbb{P}\left\{\mathbf{s}'|\mathbf{s}\right\}=\prod_{i=1}^{N}\mathbb{P}\left\{s'_i|s_i\right\}$. $\mathbb{P}\left\{s'_i|s_i\right\}$ can be obtained from the transition probabilities of the Markov chains. Observation probabilities $\mathbb{P}_A\left\{z|\mathbf{s}'\right\}$ can be obtained using the MAP criterion as described in the Appendix.

V. Simulation Results

The POMDP formulated in Section III is solved using the Approximate POMDP Planning Toolkit [20], implementing the SARSOP algorithm. [12]. In all the cases described, the code is run until the error between the value functions (discounted rewards) obtained in consecutive steps falls below a tolerance limit of $\epsilon=1$. The discount factor $\beta=0.99$, since a large value of β is needed for designing the policies that maximize the long-term average performance of the receiver. The channels at all the antennas are independent and assumed to have identical statistics, modeled by a 2-state Markov chain, as described in Section II. Results are shown here for $N=2$, and $L=4$. Similar results are obtained when $N>2$, and for channels with more than 2 states [17]. Performance is evaluated over 2×10^3 subframes, comparing the average throughput achieved—i.e., the average number of packets successfully received per frame—by the POMDP solution with other schemes, as described below. `Max. (genie aided)` gives the maximum attainable throughput when the receiver has perfect knowledge of the channel states at all the antennas. `POMDP solution` shows the performance of the policy obtained by solving the POMDP. `Myopic` is a purely greedy policy [11] that allots all L pilots to the antenna that has the highest likelihood of being in

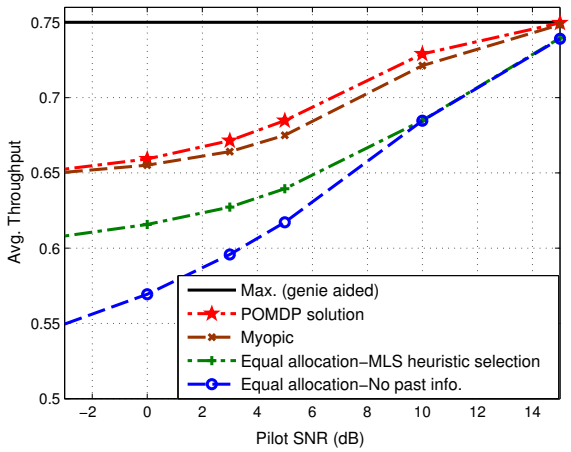

Fig. 2. Avg. Throughput vs. Pilot SNR for 2-state model ($N = 2$, $L = 4$, $p_{01} = 0.2$, $p_{11} = 0.8$).

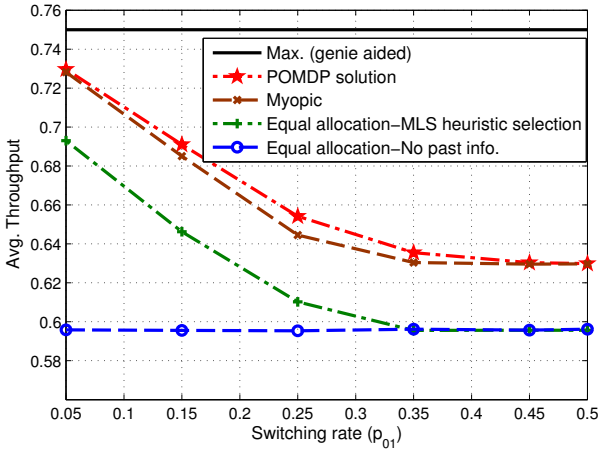

Fig. 3. Avg. Throughput vs. Switching rate $p_{01}(= 1 - p_{11} = p_{10})$ for 2-state model ($N = 2$, $L = 4$, $\bar{p}_1 = 0.5$, pilot SNR = 3 dB).

the good state, and selects the antenna using the same criterion, based on the current belief. `Equal allocation-MLS heuristic selection` uses an equal pilot allocation in all frames and the Maximum Likelihood State (MLS) heuristic solution [21] for AS. `Equal allocation-No past info` plots the performance of a scheme that uses an equal pilot allocation and makes AS decisions based only on the current training phase observation in each frame; i.e., for a given observation $\boldsymbol{\theta}$, the AS decision is $n = i$ if $\theta_i = 1$, and $\theta_j \neq 1 \forall j < i$, where $i, j \in \{1, \ldots, N\}$.

