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

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

***Abstract*—This paper considers antenna selection (AS) at a receiver equipped with multiple antenna elements but only a single radio frequency chain for packet reception. As information about the channel state is acquired using training symbols (pilots), the receiver makes its AS decisions based on noisy channel estimates. Additional information that can be exploited for AS includes the time-correlation of the wireless channel and the results of the link-layer error checks upon receiving the data packets. In this scenario, the task of the receiver is to sequentially select (a) the pilot symbol allocation, i.e., how to distribute the available pilot symbols among the antenna elements, 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 past history of allocation and selection decisions, and the corresponding noisy channel estimates and error check results. Since the channel state is only partially observed through the noisy pilots and the error checks, the joint problem of pilot allocation and AS is modeled as a partially observed Markov decision process (POMDP). The solution to the POMDP yields the policy that maximizes the long-term expected throughput. Using the Finite State Markov Chain (FSMC) model for the wireless channel, the performance of the POMDP solution is compared with that of other existing schemes, and it is illustrated through numerical evaluation that the POMDP solution significantly outperforms them.**

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

I. Introduction

Antenna selection (AS) [1]–[3] is a powerful technique, employed to reduce hardware costs at the transmitter and/or receiver of a multiple antenna wireless link. The core idea is to use a limited number of radio frequency (RF) chains, and adaptively switch them to subsets of a larger number of available antenna elements. AS achieves the same diversity order as a system that uses all the antenna elements, and hence only a small loss in data rate is suffered when the receiver uses the best possible subset of the available receive antennas [2]. AS can be employed at the transmitter, receiver, or both ends; this work focuses on *receive* AS.

Several criteria for AS have been considered, and several algorithms assuming perfect channel state information (CSI) at the receiver have been proposed ([4], [5], and references therein). In practice, it is necessary to estimate CSI using, for example, a pilot-based training scheme, and imperfect CSI can lead to both inaccurate AS and erroneous decoding of data, increasing the symbol error probability (SEP) [6]. Somewhat surprisingly, it has been shown that transmit and receive AS can achieve full diversity order even in the presence of channel estimation errors [7].

Most of the past work on AS suffers from three drawbacks. First, it assumes that the receiver equally divides the available pilot symbols by the number of antenna elements during the training phase [6], [8], [9]. However, when the channel is slowly-varying, such an equal allocation is not optimal, as the receiver can use the past estimates of the channel and the time-correlation information to reallocate pilots among antennas in subsequent training periods. Second, with packet reception, the receiver can use link-layer error checks on the data packets to glean additional information on the channel; this aspect is typically not exploited in the literature. An exception, for example, is [10], where an expression for the link-layer throughput is used as a metric for transmit antenna subset selection. Third, a quasi-static block-fading channel model is usually assumed [1], [11], which precludes the receiver from fully exploiting the temporal channel correlation. This work seeks to overcome the aforementioned drawbacks and fully exploit all the available information in deciding the optimal pilot allocation for channel estimation and AS for data packet reception.

The system model considered 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. To this end, the transmitter sends the data in frames, with each frame consisting of 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 most of the pilots out of the available L to one particular antenna, getting an accurate estimate of the channel on that antenna. However, this would lead to losing track of the channels on the other antennas that could have possibly been better. On the other hand, it could allot fewer pilots each to different antennas and keep track of the channels at all the antennas. However, now it has poorer quality estimates of channels on a larger number of antennas, which would lead to errors in subsequent selection decisions, and loss of data packets.

The actual state of the channel at each available receive antenna is known to the receiver only through noisy estimates and error checks on the data packet. The receiver can control the accuracy with which to estimate the channel at a particular

*R. G. Stephen is with the Center for Development of Telematics, Bangalore, India. He was with the Dept. of ECE, Indian Institute of Science (IISc), Bangalore, during the course of this work. Email: reubenstephen@gmail.com.

†C. R. Murthy is with the Dept. of ECE, IISc. Email: cmurthy@ece.iisc.ernet.in

‡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. Email: marceau.coupechoux@telecom-paristech.fr

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.

antenna, and can select the antenna to be used for packet reception. Thus, it has the freedom to control the partial observability of the system. These controls must be applied in such a way as to maximize some notion of long-term reward. As a consequence, in this work, the problem of joint pilot allotment and AS at the receiver in each frame is modeled as a Partially Observable Markov Decision Process (POMDP) [12]–[14], 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 channel is solved using a decision-theoretic framework. A challenge in the formulation is that it needs to be able to deal with two different kinds of actions, namely, the pilot allocation and AS decisions, and two different observations in the training and data phases. This is further elaborated in Section III.
- The POMDP is solved to obtain the joint pilot allocation and AS policy that maximizes the long-term reward such as the throughput.
- Insights are provided on the nature of the policies to be followed. For example, with 2 receive antennas and a 2-state Markov model for the wireless channel, when the channel is fast-varying, it is found, somewhat surprisingly, that the POMDP solution allots all pilot symbols to a single antenna, which allows it to glean accurate information about that antenna, and the selection decision picks the antenna that is most likely to be in a good state.
- With a Finite State Markov Chain (FSMC) model for the wireless channel, it is shown via numerical evaluation that employing the POMDP policy can lead to considerable savings in the pilot power required to achieve the same packet success rate as compared to other existing schemes. These can be up to 8 dB at a fixed average data SNR of 0 dB, and around $2-3$ dB with 3 dB pilot power boosting, for moderate values of pilot SNR. With a 2-state Markov Chain model and $N=2$ antennas and $L=4$ training symbols, it is found that the greedy myopic policy [15] is nearly optimal over a wide range of channel parameters and pilot and data SNR values.

The advantage of posing the problem as a POMDP is that it admits the use of a gamut of computationally efficient methods [16, and references therein] for solving it. Moreover, once the solution is obtained, implementing the optimal policy to optimally allot pilot symbols and select the antenna to receive the data packet is simple. One has to update the belief vector for the channel state based on the observations in every slot using Baye's 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 number of pilot symbols required to obtain a given performance, or an improvement in the average data rate, in practical AS based systems.

The paper is organized as follows. Section II gives a description of the system model. Section III describes the POMDP formulation of the problem and Section IV describes the solution techniques. This is followed by a discussion in Section V. Simulation results are presented in Section VI. Appendices A and B detail the observation models used and the calculation of observation probabilities required for POMDP planning.

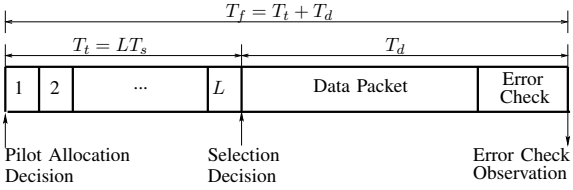

Fig. 1. Frame structure for training and data reception.

Notation In the paper, x indicates a scalar and $\mathbf{x}$, a vector. x^*, $\mathbf{x}^T$ and $\mathbf{x}^H$ denote the conjugate of x, and the transpose and Hermitian of $\mathbf{x}$ respectively. x_j is the j^{th} component of $\mathbf{x}$ and $x_{i,j}$, the j^{th} component of vector $\mathbf{x}_i$. X is a random variable (RV), and $\mathbf{X}$, a random vector, unless mentioned otherwise. $f_X(x)$ denotes the probability density function (pdf) of the continuous RV X. $\mathbf{A}_n$ indicates a square matrix of size n. $\mathcal{CN}(\mu,\sigma^2)$ and $\mathcal{N}(\mu,\sigma^2)$ denote respectively a complex and real normal distribution with mean μ and variance σ^2. $\mathbb{E}_A\{X\}$ denotes expectation of X, given condition A, and $\mathbb{P}_A\{E\}$, the probability of event E, given condition A. $\mathbb{1}_{\{A\}}$ denotes the indicator function, equal to 1 if condition A is satisfied, and 0 otherwise.

