Exploiting Channel Memory for Joint Estimation and Scheduling in Downlink Networks—a Whittle’s Indexability Analysis

Wenzhuo Ouyang, Sugumar Murugesan, Atilla Eryilmaz, and Ness B. Shroff

Abstract—We study opportunistic multiuser scheduling in downlink networks with Markov-modeled outage channels. We consider the scenario that the scheduler does not have full knowledge of the channel state information, but instead estimates the channel state by exploiting the memory inherent in the Markov channels along with Automatic-Repeat-reQuest-style feedback from the scheduled users. Opportunistic scheduling is optimized in two stages: 1) channel estimation and rate adaptation are performed to maximize the short-term throughput, i.e., the successful transmission rate of the scheduled user in the current slot and 2) user scheduling is performed, based on the short-term throughput, to maximize the overall long-term sum-throughput of the downlink. The scheduling problem is a partially observable Markov decision process with the classic exploitation versus exploration tradeoff that is difficult to quantify. We, therefore, study the problem in the framework of restless multiarmed bandit processes, and perform Whittle’s indexability analysis. Whittle’s indexability is traditionally known to be hard to establish and the index policy derived based on Whittle’s indexability is known to have optimality properties in various settings. We show that the problem of downlink scheduling under imperfect channel state information is Whittle indexable and derive the Whittle’s index policy in closed form. Through extensive numerical experiments, we show that the Whittle’s index policy has near-optimal performance and is robust against various imperfections in channel state feedback. Our work reveals that, under incomplete channel state information, exploiting channel memory for opportunistic scheduling can result in significant system-level performance gains and that almost all of these gains can be realized using the polynomial-complexity Whittle’s index policy.

Index Terms—Opportunistic scheduling, incomplete CSI, Markov channel model, channel estimation, Whittle’s index policy.

I. INTRODUCTION

The wireless channel is inherently time-varying and stochastic. It can be exploited for dynamically allocating resources to the network users, leading to the classic opportunistic scheduling principle (see [1]). Understandably, the success of opportunistic scheduling depends heavily on reliable knowledge of the instantaneous channel state information (CSI) at the scheduler. Many sophisticated scheduling strategies have been developed with provably optimal characteristics (see [2]–[6]) by assuming perfect CSI to be readily available, free of cost at the scheduler.

In realistic scenarios, however, perfect CSI is rarely, if ever, available and never cost-free, i.e., a non-trivial amount of network resource, that could otherwise be used for data transmission, must be spent in estimating the CSI [2]. This calls for jointly designing channel estimation and opportunistic scheduling strategies – an area that has recently received attention when the channel state is modeled by i.i.d. processes across time (see [7], [8]). The i.i.d. model has traditionally been a popular choice for researchers to abstract the fading channels, because of its simplicity and associated ease of analysis. On the other hand, this model fails to capture an important characteristics of the fading channels – the time-correlation or the channel memory [2].

In the presence of estimation cost, memory in the fading channels is an important resource that can be intelligently exploited for more efficient, joint estimation and scheduling strategies. In this context, Markov channel models have been gaining popularity as realistic abstractions of fading channels with memory (see [9]–[13]).

In this paper, we study joint channel estimation and scheduling using channel memory, in downlink networks. We model the downlink fading channels as two-state Markov Chains with non-zero achievable rate in both states. Assuming end-of-slot accurate feedback from the scheduled user, the scheduling decision at any time instant is associated with two potentially contradicting objectives: (1) Immediate gains in throughput via data transmission to the scheduled user; (2) Exploration of the channel of a downlink user for more informed decisions and associated throughput gains in the future. This is the classic ‘exploitation vs exploration’ trade-off often seen in sequential decision
making problems (see [14], [15]). We model the joint estimation and scheduling problem as a Partially Observable Markov Decision Process (POMDP) and study the structure of the problem, by explicitly accounting for the estimation cost. Specifically, our contributions are as follows.

- We recast the POMDP scheduling problem as a Restless Multi-armed Bandit Process (RMBP) [16] and establish its Whittle’s indexability [16] in Section IV and V. Even though Whittle’s indexability is difficult to establish in general [17], we have been able to show it in the context of our problem.

- Based on the Whittle’s indexability condition, we explicitly characterize the Whittle’s index policy for the scheduling problem in Section VI. Whittle’s index policies are known to have optimality properties in various RMBP processes and have been shown to have low-complexity (see [17], [18]).

- Using extensive numerical experiments, we demonstrate in Section VII that Whittle’s index policy in our setting has near-optimal performance and that significant system-level gains can be realized by exploiting the channel memory for estimation and scheduling. Numerical experiments also suggest that Whittle’s index policy is robust against imperfections in channel state feedback such as delays and errors. Also, the Whittle’s index policy we derive is of polynomial complexity in the number of downlink users (contrast this with the PSPACE-hard complexity of optimal POMDP solutions [19]).