1) Variation of Throughput with Pilot SNR: Figure 2 shows the variation of throughput with the pilot SNR (dB). Here, the channel transition probabilities are $p_{01} = 0.2$ and $p_{11} = 0.8$, and hence the stationary probability of being in the good state is $\bar{p}_1 = 0.5$ for each channel. For pilot SNRs from $3 - 5$ dB, `POMDP solution` offers a throughput gain of around 12% compared to `Equal allotment-No past info`. Also, for the same packet success rate, `POMDP solution` requires a $4 - 8$ dB lower pilot SNR than `Equal allotment-No past info`. `Myopic` performs slightly worse than `POMDP solution`. For channel sensing in cognitive radio, `Myopic` was shown to be optimal [11] when there are only 2 channels to choose from. Figure 2 shows that when $N = 2$, `Myopic` is a very good heuristic for AS as well. This is not surprising, as the error check on the data packet provides accurate information about the channel state, and although the actions taken by `Myopic` are suboptimal, the margin of error against `POMDP solution` is small with $N = 2$ and $L = 4$. Both `Myopic` and `POMDP solution` require belief update operations at each decision point, but `Myopic` does not require offline planning and the online table look-up operations as `POMDP solution` does. At high pilot SNRs, all the schemes tend to the maximum attainable limit.

2) Variation of Throughput with Switching Rate: The variation of average throughput with the switching rate p_{01} at 3 dB pilot SNR is shown in Figure 3. The switching rate is the probability of transiting to state 1 in the next frame, given that the current state is 0, and vice versa. In this paper, the switching rate is assumed to be known; in practice, it depends on the Doppler spread or velocity of the receiver, which can be estimated. Here, $\bar{p}_1 = 0.5$ and hence, $p_{11} = 1 - p_{01}$. With both $\bar{p}_1$ and the pilot SNR fixed, `Equal allotment-No past info` does not show any performance variation with p_{01}, as this scheme does not use information from link-layer error checks to optimize AS or pilot allotment decisions. The performance of `POMDP solution` decreases as p_{01} varies from 0 to 0.5, and it provides maximum gain ($\approx 22\%$ over `Equal allocation-No past info`) when p_{01} is low. Thus it appears that `POMDP solution` is most suited to slowly varying channel scenarios. However, even when $p_{01} = 0.5$, `POMDP solution` performs better than the equal allocation scheme, though matched by `Myopic`. Hence, an unequal pilot allocation is beneficial even when p_{01} approaches 0.5. For $p_{01} = 0.5$, the POMDP solution is observed to be somewhat simple, allotting all $L = 4$ pilots to the first antenna in every frame, and changing only the selection decision based on the current belief state. Thus, surprisingly, when the channels at the antennas are equally likely to transition to either state, it is better to put all the pilots on one antenna and track it constantly with a high accuracy, rather than use an equal allocation and get estimates that are less accurate. If the channel at this antenna is observed to be in the good state in the training phase, the receiver uses it for data reception, and otherwise, it receives the data on the other antenna. From Figure 3, when $N = 2$, close to optimal behavior can be achieved for the whole range of p_{01} by the `Myopic` policy.

VI. Conclusion

In this paper, the sequential decision problem faced by a multiple antenna receiver with a single RF chain, of determining how accurately the channel at a particular antenna should be estimated, and selecting the best antenna in each frame, so as to maximize throughput, was modeled as a POMDP. The solution to the POMDP yielded the policy based on the past

decision and observation history for making the joint decision of the number of pilot symbols to be used for estimating the channel at each antenna, and the antenna to be used for data reception. Numerical examples showed that the POMDP solution outperforms other existing schemes. The POMDP solution is particularly useful at low pilot SNRs and can save several dB of pilot power to achieve the same throughput as other existing schemes. For a 2-state Markov channel model with $N = 2$ antennas and a switching rate $p_{01} = 0.5$, the POMDP solution gave a surprising policy, where the receiver allotted all the pilots to the same antenna in all frames, and changed only the AS decision according to the current belief state. Further, with 2 receiver antennas, the myopic policy was found to perform nearly optimally, and hence can be a good alternative to finding and implementing more complex optimal policies. Future work could consider continuous observations in the training phase, selecting a subset of antennas, AS at both the transmitter and receiver, and in multi-user communication systems, all in a decision theoretic framework.

APPENDIX

Here, $h_i[k] \in \{h_0, h_1\}$, where h_0 is the bad state and h_1 the good state and their values are known to the receiver. The receiver then has a detection problem at hand in the training phase, and $h_i[k]$ can be written as $h_i[k] = x_i(h_0 - h_0) + \frac{1}{2}(h_0 + h_1)$, with $x_i \in \left\{-\frac{1}{2}, \frac{1}{2}\right\}$ being the value to be detected, where $x_i = +\frac{1}{2}$ corresponds to h_0 and $x_i = -\frac{1}{2}$ corresponds to h_1. Also, $S_i[k] = 0$ if $h_i[k] = h_0$ and $S_i[k] = 1$ if $h_i[k] = h_1$. Let $\mathbf{v} \triangleq \frac{h_0 - h_1}{|h_0 - h_1|}\frac{\mathbf{p}}{\|\mathbf{p}\|}$, where $\mathbf{p}$ is as in (1), dropping the antenna index i. Then, from (1), $\tilde{y} \triangleq \mathbf{v}^H\left[\mathbf{y} - \frac{1}{2}(h_0 + h_1)\mathbf{p}\right] = x\,|h_0 - h_1|\,\|\mathbf{p}\| + w$, where $w \sim \mathcal{CN}(0, \sigma^2)$. Since x is real-valued, $\Re\{\tilde{y}\}$ is sufficient [22] to detect x. Conditioned on x, $\Re\{\tilde{y}\}|x \sim \mathcal{N}\left(x|h_1 - h_2|\|\mathbf{p}\|, \frac{\sigma^2}{2}\right)$. In this case, obtaining a MAP decision rule is straightforward, and the observation of the channel state of antenna i is given by