II. System Model

Consider a wireless system with a single transmit antenna and N receive antenna elements, but only a single RF chain at the receiver. Time is divided into frames of fixed duration T_f. Each frame consists of a training period T_t and a data transmission period T_d. In the training period, L reference pilot symbols, each of duration T_s, are received and used by the receiver to estimate the channel gains at the N receive antennas. The training period is followed by data packet transmission, at the end of which the receiver performs an error check and hence knows whether the data packet was received without error 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 beginning of frame k. It is assumed that $h_i[k]$ is constant for the entire duration T_f of frame k, but correlated from frame to frame. This holds true if the coherence time T_c of the channel satisfies $T_c \gg T_f$, as is typically the case in practice. Also, $h_i[k]$ is independent of $h_j[k]$ for $i \neq j$. This assumption, though not necessary for the POMDP formulation of the problem, simplifies the evaluation. It holds true if the antenna elements at the receiver are spaced sufficiently apart from each other.

Consider a particular frame in which $\ell_i \in \{0,1,\ldots,L\}$ pilots are used to estimate the channel[1] h_i at receive antenna i, with $\sum_{i=1}^{N} \ell_i = L$. Here, the time overhead of switching between antennas is assumed to be negligible compared to the duration of the training phase [2], and

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

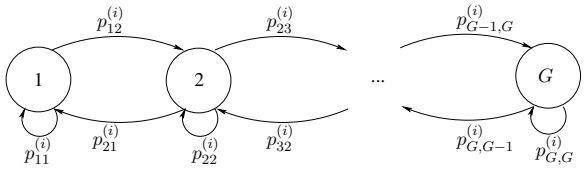

Fig. 2. The FSMC channel model for the i^{th} antenna.

hence is ignored. It is common in AS literature to assume that different pilot symbols within the training period can be received on different antenna elements [6], [8], [17]. If $\mathbf{y}_i = \begin{bmatrix} y_1 & \cdots & y_{\ell_i} \end{bmatrix}^H \in \mathbb{C}^{\ell_i}$ denotes the vector of received symbols and $\mathbf{p}_i = \sqrt{\frac{E_p}{L}} \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}^T$ is the ℓ_i-length vector of pilot symbols 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 $\ell_i > 0$, where $\mathbf{w}_i \in \mathbb{C}^{\ell_i}$ is the additive white Gaussian noise (AWGN) vector with $\mathbf{w}_i \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_{\ell_i})$. When $\ell_i = 0$, no symbol is received on the i^{th} antenna.

In the sequel, the time-correlated channel is modeled as an FSMC, as shown in Figure 2. An FSMC model has been widely used to characterize block or packet level performance measures in correlated Rayleigh fading channels [18], [19]. Zhang and Kassam [20] establish the relationship between a physical fading channel and its FSMC model for a packet transmission system, and the same approach is adopted here. The idea is to partition the received SNR values into a finite number of states according to a criterion based on the average duration of each state.

Let $\mathcal{G} = \{1, 2, \ldots, G\}$ denote the state space of the stationary Markov chain with $|\mathcal{G}| = G$ different channel states corresponding to SEPs $P_{e,j}$, $j = 1, \ldots, G$. Let $\{\gamma_1, \gamma_2, \ldots, \gamma_{G+1}\}$ denote received SNR thresholds in increasing order, with $\gamma_1 = 0$ and $\gamma_{G+1} = \infty$. State j of the Markov chain corresponds to $\gamma_j \leq \gamma < \gamma_{j+1}$. For the packet reception scenario considered in this work, a one-step transition in the model corresponds to the channel state transition after one frame period T_f. Transitions are such that if p_{ij} is the transition probability from state i to state j, $p_{ij} = 0$ $\forall j \notin \{i-1, i, i+1\}$. With Rayleigh fading in AWGN, the received instantaneous SNR γ is exponentially distributed with pdf $f_\Gamma(\gamma) = \frac{1}{\bar{\gamma}} e^{-\frac{\gamma}{\bar{\gamma}}}, (\gamma \geq 0)$, where $\bar{\gamma}$ is the average SNR. The SNR thresholds and transition probabilities can be found by solving a set of equations [20, Eq. (7)], with the requirement that the average time duration of each state, $\bar{\tau}_j$, satisfy $\bar{\tau}_j = cT_f$ for $j = 1, \ldots, G$, where $c > 1$ is a constant. 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 assumption 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 (SRS) for channel quality estimation, and demodulation reference signals (DM RS) for channel estimation during coherent demodulation [21]. The focus of this work is on finding optimal AS decision policies under partial observability of the channel state due to channel estimation errors.

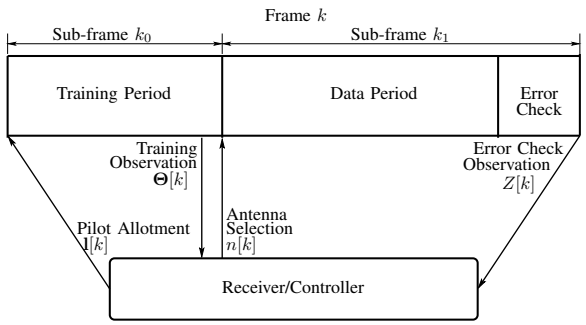

Fig. 3. Sequence of operations in frame k.

In addition to the FSMC model described in the above paragraph, the popular 2-state Markov model is also considered here [22]. In the 2-state model, the channel is reduced to two gain states h_1 and h_2, where packets are assumed to be received without error when the channel state on the selected antenna is h_2, while an error occurs with certainty when the channel state is h_1.

The observation models and corresponding probabilities in the training and data phase for the FSMC as well as the 2-state model are described in Appendices A and B. Particular models are used to derive these probabilities for concreteness in the subsequent development, but other observation models can also be used in the POMDP framework that is constructed in this work.

The sequence of operations that the receiver follows in each frame is illustrated in Figure 3 and described below. Let $\mathbf{S}[k] = \begin{bmatrix} S_1[k] & \cdots & S_N[k] \end{bmatrix}^T$ denote the state vector of the channels at the N antenna elements in frame k, with $S_i[k] \in \{1, \ldots, G\}$, $i = 1, \ldots, N$. Here, $S_i[k] \triangleq j$ if the receive SNR γ is such that $\gamma_j \leq \gamma < \gamma_{j+1}$. In Figure 3, k_0 and k_1 represent the training and data sub-frames, respectively. At the beginning 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 obtained for each antenna i, and the receiver determines the antenna $n \in \{1, \ldots, N\}$ to be used to receive the data packet. Appendix A describes how the observations $\Theta_i[k]$ are obtained, from (11) for the FSMC model, or from (12) for the 2-state model. 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 $\theta_i \in \{1, \ldots, G\}$. 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). This provides additional information on the channel state at the antenna selected in the current frame, and is used in future frames to determine the pilot allocation and make AS decisions.

III. POMDP Formulation

The sequential decision-making process described above is now formalized as a POMDP. In this particular case, two different actions, namely, pilot allocation and AS, are to be taken at different points in a single frame, and two different observations can be obtained in the training and data phase, while the channel state remains the same. In the classical POMDP framework, the actions belong to a single set at all decision points. Hence, the pilot allocation and AS decisions are combined to form a single action that has both these components, and taken at both the beginning of the training phase and the start of the data phase. Also, an observation is produced by performing an action in a particular state, i. e., an observation corresponds to a single state-action combination, and only one observation can be obtained when taking an action in a given state. This mandates a distinction between the state of the system in the training and the data phase, and hence the state space is expanded with an additional variable $m \in \{0,1\}$ that represents the two different decision points in a single frame. This state-space expansion is necessary to bring the joint pilot allocation and AS problem to a standard form, where it can be solved using available POMDP solution methods.