An important characteristic of our policy is to exploit the time correlation for joint scheduling and channel estimation. The two-state Markov channel model with known Markov transition statistics, while being simple, is a popular model for studying time-correlated fading channels. Among these investigations, one line of works considers the symmetric case where each downlink channel has known and identical Markov transition statistics. In [11], the authors proved the optimality of greedy scheduling algorithm for the symmetric system. The optimality of greedy algorithm is proved in [10] where the channel state feedback is subject to random delays. The works [12], [13] study the performances of round-robin-based scheduling algorithms for queue stability and network utility maximization. Another line of work considers the asymmetric case where the Markov transition statistics differs across channels. In [20], Whittle’s indexability analysis is performed for ON-OFF Markov channels, and a Whittle’s index policy is shown to have near-optimal throughput performance. The authors in [21] proposed a state-action frequency approach that identifies the optimal policy by solving sequence of LPs. The authors in [22] studied the asymptotic optimality of Whittle’s index policy under two-types of Markov transition statistics. The authors in [23] proposed a low-complexity queue-weighted index policy for throughput optimality in downlink queueing networks.

Our setup significantly differs from related works (see [10]–[13], [20]–[23]) in the following sense: In these works, the channels are modeled by ON-OFF Markov Chains with the OFF state corresponding to zero-achievable rate of transmission. There, once a user is scheduled, there is no need to estimate the channel of that user, since it is optimal to transmit at the constant rate allowed by the ON state irrespective of the underlying state. In contrast, in our model, the achievable rate at the lower state is, in general, non-zero and any rate above this achievable rate leads to outage. This extended model captures the realistic scenario when non-zero rates are possible with the use of sophisticated physical layer algorithms, even when the channel is bad. In this model, once a user is scheduled, the scheduler must estimate the channel of that user, with an associated cost, and adapt the transmission rate based on the estimate. The rate adaptation must balance between aggressive transmissions that lead to outage and conservative transmissions that lead to under-utilization of channels. The achievable rate expected from this process, in turn, influences the choice of the scheduled user. Thus the channel estimation and scheduling stages are tightly coupled, introducing several technical challenges to the problem, which we address in this paper.

II. SYSTEM MODEL AND PROBLEM STATEMENT

A. Channel Model

We consider a downlink system with one base station (BS) and $N$ users. Time is slotted with the time slots of all users synchronized. The channel between the BS and each user is modeled as a two-state Markov chain, i.e., the state of the channels remains static within each time slot and evolves across time slots according to Markov chain statistics. The Markov channels are assumed to be independent and, in general, non-identical across users. The state space of channel $C_i$ between the BS and user $i$ is given by $S_i = \{l_i, h_i\}$. Each state corresponds to a maximum allowable data rate. Specifically, if the channel is in state $l_i$, there exists a rate $\delta_i$, $0 \leq \delta_i < 1$, such that data transmissions at rates below $\delta_i$ succeed and transmissions at rates above $\delta_i$ fail, i.e., outage occurs. Similarly, state $h_i$ corresponds to data rate 1. Note that fixing the higher rate to be 1 across all users does not impose any loss of generality in our analysis. This will be evident as we proceed.

The Markovian channel model is illustrated in Fig. 1. For user $i$, the two-state Markov channel is characterized by a $2 \times 2$ probability transition matrix

$$P_i = \begin{bmatrix} p_i & 1 - p_i \\ r_i & 1 - r_i \end{bmatrix},$$

where $p_i := \text{prob}(C_i[t] = h_i \mid C_i[t-1] = h_i)$, $r_i := \text{prob}(C_i[t] = h_i \mid C_i[t-1] = l_i)$, with $C_i[t]$ denoting the channel state in slot $t$. Fig. 1. A two state Markov Chain.
B. Scheduling Model

We adopt the one-hop interference model, i.e., in each time slot, only one user can be scheduled for data transmission. At the beginning of the slot, the scheduler does not have exact knowledge of the channel state of the downlink users. Instead, it maintains a belief value \( \pi_i \) for channel \( i \) which is the probability that \( C_i \) is in state \( h_i \) at the time. We will elaborate on the belief values soon. Using these belief values, the scheduler picks a user, estimates its current channel state and subsequently transmits data at a rate adapted to the channel state estimate – all with an objective to maximize the overall sum-throughput of the downlink system. Specifically, in each slot, the scheduler jointly makes the following decisions: (1) Considering each user, the scheduler decides on the optimal channel estimator (that could involve the expenditure of network resources such as time, power, etc.) and rate adapter pair, and calculates the expected successful transmission rate in the current slot, i.e., the short-term throughput, for that user; (2) Based on the short-term throughput promised for each user by the previous decision, the scheduler picks a user for channel estimation, rate adaptation and subsequent data transmission, while simultaneously taking into account the effect of this decision on the long-term throughput. At the end of the slot, consistent with recent models (see [10]–[13], [20], [21]), the scheduled user sends back accurate feedback on the (high/low) state of the Markov channel in that slot. This accurate feedback is, in turn, used by the scheduler to update its belief on the channels, based on the Markov channel statistics. Note that these belief values are sufficient statistics for the underlying channel states, with respect to any past scheduling decisions and feedback [24] and are critical to our forthcoming analysis. Using \( \epsilon \) to denote the chosen estimator and \( \eta \) to denote the chosen rate adapter, the general operation is summarized in Fig. 2. The scheduling problem can be formulated as a partially observable Markov decision process [24], with the Markov channel states being the partially observable system states.