$$\Theta_i[k] \triangleq \begin{cases} 1, & \text{if } \lambda_i[k] \geq \eta_i \\ 0, & \text{otherwise,} \end{cases} \tag{6}$$

where $\lambda_i[k] \triangleq \ln \frac{\mathbb{P}_{\ell_i}\{\tilde{y}_i[k]|S_i[k]=1\}}{\mathbb{P}_{\ell_i}\{\tilde{y}_i[k]|S_i[k]=0\}} = \sqrt{\frac{\ell_i E_p}{L}}\frac{|h_0 - h_1|\Re\{\tilde{y}_i[k]\}}{\sigma^2/2}$, and

$$\eta_i \triangleq \ln \frac{\mathbb{P}\{S_i[k] = 0\}}{\mathbb{P}\{S_i[k] = 1\}} = \ln \frac{1 - p_{11}^{(i)}}{p_{01}^{(i)}}. \tag{7}$$

If $\ell_i = 0$, no observations are made on antenna i, and the belief vector is updated using the transition probabilities of the Markov chain. Now the observation probabilities $\mathbb{P}_A\{\boldsymbol{\Theta}[k_0] = \boldsymbol{\theta}|\mathbf{S}[k_0] = \mathbf{s}_0\}$ and $\mathbb{P}_A\{Z[k_1] = z|\mathbf{S}[k_1] = \mathbf{s}_1\}$ are derived. When a likelihood ratio-based detector of the channel state is used as described above, it can be shown that[3] $\mathbb{P}_A\{\Theta_i = 1|s_{0,i}\} = Q\left(\kappa_i\left(\frac{\eta_i}{\kappa_i^2} - x_i\right)\right)$ where $Q(\cdot)$ is the Gaussian Q-function, $\kappa_i = |h_0 - h_1|\sqrt{\frac{2\ell_i E_p}{L\sigma^2}}$ and η_i is given by (7). For the data reception phase, the observation $Z = 1$ only if $S_n = 1$. Thus, $\mathbb{P}_A\{Z = 1|S_n = s_{1,n}\} = \mathbb{1}_{\{s_{1,n}=1\}}$, where $\mathbb{1}_{\mathcal{A}}$ is the indicator function, taking the value 1 when event $\mathcal{A}$ is true, and zero otherwise.

[3]Frame/sub-frame indexes are dropped for convenience.

REFERENCES

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

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

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

[4] V. Kristem, N. Mehta, and A. 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.

[5] 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.

[6] T. Ramya and S. Bhashyam, "Using delayed feedback for antenna selection in MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 6059–6067, 2009.

[7] D. Gore and A. Paulraj, "Statistical MIMO antenna sub-set selection with space-time coding," in *Proc. ICC*, vol. 1. IEEE, 2002, pp. 641–645.

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

[9] R. Smallwood and E. Sondik, "The optimal control of partially observable Markov processes over a finite horizon," *Oper. Res.*, pp. 1071–1088, 1973.

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

[11] S. Ahmad, M. Liu, T. Javidi, Q. Zhao, and B. Krishnamachari, "Optimality of myopic sensing in multichannel opportunistic access," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 4040–4050, 2009.

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

[13] V. Kristem, N. Mehta, and A. Molisch, "Training and voids in receive antenna subset selection in time-varying channels," *IEEE Trans. Wireless Commun.*, no. 99, pp. 1–12, 2011.

[14] E. Gilbert *et al.*, "Capacity of a burst-noise channel," *Bell Syst. Tech. J*, vol. 39, no. 9, pp. 1253–1265, 1960.

[15] M. Zorzi, R. Rao, and L. Milstein, "On the accuracy of a first-order Markov model for data transmission on fading channels," in *Proc. Int. Conf. Universal Personal Commun.* IEEE, 1995, pp. 211–215.

[16] A. Chockalingam, M. Zorzi, L. Milstein, and P. Venkataram, "Performance of a wireless access protocol on correlated Rayleigh-fading channels with capture," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 644–655, 1998.

[17] 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.*, (Submitted).

[18] S. Sesia, I. Toufik, and M. Baker, *LTE-the UMTS long term evolution: from theory to practice*. Wiley, 2011.

[19] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 2nd ed. Athena Scientific, 2000.

[20] [Online]. Available: http://bigbird.comp.nus.edu.sg/pmwiki/farm/appl/index.php?n=Main.HomePage

[21] R. Simmons and S. Koenig, "Probabilistic robot navigation in partially observable environments," in *International Joint Conference on Artificial Intelligence*, vol. 14, 1995, pp. 1080–1087.

[22] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge Univ. Pr., 2005.