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 assumed to be constant over a frame, transitions are naturally restricted so that

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

where $\tilde{\mathbf{s}}, \mathbf{s} \in \{1,\ldots,G\}^N$, $\mathbf{s}_1 \triangleq \begin{bmatrix}\mathbf{s} & 1\end{bmatrix}^T$ and $\mathbf{s}_0 \triangleq \begin{bmatrix}\mathbf{s} & 0\end{bmatrix}^T$. Here $\mathbf{s} = \begin{bmatrix}s_1 & \cdots & s_N\end{bmatrix}^T$, 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

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

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

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

2) Action Space: The action consists of two parts within a frame:

- A pilot allocation vector $\mathbf{l} = [\ell_i]_{i=1}^N \in \mathcal{L}$, where $\mathcal{L} \triangleq \left\{\mathbf{l} : \ell_i \in \{0,\ldots,L\}, \sum_{i=1}^N \ell_i = L\right\}$. For a given N and L, $|\mathcal{L}| = \binom{N+L-1}{L}$.
- 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 $k_m = 1_0, 1_1, 2_0, 2_1, \ldots$ 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 consists of 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 $\Omega_0 \triangleq \{1,\ldots,G\}^N$, 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 $|\Omega| = G^N + 2$ for the G-state channels considered here. The probabilities of observing $z \in \Omega_m$ satisfy $\mathbb{P}_A\left\{z \in \Omega_1 | \mathbf{s}_0\right\} = \mathbb{P}_A\left\{z \in \Omega_0 | \mathbf{s}_1\right\} = 0$.

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:

$$R(\mathbf{s}_m, A[k_m]) = m\mathbb{P}_A\left\{Z[k_m] = 1 | \mathbf{s}_m\right\} \cdot B. \tag{4}$$

In the sequel, $B = 1$ is assumed without loss of generality. Thus, $R(\mathbf{s}_0, A[k_0]) = 0\ \forall k$, since 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 $\mathbf{l}[k_0]$, does indirectly affect 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, after applying a discounting factor for future rewards, that can be delivered.

5) Belief Vector: With a Markovian evolution of the states, it is known that the entire decision and observation history can be encapsulated in a belief vector $\mathbf{b}[k_m] \triangleq [b_{\mathbf{s}_m}[k_m]]_{\mathbf{s}_m \in \mathcal{S}}$ [13]. 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 channel state vector 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 [14], 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 long-term reward, measured as 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, the optimal policy is given by

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

where $\beta \in [0,1)$ is the discount factor [23], and the exponent $q \triangleq 2(k-1)+m$, $\forall k, m$.

IV. Solving the POMDP

Let $V(\mathbf{b}[k_m])$ denote the value function [14], which represents the *maximum* expected discounted reward that can be obtained, starting in the belief state $\mathbf{b}[k_m]$. According to the notation introduced here, when the receiver takes action $A[k_m+1] = A \in \mathcal{A}$ and observes $z[k_m+1] = z \in \Omega_{m'}$, where $m' = 1-m$, $m \in \{0,1\}$, the reward that can be accumulated starting from point k_m+1 consists of two parts:

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

where k_m+1 is as defined in (3) and $\mathbf{s}'_{m'}$ denotes 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:

$$V\left(\mathbf{b}[k_0]\right) = \max_{A \in \mathcal{A}} \sum_{\mathbf{s}_0 \in \mathcal{S}} b_{\mathbf{s}_0}[k_0] \sum_{z \in \Omega_1} \mathbb{P}_A\left\{z|\mathbf{b}[k_0]\right\} \left[z \cdot 1 + \beta V(f(\mathbf{b}[k_0], A, z))\right], \quad (5)$$

and

$$V(\mathbf{b}[k_1]) = \max_{A \in \mathcal{A}} \sum_{\mathbf{s}_1 \in \mathcal{S}} b_{\mathbf{s}_1}[k_1] \sum_{\boldsymbol{\theta} \in \Omega_0} \beta \mathbb{P}_A\left\{\boldsymbol{\theta}|\mathbf{b}[k_1]\right\} V(f(\mathbf{b}[k_1], A, \boldsymbol{\theta})). \quad (6)$$

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

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

Note that two equations are needed to represent the value function updates in the training and data phases to account for the dual observations, actions, decision points and immediate rewards, and these need to be simultaneously satisfied, as opposed to the traditional POMDP value updates, where only one equation is needed.

The term $\mathbb{P}_A\left\{\boldsymbol{\theta}|\mathbf{s}'_{m'}\right\} = \mathbb{P}_A\left\{\boldsymbol{\Theta}[k_m+1] = \boldsymbol{\theta}|\mathbf{S}[k_m+1] = \mathbf{s}'_{m'}\right\}$ denotes the conditional probability mass function (pmf) of the channel state observation vector, given the landing state $\mathbf{S}[k_m+1] = \mathbf{s}'_{m'}$ and action $A[k_m+1] = A$, and $\mathbb{P}_A\left\{z|\mathbf{s}'_{m'}\right\} = \mathbb{P}_A\left\{Z[k_m+1] = z|\mathbf{S}[k_m+1] = \mathbf{s}'_{m'}\right\}$ denotes the corresponding 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

$$\mathbb{P}_A\left\{\boldsymbol{\Theta}[k_0] = \boldsymbol{\theta}|\mathbf{s}_0\right\} = \prod_{i=1}^{N} \mathbb{P}_{\ell_i}\left\{\Theta_i[k_0] = \theta_i|s_{0,i}\right\}. \quad (7)$$

Similarly,

$$\mathbb{P}_A\left\{Z[k_1] = z|\mathbf{s}_1\right\} = \mathbb{P}_n\left\{Z[k_1] = z|s_{1,n}\right\}, \quad (8)$$

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

$$b_{\mathbf{s}'_{m'}}[k_m+1] = \mathbb{P}\left\{\mathbf{S}[k_m+1] = \mathbf{s}'_{m'}|\mathbf{b}[k_m], A, z\right\} = \frac{\mathbb{P}_A\left\{z|\mathbf{s}'_{m'}\right\} \sum_{\mathbf{s}_m \in \mathcal{S}} b_{\mathbf{s}_m}[k_m] \mathbb{P}\left\{\mathbf{s}'_{m'}|\mathbf{s}_m\right\}}{\sum_{\mathbf{s}'_{m'} \in \mathcal{S}} \mathbb{P}_A\left\{z|\mathbf{s}'_{m'}\right\} \sum_{\mathbf{s}_m \in \mathcal{S}} b_{\mathbf{s}_m}[k_m] \mathbb{P}\left\{\mathbf{s}'_{m'}|\mathbf{s}_m\right\}} \quad (9)$$

Since the channels at the antennas are assumed to be independent, $\mathbb{P}\left\{\mathbf{s}'|\mathbf{s}\right\} = \prod_{i=1}^{N} \mathbb{P}\left\{s'_i|s_i\right\}$. Probabilities $\mathbb{P}\left\{s'_i|s_i\right\}$ can be obtained from the transition probabilities of the FSMCs. Observation probabilities $\mathbb{P}_A\left\{z|\mathbf{s}'\right\}$ can be obtained using the ML criterion as described in Appendix B, from (18), (19) and (20) for the data phase, or from (7) and (17) for the training phase.