As noted in Section I, the scheduling decision in each slot involves two objectives: data transmission to the scheduled user and probing the channel of the scheduled user (through the accurate end-of-slot (high/low) channel state feedback). On one hand, the scheduler can transmit data to the user that promises the best short-term throughput and hence realize immediate performance gains. On the other hand, the scheduler can schedule possibly another user and use the channel feedback from that user to gain a better understanding of the overall downlink system, which, in turn, could result in more informed future scheduling decisions with corresponding gains in the future.

C. Formal Problem Statement

We now proceed to formally introduce the expected immediate reward. Recall from Section II-B that, the short-term throughput depends on the choice of both the channel estimator \( \epsilon \) and rate adapter \( \eta \). We hence let \( u := \{\epsilon, \eta\} \) denote an arbitrary estimator and rate adapter pair and \( U \) denote the set of such pairs. Recall from the discussion on the scheduling model that, at the end of the slot, the scheduled user sends back accurate feedback on its (high/low) Markov channel state in that slot. With this setup, once a user is scheduled, the choice of the channel estimator and rate adapter pair does not affect the future paths of the scheduling process. Thus, within each slot, it is optimal to design this pair to maximize the short-term throughput of the user scheduled in that slot. Henceforth, in the language of POMDPs, we call this maximized short-term throughput the expected immediate reward. We now proceed to formally introduce the expected immediate reward. We let \( \pi_i \) denote the current belief value of the channel of user \( i \). The optimal estimator and rate adapter pair, \( u^*_{i,\pi_i} = (\epsilon^*_{i,\pi_i}, \eta^*_{i,\pi_i}) \), for user \( i \), as a function of the belief \( \pi_i \), is given by

\[
    u^*_{i,\pi_i} = \arg \max_{u \in U} E_{C_i}[\gamma_i(C_i, u)],
\]

where the quantity \( \gamma_i(C_i, u) \) is the short-term throughput of user \( i \) when the underlying Markov channel is in state \( C_i \) and the estimator and rate adapter pair \( u \) is deployed. The expectation in (1) is taken over the underlying channel state \( C_i \), with distribution characterized by belief value \( \pi_i \), i.e.,

\[
    C_i = \begin{cases} 
        h_i & \text{with probability } \pi_i, \\
        l_i & \text{with probability } 1 - \pi_i.
    \end{cases}
\]

The expected immediate reward when user \( i \) is scheduled is thus given by

\[
    R_i(\pi_i) = E_{C_i}[\gamma_i(C_i, u^*_{i,\pi_i})].
\]

Note that our model is very general in the sense that we do not restrict to any specific estimation, data transmission structure, corresponding to the estimator and rate adapter pair \( u \) is illustrated in Fig. 3. Here a pilot-aided estimation [2] is performed for a fraction of the time slots.
followed by data transmission at an adapted rate in the rest of the time slots. Correspondingly, $u^*_{\pi}$ represents the optimal allocation of time/power for training and optimal transmission rate that maximize the expected immediate reward.

We now introduce the optimality equations for the scheduling problem. Let $\vec{\pi}[t] = (\pi_1[t], \ldots, \pi_N[t])$ denote the vector of current belief values of the channels at the beginning of slot $t$. A stationary scheduling policy, $\Psi$, is a stationary mapping $\Psi: \vec{\pi} \rightarrow I$ between the belief vector and the index of the user scheduled for data transmission in the current slot. Our performance metric is the infinite horizon, discounted sum-throughput of the downlink (henceforth simply the expected discounted reward in the language of POMDPs), formally defined next.

For a stationary policy $\Psi$, the expected discounted reward under initial belief $\vec{\pi}$ is given by

$$V(\Psi, \vec{\pi}) = \sum_{t=0}^{\infty} \beta^t \mathbb{E}_{\vec{\pi}[t]}[R_I(t) = \Psi(\vec{\pi}[t])((\pi_I[t]|I[t])$$

where $\vec{\pi}[t]$ is the belief vector in slot $t$, $\pi_I[t]$ denotes the belief value of user $i$ in slot $t$, $\vec{\pi}[0] = \vec{\pi}$, $I[t]$ denotes the index of the user scheduled in slot $t$. The discount factor $\beta \in (0,1)$ provides relative weighing between the immediately realizable rates and future rates. For any initial belief $\vec{\pi}$, the optimal expected discounted reward, $V(\vec{\pi}) = \max_{\Psi} V(\Psi, \vec{\pi})$, is given by the Bellman equation [25], [26]

$$V(\vec{\pi}) = \max_{\pi} \{R_I(\pi_I) + \beta \mathbb{E}_{\vec{\pi}[t+1]}[V(\vec{\pi}[t+1])].$$

Here $\vec{\pi}[t+1]$ denotes the belief vector in the next slot when the current belief is $\vec{\pi}$. The belief evolution $\vec{\pi}[t] \rightarrow \vec{\pi}[t+1]$ proceeds as follows:

$$\pi_{i}[t+1] = \begin{cases} p_i & \text{if } I[t] = i \text{ and } C_i[t] = h_i, \\ r_i & \text{if } I[t] = i \text{ and } C_i[t] = l_i, \\ Q_i(\pi_i | I[t]) & \text{if } I[t] \neq i, \end{cases}$$

where $Q_i(\pi_i) = p_i p_i + (1 - p_i) r_i$ is the belief evolution operator for user $i$ when it is not scheduled in the current slot. A stationary scheduling policy $\Psi^*$ is optimal if and only if $V(\Psi^*, \vec{\pi}) = V(\vec{\pi})$ for all $\vec{\pi}$ [25].