Except for small problems with less than 10 states and actions, exact algorithms [13], [24], [25] for solving POMDPs are computationally infeasible, and hence for larger problems, point-based algorithms [16], [26], [27] are preferred. The latter apply the value backup operation in (5) and (6) only on a finite subset of carefully chosen belief points. An example is the Successive Approximation of the Reachable Space under Optimal Policies (SARSOP) algorithm [16], which is used in this work to solve the POMDP formulated in Section III. SARSOP has been demonstrated to outperform recent point-based POMDP solvers in terms of computational efficiency [16]. However, any other efficient POMDP solving algorithm may also be used to solve the problem under consideration. In this context, it is important to note that the complexity of finding the solution does not impose a computational burden on the receiver, as this can be performed offline given the channel model. The real-time computations that need to be performed at the receiver involve updating the belief vector using (9), and then applying the solution to the POMDP corresponding to the updated belief, possibly by using a look-up table.

V. Discussion

The joint problem of pilot allocation and antenna selection in a time-correlated channel is solved using a decision theoretic framework for the first time in this work. A POMDP solution fully exploits the information from past pilot allocation and AS decisions, training and data observations, and the statistics of the channel variations at the antenna elements, and hence outperforms all other methods as illustrated in Section VI.

As mentioned in Section II, the time overhead for switching between antennas is neglected in this work. However, with the model under consideration, this can be included by accounting for the loss of, for example, M pilot symbols for each switch between antennas in the training phase. This would imply a correspondingly lower pilot SNR on the different antennas for the particular frame, depending on the pilot allocation chosen.

In this paper, the pilot allocation for the whole training period is decided at the start of the frame. Alternatively, one could consider performing pilot allotment on a symbol-by-symbol basis in the training period, based on beliefs that are updated upon reception of each pilot symbol. An in-depth analysis of such a scheme requires a new study, and is relegated to future work.

The assumption that the channels at the antennas are mutually independent is used to factorize the probabilities in (7) and (8), and this simplifies the derivations in Appendix B. Relaxing this assumption would affect the calculation of the observation probabilities and the behavior of the optimal policy. However, the POMDP framework in Section III can still be applied when the channels are correlated.

The observations in the training phase are restricted to a finite set in this work. This is necessary for POMDP planning using existing algorithms. Dealing with continuous observations requires different planning techniques [28] and is a topic for future work. Further, the POMDP framework is a model-based approach and needs a Markovian model for the system. Model-free approaches require techniques like reinforcement learning [29].

Given the large size of the state space, it is, in general, difficult to obtain closed-form analytical solutions to the POMDP. Here the number of variables is large due to the several different pilot allocations possible and the number of antennas. The relative impact of these on the observability of the channels and other factors like channel correlation, discount factor, etc. make it hard to analyze the nature of the POMDP solution. However, one can come up with heuristic policies that perform well, which give insights into the optimal solution. This is elaborated in Section VI.

In the unrelated context of Cognitive Radios (CR), a POMDP formulation has been followed earlier [30], where the CR is required to sense a subset of potentially available channels, but can access channel(s) for data transmission only from the subset of channels that were sensed in the current slot. In contrast, in this work the receiver can estimate the channels at some or all the available antennas with varying degrees of accuracy by using $0, 1, \ldots, L$ pilot symbols on each antenna. Also, it has the freedom to select an antenna that it did not estimate during the training phase for data reception. The problem is thus more general than that in Chen et al. [30].

VI. Simulation Results

The POMDP formulated in Section III is solved using the Approximate POMDP Planning Toolkit [31], implementing the SARSOP algorithm [16]. In all cases described below, the code is run until the error between value functions obtained in consecutive steps falls below a tolerance limit of $\epsilon = 1$. The discount factor $\beta = 0.99$ in all cases, since a large value of β is relevant 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 G-state Markov chain as described in Section II. The number of antenna elements $N = 2$ and pilot symbols in each frame $L = 4$ in all cases, except in Figures 5 and 6, where $N = 4$ and $L = 4$.

Performance is evaluated over 2×10^3 sub-frames and the POMDP solution is compared with other existing schemes. In the evaluations, the following curves are shown.

1) `Max. (genie aided)` shows the maximum attainable throughput when the receiver has perfect knowledge of the channel states at all the antennas.
2) `POMDP solution` shows the performance of the policy obtained by solving the POMDP.
3) `Equal allocation-POMDP selection` is a scheme that always uses an equal pilot allocation of $\lfloor \frac{L}{N} \rfloor$ pilots on each antenna, but selects the antenna optimally in each frame. This is a straightforward modification of the POMDP solution, and is obtained by solving a POMDP with a restricted action set that has only one possible pilot allocation.
4) For the 2-state channel, a purely greedy policy [15] labeled `Myopic` is also shown, which allots all L pilots to the antenna that has the maximum likelihood of being in the good state, and selects the antenna using the same criterion, based on the current belief.
5) `Equal allocation-MLS heuristic selection` uses an equal pilot allocation and the Maximum Likelihood State (MLS) heuristic [32] for AS. The MLS heuristic is a popular method used to find heuristic solutions to POMDPs.
6) `Equal allocation-No past info` plots the performance of a scheme that uses an equal pilot allocation and makes AS decisions in each frame based solely on the current training phase observation; i.e., for a given observation $\boldsymbol{\theta}$, the AS decision is $n = \arg\max_i \theta_i$, where $\theta_i \in \{1, \ldots, G\}$. This is a natural approach to antenna selection in a quasi-static block-fading environment [33].

The effects of pilot power, the history as captured by the belief updates and the future rewards, on the optimal policy are investigated. Schemes 2 and 3 take into account both history and future rewards, while 4, 5 and 6 do not account for future rewards. Also, Schemes 3, 5 and 6 use equal pilot allocation and focus on the AS decisions, while 2 and 4 use variable pilot symbol allocation in addition to the AS decisions.

A. Variation of Throughput with Pilot SNR

1) 2-state Model: Figure 4 shows the variation of throughput with the pilot SNR (dB). Here, the channel transition probabilities are $p_{12} = 0.2$ and $p_{22} = 0.8$, and hence the stationary probability of being in the good state is $\bar{p}_2 = 0.5$ for each channel. At pilot SNRs of $3-5$ dB, `POMDP solution` offers a throughput gain of around 12% compared to `Equal allotment-No past info`. Also, for the same packet

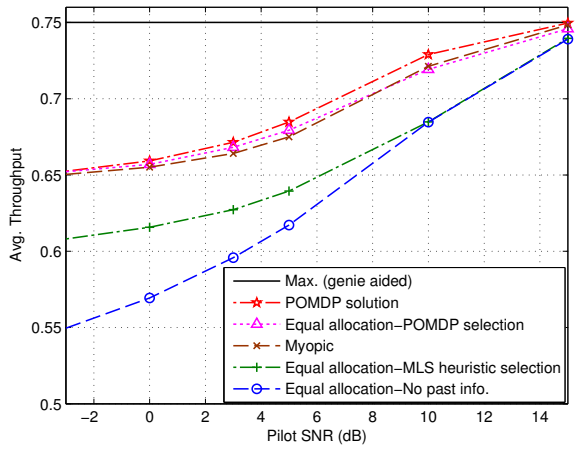

Fig. 4. Avg. Throughput vs. Pilot SNR for 2-state model ($N = 2$, $L = 4$, $p_{01} = 0.2$, $p_{11} = 0.8$).

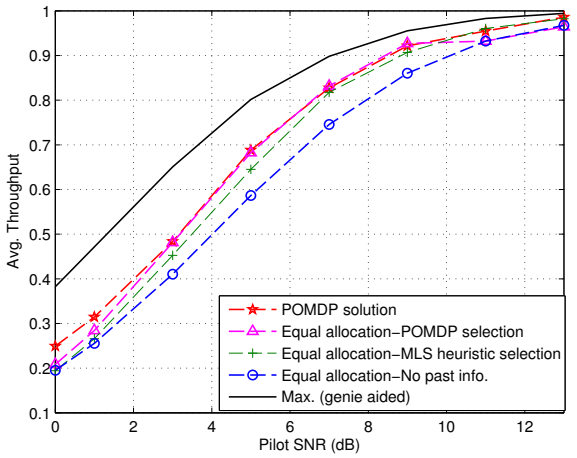

Fig. 5. Avg. Throughput vs. Pilot SNR for FSMC model ($N = 4$, $L = 4$, $G = 4$, data SNR = pilot SNR − 3 dB, $f_d T_f = 0.01$).

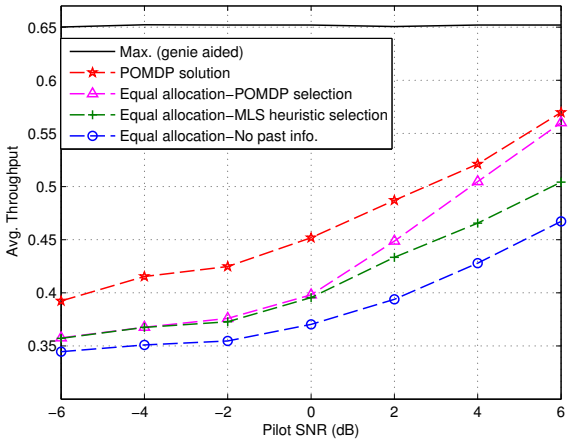

Fig. 6. Avg. Throughput vs. Pilot SNR for FSMC model ($N = 4$, $L = 4$, $G = 4$, data SNR = 0 dB, $f_d T_f = 0.01$).

success rate, `POMDP solution` requires a pilot SNR around $4-8$ dB lower than that required by `Equal allotment-No past info`. `Equal allocation-POMDP selection` performs only slightly worse than `POMDP solution`, suggesting that making the right AS decisions is more important than making the right pilot allocation decisions. In the 2-state model, the receiver gains perfect knowledge of the channel at a selected antenna, and hence learns much more about the channel at a particular antenna from the error check observation rather than the training phase observation. `Myopic` performs slightly worse than `POMDP solution`. For channel sensing in CR, `Myopic` was shown to be optimal when there are only 2 channels to choose from [15]. From Figure 4, it can be seen that when $N = 2$, the `Myopic` policy is a very good heuristic for AS in the 2-state channel model as well. This is not surprising, since the CRC bit provides accurate information about the channel state, which makes a myopic policy that is primarily based on observing the CRC bit nearly optimal. At high pilot SNRs, all the schemes tend to the maximum attainable limit.

2) FSMC Model: With the FSMC model, the packet success rate depends on the SNR state the channel is in, and hence, the variation of throughput is plotted against pilot SNR for two cases. In Figure 5, the data SNR is also varied with the pilot SNR, with $N = 4$ antennas and a pilot power boosting of 3 dB relative to the data power. Here, the normalized Doppler spread, $f_d T_f = 0.01$, which corresponds to a speed of 3 m/s at a carrier frequency $f_c = 1$ GHz, with $T_f = 1$ ms, where f_d is the maximum Doppler frequency. It can be seen that `Equal allocation-POMDP selection` performs as well as `POMDP solution`. As the data SNR ultimately determines the achievable throughput, an optimal selection policy is sufficient to ensure near-optimal performance, provided the channel is slowly-varying and the pilot SNR is sufficiently higher than the data SNR.

In Figure 6, the data SNR is kept fixed at 0 dB while the pilot SNR is varied, i.e., without pilot power boosting. Here `POMDP solution` offers considerable savings in pilot power, to achieve the same throughput as compared to `Equal allocation-No past info`. At lower pilot powers, the difference between the schemes is more pronounced. `Equal allocation-POMDP selection` performs worse than `POMDP solution` at low pilot SNRs, showing that pilot allocation is important in the low pilot SNR regime. `Equal allocation-MLS heuristic selection` offers considerable improvement over `Equal allocation-No past info`, and hence it is a useful scheme to follow in the moderate pilot power range.

From Figures 5 and 6, it can be concluded that with pilot power boosting, equal allocation is sufficient provided AS is done optimally and the channel is slowly varying. Varying the pilot allocation is useful when there is no pilot power boosting.

B. Variation of Throughput with Switching Rate/Doppler Spread

The switching rate between channel states is a measure of the time correlation of the channel in the 2-state case. When the switching rate p_{12} is low, the correlation between successive states is high and it is important to take into account

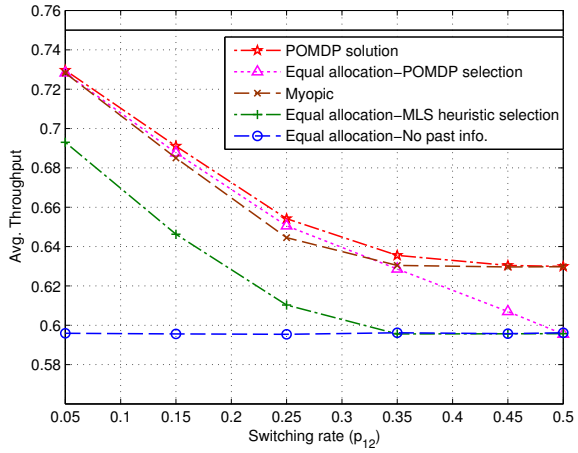

Fig. 7. Avg. Throughput vs. Switching rate $p_{12}(= 1 - p_{22} = p_{21})$ for 2-state model ($N = 2$, $L = 4$, $\bar{p}_2 = 0.5$, pilot SNR = 3 dB).

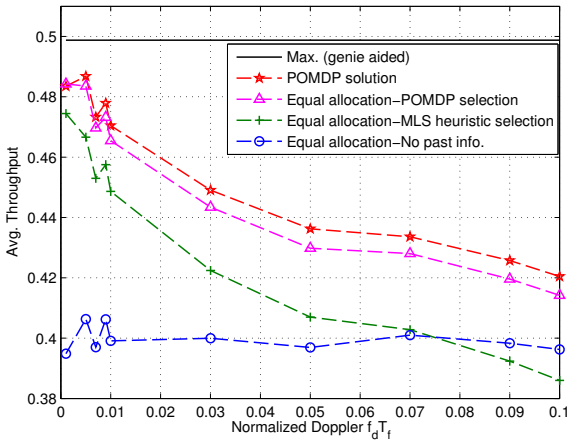

Fig. 8. Avg. Throughput vs. f_dT_f for FSMC model ($N = 2$, $L = 4$, $G = 4$, pilot SNR = 3 dB, data SNR = 0 dB).

past information to make optimal decisions. When it is high, greedy policies are expected to provide easily implementable good solutions. In the FSMC case, varying the Doppler spread has a similar effect.

1) 2-state Model: The variation of average throughput with p_{12} is shown in Figure 7. Here, $\bar{p}_2 = 0.5$ and the pilot SNR is fixed at 3 dB. Hence, $p_{22} = 1-p_{12}$. With both $\bar{p}_2$ and the pilot SNR fixed, `Equal allotment-No past info` does not show any performance variation with p_{12}. This is expected, since this scheme does not use information from link-level error checks to optimize AS or pilot allotment decisions. The performance of `POMDP solution` decreases as p_{12} varies from 0 to 0.5, and it provides maximum gain ($\approx 22\%$ over `Equal allocation-No past info`) when p_{12} is low.