In the introduction, we briefly contrasted our setup with those in [10]-[13], [20], and [21]. We provide a rigorous comparison here. The works [10]-[13], [20], [21] studied opportunistic scheduling with the channels modeled by ON-OFF Markov chains. In these works, the lower state is an ‘OFF’ state, i.e., it does not allow transmission at any non-zero data rate. Contrast this with our model where, at the lower state $l_i$, a possibly non-zero rate $\delta_i$ is achievable and outage occurs at any rate above $\delta_i$. We now further explain how these two models are fundamentally different.

In the ON-OFF channel model, the scheduler does not need a channel estimator and rate adapter pair. The scheduler can aggressively transmit at rate 1, since it has nothing to gain by transmitting at a lower rate – a direct consequence of the ‘OFF’ nature of the lower state. On the other hand, transmitting at a rate lesser than 1 can lead to losses due to under-utilization of the channel.

- In contrast, in our model, when $\delta > 0$, the scheduler must strike a balance between aggressive and conservative rates of transmission. An aggressive strategy (transmit at rate 1) can lead to losses due to outages, while a conservative strategy can lead to losses due to under-utilization of the channel. This underscores the importance of the knowledge of the underlying channel state and, therefore, the need for intelligent estimation and rate adaptation mechanisms.

- As a direct consequence of the preceding arguments, the expected immediate reward in our model is not a trivial $\delta$-shift of the expected immediate reward when the rates supported by the channel states are 0 and 1 – $\delta$. Formally, $R(\pi) = R_0(\pi), \delta > 0$.

We believe that our channel model, in contrast to the ON-OFF model, better captures realistic communication channels where, using appropriate physical layer algorithms, it is possible to transmit at a non-zero rate even at the lowest state of the channel model and the same physical layer algorithms may impose outage behavior when this allowed rate is exceeded.

III. OPTIMAL EXPECTED TRANSMISSION RATE – STRUCTURAL PROPERTIES

In this section, we study the structural properties of the expected immediate reward, $R(\pi)$, defined in equation (2). These properties will be crucial for our analysis in subsequent sections. For notational convenience, we will drop the suffix $i$ in the rest of this section.

Lemma 1: The expected immediate reward $R(\pi)$ has the following properties:

(a) $R(\pi)$ is convex and increasing in $\pi$ for $\pi \in [0, 1]$.

(b) $R(\pi)$ is bounded as follows:

$$\max\{\pi, 1 - \pi\} \leq R(\pi) \leq (1 - \delta)\pi + \delta.$$  (4)

Proof: Let $U^*$ be the set of optimal estimator and rate adapter pairs for all $\pi \in [0, 1]$, i.e., $U^* = \{u^*_\pi, \pi \in [0, 1]\}$. The expected immediate reward, provided in (2), can now be rewritten as

$$R(\pi) = \max_{u \in U^*} \mathbb{E}_C[\gamma(C, u)] = \max_{u \in U^*} \mathbb{E}_C[\pi \gamma(h, u) + (1 - \pi) \gamma(l, u)],$$

where $\gamma(s, u)$ denotes the average rate of successful transmission when the channel state is $s \in \{l, h\}$. Note that, for fixed $u$, the average rate $\pi \gamma(h, u) + (1 - \pi) \gamma(l, u)$ is linear in $\pi$. Thus, $R(\pi)$ is given as a point-wise maximum over a family of linear functions, which is convex [27]. $R(\pi)$ is therefore convex in $\pi$, establishing the convexity statement in (a). We proceed to derive the bounds to $R(\pi)$. From equation (2),

$$R(\pi) = \max_{u \in U^*} \mathbb{E}_C[\gamma(C, u)] \geq \max_{\eta} \mathbb{E}_C[\gamma(C, \{\pi, \eta\})]$$
where }\gamma(C, [* , \eta])\text{ denotes expected transmission rate where only rate adaptation is considered by channel estimation. This explains the last inequality. Note that without the estimator, the rate adaptation is solely a function of the belief value }\pi.\text{ Thus, the average rate achieved under the rate adapter, conditioned on the underlying channel state, can be expressed simply by indicator functions, as seen below:}

\[
\max_{\eta} \mathbb{E}C[\gamma(C, [* , \eta])]
= \max_{\eta} [P(C = l)\eta \cdot 1(\eta \leq \delta) + P(C = h)\eta \cdot 1(\eta \leq 1)]
= \max_{\eta}[P(C = l) \cdot 1(\eta \leq \delta) + P(C = h) \cdot 1(\eta \leq 1)]
= \max \{\delta, \pi\}.
\]

This establishes the lower bound in (b).

The upper bound in (b) corresponds to the expected immediate reward when full channel state information is available at the scheduler.