`POMDP solution` performs better than the equal allocation schemes even when $p_{12} = 0.5$, and is matched by `Myopic` as well. Thus, an unequal pilot allocation is beneficial when p_{12} approaches 0.5. For $p_{12} = 0.5$, 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. Also, when an equal pilot allocation is used, `Equal allocation-POMDP selection` approaches `Equal allocation-No past info`, and hence the optimal selection policy tends to a greedy policy as p_{12} approaches 0.5. From Figure 7, close to optimal behavior can be achieved for the whole range of p_{12} by the `Myopic` policy.

2) FSMC Model: For the FSMC model, the variation of throughput with f_dT_f is shown in Figure 8. As f_dT_f increases, the channel approaches a block fading one, as the correlation between successive frames decreases. Due to this, greedy AS schemes such as `Equal allocation-No past info` perform well for large f_dT_f.

Thus, the simulation results illustrate that for receive AS, the proposed POMDP approach leads to an improvement in the long-term discounted throughput performance or a reduction of the pilot SNR required to achieve a given performance, relative to existing policies. The benefits arise from the fact that the POMDP is able to fully exploit the information obtained from past actions and observations, as well as the statistical knowledge of the channel fading processes, to make optimal pilot allocation and AS decisions.

VII. 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. Through numerical examples, it was shown that for the channel models considered, the solution to the POMDP outperformed other existing schemes. The POMDP solution is particularly useful when there is no pilot power boosting, 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 $p_{12} = 0.5$, 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, in the 2-state channel with 2 receiver antennas, simple greedy and myopic policies were found to perform nearly optimally, and hence could 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, antenna

selection at both the transmitter and receiver and in multi-user communication systems.

APPENDIX A
OBSERVATION MODELS FOR THE TRAINING PHASE

Here, the observation models used for the FSMC and the 2-state channels are described.

1) FSMC Model: For the FSMC model of the channel, when $\ell_i > 0$ pilots are used on the i^{th} antenna, the received signal is $\mathbf{y}_i = \sqrt{\gamma_i}e^{j\phi_i}\mathbf{p} + \mathbf{w}_i$, where it is assumed that $\mathbf{p} = \sqrt{\frac{E_p}{L}}\,[1 \quad 1 \quad \cdots \quad 1]^T$ and $\mathbf{w}_i \sim \mathcal{CN}(0, \sigma^2\mathbf{I}_{\ell_i})$. Dropping the subscript i for clarity, the ML estimate for $h_i = \sqrt{\gamma_i}e^{j\phi_i}$ is a scaled version of $y' = \sqrt{\gamma}e^{j\phi}\sqrt{\frac{E_p}{L}}\ell + w'$, where $w' \sim \mathcal{CN}\left(0, \ell\sigma^2\right)$, or, equivalently,

$$y = \sqrt{\gamma'}e^{j\phi} + w, \tag{10}$$

with $w \sim \mathcal{CN}(0,2)$, and $\gamma' = \gamma\frac{2\ell E_p}{L\sigma^2}$. By the functional invariance property of the ML, $|y|^2$ yields an ML estimate of γ'. However, the ML estimate is biased, since $\mathbb{E}\left\{|y|^2\right\} - \mathbb{E}\{\gamma'\} = \mathbb{E}\{w^*w\} = 2$. Hence, the quantity $|y|^2 - 2$ is an unbiased estimate for γ', and is used to estimate the received SNR state. In order to build a detector for the received SNR states, the value of $|y|^2 - 2$ is compared against the received SNR thresholds for the model, scaled appropriately. The detected state is d if $\gamma'_d \le |y|^2 - 2 < \gamma'_{d+1}$. The observation at antenna i in frame k is thus defined as

$$\Theta_i[k] \triangleq d \quad \text{if} \quad \gamma'_d \le |y_i[k]|^2 - 2 < \gamma'_{d+1}, \quad d = 1, \ldots, G, \tag{11}$$

where $y_i[k]$ is given by (10). If $\ell_i = 0$ pilots are used on some antenna i, then no observations are obtained on that particular antenna, and the transition probabilities of the channel state are used to update the belief state in Sec. IV.

2) 2-state model: For the 2-state model, $h_i[k] \in \{h_1, h_2\}$, with the values of h_1 and h_2, being known to the receiver. The receiver then has a detection problem in the training phase, and $h_i[k]$ can be written as $h_i[k] = x(h_1 - h_2) + \frac{1}{2}(h_1 + h_2)$, with $x \in \left\{-\frac{1}{2}, \frac{1}{2}\right\}$ being the value to be detected. Specifically, $x = +\frac{1}{2}$ corresponds to h_1, while $x = -\frac{1}{2}$ corresponds to h_2.

Define $S_i[k] \triangleq 1$ if $h_i[k] = h_1$ and $S_i[k] \triangleq 2$ if $h_i[k] = h_2$. Let $\mathbf{v} \triangleq \frac{h_1 - h_2}{|h_1 - h_2|}\frac{\mathbf{p}}{\|\mathbf{p}\|}$, where $\mathbf{p}$ is as in (1), dropping the antenna index i. Then, from (1),

$$\tilde{y} \triangleq \mathbf{v}^H\left[\mathbf{y} - \frac{1}{2}(h_1 + h_2)\mathbf{p}\right] = x\,|h_1 - h_2|\,\|\mathbf{p}\| + w,$$

where $w \sim \mathcal{CN}(0, \sigma^2)$. Since x is real-valued, $\Re\{\tilde{y}\}$ is sufficient [34] 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

$$\Theta_i[k] \triangleq \begin{cases} 2, & \text{if } \lambda_i[k] \ge \eta_i \\ 1, & \text{otherwise,} \end{cases} \tag{12}$$

where

$$\begin{aligned} \lambda_i[k] &\triangleq \ln\frac{\mathbb{P}_{\ell_i}\{\tilde{y}_i[k]|S_i[k] = 2\}}{\mathbb{P}_{\ell_i}\{\tilde{y}_i[k]|S_i[k] = 1\}} \\ &= \sqrt{\frac{\ell_i E_p}{L}}\frac{|h_1 - h_2|\,\Re\{\tilde{y}_i[k]\}}{\sigma^2/2}, \end{aligned}$$

and

$$\eta_i \triangleq \ln\frac{\mathbb{P}\{S_i[k] = 1\}}{\mathbb{P}\{S_i[k] = 2\}} = \ln\frac{1 - p_{22}^{(i)}}{p_{12}^{(i)}}. \tag{13}$$

As before, if $\ell_i = 0$ is used for some i, no observations are obtained on that antenna, and the belief states are updated using the transition probabilities of the Markov chain. In Appendix B 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, for both the FSMC as well as the 2-state channel models.

APPENDIX B
DERIVATION OF OBSERVATION PROBABILITIES FOR TRAINING AND DATA PHASE

It can be seen from (7) and (8) that when the channels at the antennas are mutually independent, it is sufficient to evaluate $\mathbb{P}_{\ell_i}\{\Theta_i[k_0] = \theta_i|s_{0,i}\}$ and $\mathbb{P}_n\{Z[k_1] = z|s_{1,n}\}$ in order to find the observation probabilities $\mathbb{P}_A\{\boldsymbol{\theta}|\mathbf{s}_0\}$ and $\mathbb{P}_A\{z|\mathbf{s}_1\}$ as discussed in Section IV.