It is clear from the upper and lower bounds that \(\delta \leq R(\pi) \leq 1\). Note that when }\pi=0\text{ or }\pi=1,\text{ there is no uncertainty in the channel, hence }R(0)=\delta\text{ and }R(1)=1.\text{ Using these properties, along with the convexity property of }R(\pi),\text{ we see that }R(\pi)\text{ is monotonically increasing in }\pi,\text{ establishing the monotonicity of (a). The lemma thus follows.}

\textbf{Remark:} Here we present some insights into the effect of the non-zero rate }\delta\text{ on the channel estimation and rate adaptation mechanisms by studying the upper and lower bounds to }R(\pi)\text{ provided in Lemma 1. The upper bound essentially corresponds to the case when perfect channel state information is available at the scheduler at the beginning of each slot. Here, no channel estimation and rate adaptation is necessary. The lower bound, on the other hand, corresponds to the case when the channel estimation stage is eliminated and rate adaptation is performed solely based on the belief value }\pi\text{ of the scheduled user.}

Fig. 4 plots the lower and upper bounds to }\overline{R}(\pi)\text{ for different values of }\delta.\text{ Note that the lower bound approaches the upper bound in both directions, i.e., when }\delta \to 0\text{ or when }\delta \to 1.\text{ This behavior can be explained as follows: (1) }\delta \to 1\text{ essentially means that the states of the Markov channel move closer to each other. This progressively reduces the channel uncertainty and hence the need for channel estimation (and, consequently, rate adaptation), essentially bringing the bounds closer. (2) As }\delta \to 0,\text{ the channel uncertainty increases. At the same time, the impact of the channel estimator and rate adapter pair decreases. This is because, as }\delta \to 0,\text{ the loss in immediate reward due to outage (transmitting at 1 when channel is in state }\delta)\text{ is less severe than the loss due to under-utilization of the channel (transmitting at rate }\delta\text{ when the channel is in state 1), essentially making it optimal for the rate adaptation scheme to be progressively more aggressive (transmit at rate 1). Thus channel estimation loses its significance as }\delta \to 0.\text{ This brings the bounds closer as }\delta \to 0.\text{ It can be verified that the separation between the lower and upper bounds is at its peak when }\delta = 0.5.\text{ This, along with the preceding discussion, indicates the potential for rate improvement when intelligent channel estimation and rate adaptation is performed under moderate values of }\delta.\text{ }

IV. RESTLESS MULTI-ARMED BANDIT PROCESSES, WHITTLE’S INDEXABILITY AND INDEX POLICIES

A direct analysis of the downlink scheduling problem appears difficult due to the complex nature of the ‘exploitation vs exploration’ tradeoff. We therefore establish a connection between the scheduling problem and the Restless Multiarmed Bandit Processes (RMBP) [16] and make use of the established theory behind RMBP in our analysis. We briefly overview RMBPs and the associated theory of Whittle’s indexability in this section.

RMBPs are defined as a family of sequential dynamic resource allocation problems in the presence of several competing, independently evolving projects. In RMBPs, a subset of the competing projects are served in each slot. The states of each project in the system stochastically evolve in time based on the current state of the projects and the action taken (whether the project is served or not in the current slot). Once a project is served, a reward dependent on the states of the served projects and the action taken is accrued by the controller. Hence, the RMBPs are characterized by a fundamental tradeoff between decisions guaranteeing high immediate rewards versus those that sacrifice immediate rewards for better future rewards. Solutions to RMBPs are, in general, known to be PSPACE-hard [19].

Under an average constraint on the number of projects scheduled per slot, a low complexity index policy developed by Whittle [16], commonly known as Whittle’s index policy, is optimal. Under stringent constraint on the number of users scheduled per slot, Whittle’s index policy may not exist and, if it does exist, its optimality properties are, in general, lost. However, Whittle’s index policies, upon existence, are known to have near optimal performance in various RMBPs (see [17], [18]). For an RMBP, Whittle’s index policy exists if and only if the RMBP satisfies a condition known as Whittle’s indexability [16], defined next.

Consider the following Whittle indexability setup: for each project }P\text{ in the system, consider a virtual system where, in each slot, the controller must make one of two decisions: (1) Serve project }P\text{ and accrue an immediate reward that is a function of the state of the project. This reward structure reflects the one in the original RMBP for project }P.\text{ (2) Do not serve project }P,\text{ i.e., stay passive and accrue an immediate...}
reward for passivity \( \omega \). The state of the project \( P \) evolves in the same fashion as it would in the original RMBP, as a function of its current state and current action. In this setup, the design goal is to maximize the infinite horizon reward by balancing (1) immediate reward from serving the project, and (2) subsidy for passivity, i.e., not serving the project, in each slot. We let \( D(\omega) \) be the set of states of project \( P \) in which it is optimal to stay passive. The Whittle’s indexability condition is defined as follows.

**Project P is Whittle indexable if and only if**, as \( \omega \) increases from \(-\infty\) to \( \infty \), the set \( D(\omega) \) monotonically expands from \( \emptyset \) to \( S \), the state space of project \( P \). The RMBP is Whittle indexable if and only if all the projects in the RMBP are Whittle indexable.

For each state, \( s \), of a project, Whittle’s index, \( W(s) \), is given by the infimum of the value of \( \omega \) in which it is optimal to stay idle in the \( \omega \)-sizedized system, i.e.,

\[
W(s) = \inf\{\omega : s \in D(\omega)\}.
\]

The notion of indexability gives a consistent ordering of states with respect to the indices. For instance, if \( W(s_1) > W(s_2) \) and if it is optimal to serve the project at state \( s_1 \), then it is optimal to serve the project at \( s_2 \). This natural ordering of states based on the indices renders the near-optimality properties to Whittle’s index policy (see [17], [18]).