1) FSMC Model: For the training phase in the FSMC case, the observation probabilities are

$$\begin{aligned} &\mathbb{P}_A\{\Theta_i = \theta_i|S_i = s_{0,i}\} \\ &\quad = \mathbb{P}\left\{\gamma'_d \le |y|^2 - 2 < \gamma'_{d+1}\middle|\gamma'_a \le \gamma' < \gamma'_{a+1}\right\} \\ &\quad = \mathbb{P}\left\{\tilde{\gamma}_d \le |y|^2 < \tilde{\gamma}_{d+1}\middle|\gamma'_a \le \gamma' < \gamma'_{a+1}\right\}, \end{aligned} \tag{14}$$

where the first equality is written taking $\theta_i = d$ and $s_{0,i} = a$ for notational simplicity, and $\tilde{\gamma} = \gamma' + 2$. Since γ is exponentially distributed with mean $\bar{\gamma}$, γ' is also exponentially distributed with mean $\bar{\gamma'} = \frac{2\ell E_p}{L\sigma^2}\bar{\gamma}$. Thus,

$$\begin{aligned} &\mathbb{P}_A\{\Theta_i = \theta_i|S_i = s_{0,i}\} \\ &= \mathbb{P}\left\{\tilde{\gamma}_d \le \left|\sqrt{\gamma'}e^{j\phi} + w\right|^2 < \tilde{\gamma}_{d+1}\middle|\gamma'_a \le \gamma' < \gamma'_{a+1}\right\} \\ &= \mathbb{P}\left\{\tilde{\gamma}_d \le \left|\sqrt{\gamma'} + w\right|^2 < \tilde{\gamma}_{d+1}\middle|\gamma'_a \le \gamma' < \gamma'_{a+1}\right\} \\ &= \int_{\gamma'_a}^{\gamma'_{a+1}} \frac{\mathbb{P}\left\{\tilde{\gamma}_d \le \left|\sqrt{\gamma'} + w\right|^2 < \tilde{\gamma}_{d+1}\middle|\gamma'\right\} f_{\Gamma'}(\gamma')}{e^{-\frac{\gamma'_a}{\bar{\gamma'}}} - e^{-\frac{\gamma'_{a+1}}{\bar{\gamma'}}}}\mathrm{d}\gamma' \end{aligned} \tag{15}$$

where the second equality follows since $|y| = \left|\sqrt{\gamma'}e^{j\phi} + w\right| = \left|\sqrt{\gamma'} + we^{-j\phi}\right|$, from (10) and the statistical equivalence of w and $we^{-j\phi}$. Now, $\left|\sqrt{\gamma'} + w\right|^2$ has a normalized non-central chi-squared distribution with 2 degrees of freedom (d.o.f.) and non-centrality parameter γ',

and hence

$$\mathbb{P}_A\left\{\Theta_i=\theta_i|S_i=s_{0,i}\right\}$$
$$=\frac{1}{\left(e^{-\frac{\gamma'_a}{\bar{\gamma}'}}-e^{-\frac{\gamma'_{a+1}}{\bar{\gamma}'}}\right)}\Bigg[\int_{\gamma'_a}^{\gamma'_{a+1}}Q_1\left(\sqrt{\gamma'},\sqrt{\tilde{\gamma}_d}\right)\frac{e^{-\frac{\gamma'}{\bar{\gamma}'}}}{\bar{\gamma}'}\mathrm{d}\gamma'$$
$$-\int_{\gamma'_a}^{\gamma'_{a+1}}Q_1\left(\sqrt{\gamma'},\sqrt{\tilde{\gamma}_{d+1}}\right)\frac{e^{-\frac{\gamma'}{\bar{\gamma}'}}}{\bar{\gamma}'}\mathrm{d}\gamma'\Bigg], \quad (16a)$$

where $Q_1\left(s,\sqrt{y}\right)$ is the first order Marcum-Q function [35, Eq. (4.10)] that gives $P\{Y\geq\sqrt{y}\}$ for a normalized non-central chi-square distributed RV Y with 2 d.o.f. and non-centrality parameter s^2.

Performing a change of variable with $x^2=\gamma'$, the first integral in (16a) can be written as $T\left(\tilde{\gamma}_d\right)$, and the second as $T\left(\tilde{\gamma}_{d+1}\right)$, where

$$T(\zeta)=\int_{\sqrt{\gamma'_a}}^{\sqrt{\gamma'_{a+1}}}2xQ_1\left(x,\sqrt{\zeta}\right)\frac{e^{-\frac{x^2}{\bar{\gamma}'}}}{\bar{\gamma}'}\mathrm{d}x, \quad (16b)$$

which using the available result [36, Eq. (B.18)], becomes

$$T(\zeta)=e^{-\frac{\gamma'_a}{\bar{\gamma}'}}Q_1\left(\sqrt{\gamma'_a},\sqrt{\zeta}\right)-e^{-\frac{\gamma'_{a+1}}{\bar{\gamma}'}}Q_1\left(\sqrt{\gamma'_{a+1}},\sqrt{\zeta}\right)$$
$$+e^{-\frac{\zeta}{\bar{\gamma}'\left(1+\frac{2}{\bar{\gamma}'}\right)}}\Bigg[Q_1\left(\sqrt{\gamma'_{a+1}\left(1+\frac{2}{\bar{\gamma}'}\right)},\sqrt{\frac{\zeta}{1+\frac{2}{\bar{\gamma}'}}}\right)$$
$$-Q_1\left(\sqrt{\gamma'_a\left(1+\frac{2}{\bar{\gamma}'}\right)},\sqrt{\frac{\zeta}{1+\frac{2}{\bar{\gamma}'}}}\right)\Bigg] \quad (16c)$$

Hence, using (16a) and (16c), with $\theta_i=d$,

$$\mathbb{P}_A\left\{\Theta_i=d\middle|S_i=s_{0,i}=a\right\}=\frac{T\left(\tilde{\gamma}_d\right)-T\left(\tilde{\gamma}_{d+1}\right)}{\left(e^{-\frac{\gamma'_a}{\bar{\gamma}'}}-e^{-\frac{\gamma'_{a+1}}{\bar{\gamma}'}}\right)}. \quad (17)$$

For the data reception phase,

$$\mathbb{P}_A\left\{Z=1|S_n=s_{1,n}\right\}=1-\bar{P}_{e,j}, \quad (18)$$

where $\bar{P}_{e,j}=\mathbb{P}\left\{\text{Packet Error}|S_n=s_{1,n}=j\right\}$. Assuming that the channel fading is frequency non-selective and slow compared to the frame duration, the average SEP for each state for a particular modulation scheme, $P_{e,j}$, can be obtained as $P_{e,j}=\frac{1}{\bar{p}_j}\int_{\gamma_j}^{\gamma_{j+1}}P_e(\gamma)f_\Gamma(\gamma)\mathrm{d}\gamma$, where γ denotes the received SNR for the data, and $P_e(\gamma)$ is the SEP for a particular value of γ. Assuming an uncoded system with B data symbols in a packet,

$$\bar{P}_{e,j}=1-\left(1-P_{e,j}\right)^B. \quad (19)$$

With BPSK signaling, $P_e(\gamma)=Q(\sqrt{2\gamma})$, and thus

$$P_{e,j}=\frac{1}{\bar{p}_j}\int_{\gamma_j}^{\gamma_{j+1}}Q\left(\sqrt{2\gamma}\right)\frac{1}{\bar{\gamma}}e^{-\frac{\gamma}{\bar{\gamma}}}\mathrm{d}\gamma=\frac{I\left(\gamma_j\right)-I\left(\gamma_{j+1}\right)}{\bar{p}_j}, \quad (20)$$

where,

$$I(\zeta)\triangleq\int_\zeta^\infty Q\left(\sqrt{2\gamma}\right)\frac{1}{\bar{\gamma}}e^{-\frac{\gamma}{\bar{\gamma}}}\mathrm{d}\gamma$$
$$=\left[e^{-\frac{\zeta}{\bar{\gamma}}}-\left(1+\frac{1}{\bar{\gamma}}\right)^{-\frac{1}{2}}\right]Q\left(\sqrt{2\zeta}\right) \quad (21)$$

$\mathbb{P}_A\left\{Z=1|S_n=s_{1,n}\right\}$ can thus be found using (18), (19) and (20).