The downlink scheduling problem we have considered is in fact an RMBP process. Here, each downlink user, along with the belief value of its channel, corresponds to a project in the RMBP, and the project is served when the corresponding user is scheduled for data transmission. Now, referring to our earlier discussion on the RMBPs, we see that Whittle’s index policy is very attractive from an optimality point of view. The attractiveness of the index policy can be attributed to the natural ordering of states (and hence projects) based on indices, as guaranteed by Whittle’s indexability. In the rest of the paper, we establish that this advantage carries over to the downlink scheduling problem at hand. As a first step in this direction, in the next section, we study the scheduling problem in Whittle’s indexability framework and show that the downlink scheduling problem is, in fact, Whittle indexable.

**V. WHITTLE’S INDEXABILITY ANALYSIS OF THE DOWNLINK SCHEDULING PROBLEM**

In this section, we study the Whittle’s indexability of our joint scheduling and estimation problem. To that end, we first describe the downlink scheduling setup:

At the beginning of each slot, based on the current belief value \( \pi \) (we drop the user index \( i \) in this section since only one user is considered throughout), the scheduler takes one of two possible actions: schedules data transmission to the user (action \( a = 1 \)) or stays idle (\( a = 0 \)). Upon an idle decision, a subsidy of \( \omega \) is obtained. Otherwise, optimal channel estimation and rate adaptation is carried out, with a reward equal to \( R(\pi) \) (consistent with the immediate reward seen in previous sections). The belief value is updated based on the action taken and feedback from the user (upon transmit decision). This belief update is consistent with that in the Section II. The optimal scheduling policy (henceforth, the \( \omega \)-subsidy policy) maximizes the infinite horizon discounted reward, parameterized by \( \omega \). The optimal infinite horizon discounted reward is given by the Bellman equation [25]

\[
V_\omega(\pi) = \max\{R(\pi) + \beta(\pi V_\omega(p) + (1 - \pi)V_\omega(r))\},
\]

\[
[\omega + \beta V_\omega(Q(\pi))],
\]

where, recall from Section II, \( Q(\pi) \) is the evolution of the belief value when the user is not scheduled. The first quantity inside the max operator corresponds to the infinite horizon reward when a transmit decision is made in the current slot and optimal decisions are made in the future slot. The second element corresponds to idle decision in the current slot and optimal decisions in all future slots.

We note that the indexability analysis in the rest of this section bears similarities to that in [20], where the authors studied indexability of a sequential resource allocation problem in a cognitive radio setting. This problem is mathematically equivalent to our downlink scheduling problem when \( \delta = 0 \). We have already discussed in detail (in Section II) that the structure of the immediate reward \( R(\pi) \) when \( \delta > 0 \) is very different than when \( \delta = 0 \), due to the need for channel estimation and rate adaptation in the former case. Consequently, in the Whittle’s indexability setup, the infinite horizon discounted reward \( V_\omega(\pi) \) in our problem is different (and more general) than that in [20], underscoring the significance of our results.

As a crucial preparatory result, we now proceed to show that the \( \omega \)-subsidy policy is thresholdable.

**A. Thresholdability of the \( \omega \)-Subsidy Policy**

We first record our result on the convexity property of the infinite horizon discounted reward, \( V_\omega(\pi) \), of (5) in the following proposition.

**Proposition 1**: The infinite horizon discounted reward, \( V_\omega(\pi) \), is convex in \( \pi \in [0, 1] \).

**Proof**: We first consider the discounted reward for finite horizon \( \omega \)-subsidy problem. We let \( v^1(\pi) = R(\pi) \) and \( v^0(\pi) = \omega \) represent the immediate reward corresponding to active and idle decisions, respectively. The reward function associated with \( M \)-stage finite horizon process is expressed as

\[
\hat{V}_M(\pi[0]) = \max_{a[1]} \mathbb{E}\left[\sum_{t=0}^{M-1} \beta^t v^a(t)[\pi[t]] \mid \pi[0]\right]
\]

Let \( \hat{V}_{\omega,t}(\pi) \) be the reward at time \( t \) with belief value \( \pi[t] = \pi \). Hence \( \hat{V}_M(\pi[0]) = \hat{V}_{\omega,0}(\pi[0]) \) and the last stage value function \( \hat{V}_{\omega,M-1}(\pi[M-1]) \) is given by

\[
\hat{V}_{\omega,M-1}(\pi[M-1]) = \max_{a[M-1]} \{v^a[M-1]\pi[M-1]\} = \max(\omega R(\pi[M-1])).
\]

Therefore, \( \hat{V}_{\omega,M-1}(\pi) \) is convex with \( \pi \) since it is the maximum of a constant and a convex function. For any time \( 0 \leq t < M-1 \), the Bellman [25] equation can be written as

\[
\hat{V}_{\omega,t}(\pi[t]) = \max\{\hat{V}_{\omega,t}^0(\pi[t]), \hat{V}_{\omega,t}^1(\pi[t])\}.
\]
where
\[ \hat{V}_{0,t+1}^0(\pi) = \omega + \beta \hat{V}_{0,t+1}(Q(\pi)), \]
\[ \hat{V}_{0,t+1}^1(\pi) = R(\pi) + \beta \left( \pi \hat{V}_{0,t+1}(p) + (1 - \pi) \hat{V}_{0,t+1}(r) \right). \]

Suppose now \( \hat{V}_{0,t+1}^0(\pi) \) is convex with \( \pi \). If \( a[t] = 1 \), it is clear from (7) that \( \hat{V}_{0,t+1}^1(\pi) \) is convex function of \( \pi \) since it is a summation of a convex function and a linear function of \( \pi \). If \( a[t] = 0 \), \( \hat{V}_{0,t+1}^0(\pi) \), expressed in (6), is also a convex function, because composition of convex function \( \hat{V}_{0,t+1}^0(\pi) \) and linear function \( Q(\pi) \) is convex [27]. Therefore \( \hat{V}_{0,t}(\pi) \) is convex with \( \pi \) as maximum of two convex functions. By induction, the convexity of \( \hat{V}_{0,0}(\pi) \) is thus established.

Since \( \hat{V}_M(\pi) = V_{0,0}(\pi), \hat{V}_M(\pi) \) is convex with \( \pi \). For discounted problem with bounded reward per slot, the infinite horizon reward is the limit of \( V_{0,0}(\pi) \) [25]. Therefore \( V_0(\pi) = \lim_{M \to \infty} V_{0,M}(\pi) \). Upon point-wise convergence, point-wise limit of convex functions is convex [27]. Hence \( V_0(\pi) \) is a convex function of \( \pi \).

In the next proposition, we show that the optimal \( \omega \)-subsidy policy is a threshold policy.

**Proposition 2:** The optimal \( \omega \)-subsidy policy is thresholdable in the belief space \( \pi \). Specifically, there exists a threshold \( \pi^*(\omega) \) such that the optimal action \( a \) is 1 if the current belief \( \pi > \pi^*(\omega) \) and the optimal action \( a \) is 0, otherwise. The value of the threshold \( \pi^*(\omega) \) depends on the subsidy \( \omega \), partially characterized below.

(i) If \( \omega > 1 \), \( \pi^*(\omega) = 1 \);
(ii) If \( \omega \leq \delta \), \( \pi^*(\omega) = \kappa \) for some arbitrary \( \kappa < 0 \);
(iii) If \( \delta < \omega < 1 \), \( \pi^*(\omega) \) takes value within interval \((0, 1)\).

**Proof:** Consider the Bellman equation (5), let \( V_{0}^1(\pi) \) be the reward corresponding to transmit decision and \( V_{0}^0(\pi) \) be the reward corresponding to idle decision, i.e.,
\[ V_{0}^1(\pi) = R(\pi) + \beta (\pi V_{0}(p) + (1 - \pi) V_{0}(r)), \]
\[ V_{0}^0(\pi) = \omega + \beta V_{0}(Q(\pi)) = \omega + \beta V_{0}(p + (1 - \pi)r). \]

It is clear from the Bellman equation (5) that the optimal action depends on the relationship between \( V_{0}^1(\pi) \) and \( V_{0}^0(\pi) \), presented as follows.

**Case (i):** If \( \omega \geq 1 \), since \( R(\pi) \leq 1 \), in each slot, the immediate reward for being idle gives the maximum reward for being active. Hence it will be optimal to always stay idle. We can thus set the threshold to 1.

**Case (ii):** If \( \omega \leq \delta \), then for any \( \pi \in [0, 1] \), we have
\[ V_{0}^0(\pi) = \omega + \beta V_{0}(\pi p + (1 - \pi)r) \leq R(\pi) + \beta (\pi V_{0}(p) + (1 - \pi) V_{0}(r)), \]
\[ = V_{0}^1(\pi), \]
where the inequality is due to \( \delta \leq R(\pi) \) along with Jensen's inequality [27] due to the convexity of \( V_{0}(\pi) \) from Proposition 2. Hence, it is optimal to stay active. Consistent with the threshold definition, we can set \( \pi^*(\omega) = \kappa \) for any \( \kappa < 0 \).

**Case (iii):** If \( \delta < \omega < 1 \), then at the extreme values of belief,
\[ V_{0}^0(0) = \omega + \beta V_{0}(r) > \delta + \beta V_{0}(r) = V_{0}^0(0) \]
\[ V_{0}^0(1) = \omega + \beta V_{0}(p) < 1 + \beta V_{0}(p) = V_{0}^0(1) \]

Note that the relationship of \( V_{0}^0(\pi) \) and \( V_{0}^1(\pi) \) is reversed at the end points 0 and 1, and they are both convex functions of \( \pi \). Thus, there must exist a threshold \( \pi^*(\omega) \) within \((0, 1)\) such that \( a = 1 \) whenever \( \pi > \pi^*(\omega) \).

**B. Whittle’s Indexability of Downlink Scheduling**

Having established that the \( \omega \)-subsidy policy is thresholdable in Proposition 2, Whittle’s indexability, defined in Section IV, is re-interpreted for the downlink scheduling problem as follows: the downlink scheduling problem is Whittle indexable if the threshold boundary \( \pi^*(\omega) \) monotonically increases with subsidy \( \omega \).