2) 2-state Model: In the training phase, when the complex channel gain in the good (bad) state is h_2 (h_1) respectively, and a likelihood ratio-based detector of the channel state is used as described in Section A-2, it can be shown that[2]

$$\mathbb{P}_A\left\{\Theta_i=2|s_{0,i}\right\}=Q\left(\kappa_i\left(\frac{\eta_i}{\kappa_i^2}-x_i\right)\right) \quad (22)$$

where $Q(\cdot)$ is the Gaussian Q-function, $\kappa_i=|h_0-h_1|\sqrt{\frac{2\ell_i E_p}{L\sigma^2}}$, η_i is given by (13), and $x_i=-\frac{1}{2}$ for $s_{0,i}=1$ and $x_i=+\frac{1}{2}$ for $s_{0,i}=2$. For the data reception phase, the observation $Z=1$ only if $S_n=2$. Thus, $\mathbb{P}_A\left\{Z=1|S_n=s_{1,n}\right\}=\mathbb{1}_{\{s_{1,n}=2\}}$.

REFERENCES

[1] A. Gorokhov, D. Gore, and A. Paulraj, "Receive antenna selection for MIMO flat-fading channels: theory and algorithms," *IEEE Trans. Inf. Theory*, pp. 2687 – 2696, Oct. 2003.

[2] A. Molisch and M. Win, "MIMO systems with antenna selection," *IEEE Microw. Mag.*, pp. 46 – 56, March 2004.

[3] S. Sanayei and A. Nosratinia, "Antenna selection in MIMO systems," *IEEE Commun. Mag.*, pp. 68 – 73, Oct. 2004.

[4] B. H. Wang, H. T. Hui, and M. S. Leong, "Global and fast receiver antenna selection for MIMO systems," *IEEE Trans. Commun.*, pp. 2505 –2510, Sept. 2010.

[5] H. Saleh, A. Molisch, T. Zemen, S. Blostein, and N. Mehta, "Receive antenna selection for time-varying channels using discrete prolate spheroidal sequences," *IEEE Trans. Wireless Commun.*, pp. 2616–2627, July 2012.

[6] 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.*, pp. 2023 –2034, July 2010.

[7] T. Gucluoglu and E. Panayirci, "Performance of transmit and receive antenna selection in the presence of channel estimation errors," *IEEE Commun. Lett.*, pp. 371 –373, May 2008.

[8] V. Kristem, N. Mehta, and A. Molisch, "Training and voids in receive antenna subset selection in time-varying channels," *IEEE Trans. Wireless Commun.*, pp. 1992 –2003, June 2011.

[9] T. Ramya and S. Bhashyam, "Using delayed feedback for antenna selection in MIMO systems," *IEEE Trans. Wireless Commun.*, pp. 6059 –6067, Dec. 2009.

[10] J. Vicario, M.-A. Lagunas, and C. Anton-Haro, "A cross-layer approach to transmit antenna selection," *IEEE Trans. Wireless Commun.*, pp. 1993–1997, Aug. 2006.

[11] D. Gore and A. Paulraj, "Statistical MIMO antenna sub-set selection with space-time coding," in *Proc. ICC*, April/May 2002, pp. 641 –645.

[12] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artificial Intelligence*, pp. 99 – 134, May 1998.

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

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

[15] S. Ahmad, M. Liu, T. Javidi, Q. Zhao, and B. Krishnamachari, "Optimality of myopic sensing in multichannel opportunistic access," *IEEE Trans. Inf. Theory*, pp. 4040–4050, Sept. 2009.

[2] frame/sub-frame indices are dropped for convenience.

[16] 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*, June 2008.

[17] V. Kristem, N. Mehta, and A. Molisch, "A novel, balanced, and energy-efficient training method for receive antenna selection," *IEEE Trans. Wireless Commun.*, pp. 2742 –2753, Sept. 2010.

[18] 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.*, Nov. 1995, pp. 211 –215.

[19] 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.*, pp. 644 –655, May 1998.

[20] Q. Zhang and S. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, pp. 1688 –1692, Nov. 1999.

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

[22] E. Gilbert *et al.*, "Capacity of a burst-noise channel," *Bell Syst. Tech. J.*, pp. 1253–1265, Sept. 1960.

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

[24] M. Littman, "The witness algorithm: Solving partially observable Markov decision processes," *Brown University, Providence, RI*, 1994.

[25] A. Cassandra, M. Littman, and N. Zhang, "Incremental pruning: A simple, fast, exact method for partially observable Markov decision processes," in *Proc. Conf. Uncertainty in Artificial Intelligence*, Aug. 1997, pp. 54–61.

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

[27] T. Smith and R. Simmons, "Point-based POMDP algorithms: Improved analysis and implementation," in *Proc. Conf. Uncertainty in Artificial Intelligence*, July 2005, pp. 542–549.

[28] J. Hoey and P. Poupart, "Solving POMDPs with continuous or large discrete observation spaces," in *Proc. Int. Joint Conf. Artificial Intelligence*, July-Aug. 2005, pp. 1332–1338.

[29] R. Sutton and A. Barto, *Reinforcement learning: An introduction*. Cambridge University Press, 1998.

[30] Y. Chen, Q. Zhao, and A. Swami, "Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors," *IEEE Trans. Inf. Theory*, pp. 2053 –2071, May 2008.

[31] [Online]. Available: http://bigbird.comp.nus.edu.sg/pmwiki/farm/motion/index.php?n=Site.PomdpPlanning

[32] R. Simmons and S. Koenig, "Probabilistic robot navigation in partially observable environments," in *Proc. Int. Joint Conf. Artificial Intelligence*, Aug. 1995, pp. 1080–1087.

[33] A. Sanei, A. Ghrayeb, and Y. Shayan, "Antenna selection for space-time trellis codes over block Rayleigh fading channels," in *Proc. VTC*, Sept. 2006, pp. 1 –5.

[34] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge University Press, 2005.

[35] M. Simon and M. Alouini, *Digital communication over fading channels*. Wiley-IEEE Press, 2005.

[36] M. Simon, *Probability distributions involving Gaussian random variables: A handbook for engineers and scientists*. Springer Verlag, 2006.

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 Electrical Communication Engineering, Indian Institute of Science, Bangalore, India, in 2012.

From Aug. 2012, he is working as a Research Engineer in the Center for Development of Telematics (C-DoT), Bangalore, India. His research interests are in the areas of MIMO wireless systems, machine learning, and wireless sensor networks.

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

From 2000 to 2002, he worked as an engineer for Qualcomm Inc., San Jose, CA, 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., Bangalore, India, on advanced receiver architectures for the 802.16e Mobile WiMAX standard. In Sept. 2007, he joined as an assistant professor at the Department of Electrical Communication Engineering at the Indian Institute of Science, Bangalore, India, where he is currently working. His research interests are in the areas of Cognitive Radio, Energy Harvesting Wireless Sensors and MIMO systems with channel-state feedback. He is currently serving as an associate editor for the IEEE Signal Processing Letters.

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.