Using our discussion in Section IV, the index of the belief value \( \pi \), i.e., \( W(\pi) \) is the infimum value of the subsidy \( \omega \) such that it is optimal to stay idle, i.e.,
\[ W(\pi) = \inf \{ \omega : V_{0}^0(\pi) \geq V_{0}^1(\pi) \} = \inf \{ \omega : \pi^*(\omega) = \pi \}. \]

To establish indexability, we need to investigate the infinite horizon discounted reward \( V_{0}(\pi) \), given by (5). We can observe from (5) that given the value of \( V_{0}(p) \) and \( V_{0}(r) \), \( V_{0}(\pi) \) can be calculated for all \( \pi \in [0, 1] \). Let \( \pi^0 \) denote the steady state probability of being in state \( h \). The next lemma provides closed form expressions for \( V_{0}(p) \) and \( V_{0}(r) \) and is critical to the proof of indexability.

**Lemma 2:** The closed form expression of the discounted rewards \( V_{0}(p) \) and \( V_{0}(r) \) is expressed as follows.

**Case 1:** \( p > r \) (positive correlation)
\[ V_{0}(p) = \begin{cases} \sum_{k=0}^{\infty} \beta R \frac{R + (p - r)^k (1 - p)}{1 - (1 - \beta) (1 - \beta^k)} & \text{if } \pi^0(\omega) < p^0 \\ \frac{1 - \beta}{\omega} & \text{if } \pi^0(\omega) \geq p^0 \end{cases} \]

**Case 2:** \( p \leq r \) (negative correlation)
\[ V_{0}(r) = \begin{cases} \sum_{k=0}^{\infty} \beta R \frac{R + (p - r)^k (1 - p)}{1 - (1 - \beta) (1 - \beta^k)} & \text{if } \pi^0(\omega) < p^0 \\ \Theta & \text{if } \pi^0(\omega) \geq p^0 \end{cases} \]
The expression of $\Theta$ is given by (11), as shown at the top of the page, where $Q^n$ denotes $n^{th}$ iteration of $Q$ and $L(\pi, \pi^*(\omega))$ is a function of $\pi$ and $\pi^*(\omega)$. Their expressions are given in Appendix A.

Proof: The derivation of $V_\omega(p)$ and $V_\omega(r)$ follows from substituting $p$ and $r$ in (5). Together with the expression of $Q(\pi)$ given by in Section II, the expression of $V_\omega(p)$ and $V_\omega(r)$ can be obtained. For details, please refer to Appendix A.

We note that the value function expression depends on the correlation type of the Markov chain, because the transition function $Q(\pi)$ given in Section II behaves differently with the correlation type of the channel.

The closed form expression of the value function given by the previous lemma serves as a useful tool for us to establish indexability, which is given by the next proposition.

**Proposition 3:** The threshold value is strictly increasing with $\omega$. Therefore, the problem is Whittle indexable.

Proof: The proof of indexability follows the lines of [20]. Details are provided in Appendix B.

VI. WHITTLE’S INDEX POLICY

In this section, we explicitly characterize Whittle’s index policy for the downlink scheduling problem. For user $i$, recall that $\pi_i^0$ denotes the steady state probability of being in state $h_i$, and we let $V_{i,\omega}(\pi_i)$ denote the reward function for its $\omega$-subsidy problem in (5). We first characterize the Whittle’s index as follows.

**Proposition 4:** For user $i$, the index value at state $\pi_i$, i.e., $W_i(\pi_i)$ is characterized as follows,

- **Case 1:** Positively correlated channel ($p_i > r_i$)

  $$W_i(\pi_i) = \frac{R_i(\pi_i)}{1 - \beta R_i(\pi_i) + (1 - \beta p_i) R_i(\pi_i)}$$

  if $\pi_i^0 \leq \pi_i < p_i$

  $$W_i(\pi_i) = \frac{R_i(\pi_i)}{1 - \beta R_i(\pi_i) - \beta p_i}$$

  if $\pi_i^0 \geq \pi_i < p_i$

- **Case 2:** Negatively correlated channel ($p_i \leq r_i$)

  $$W_i(\pi_i) = \frac{R_i(\pi_i)}{1 - \beta R_i(\pi_i) + (1 - \pi_i) V_{i, W_i(\pi_i)}(r_i)}$$

  if $Q_i(p_i) \leq \pi_i < r_i$

  $$W_i(\pi_i) = \frac{R_i(\pi_i) + \beta[(1 - \pi_i) - (1 - Q_i(\pi_i))]}{1 - \beta R_i(\pi_i) - \beta p_i}$$

  if $\pi_i^0 \leq \pi_i < Q_i(p_i)$

  $$W_i(\pi_i) = \frac{R_i(\pi_i) + \beta[(1 - \pi_i) - \beta(1 - Q_i(\pi_i))] + \beta[1 - \beta(1 - Q_i(\pi_i))]}{1 - \beta R_i(\pi_i) - \beta p_i}$$

  if $\pi_i^0 \geq \pi_i < Q_i(p_i)$

Fig. 5. Index value evolution of user $i$, with $\pi_i[0] = 0.3$. (a) Positive correlation, $p_i = 0.8$, $r_i = 0.2$; (b) Negative correlation, $p_i = 0.2$, $r_i = 0.8$.

Proof: The derivation of the index value follows from substituting the expression of $V_{i,\omega}(p_i)$ and $V_{i,\omega}(r_i)$ (given in Lemma 2) in (5). Details of the proof are provided in Appendix C.

Remark: Notice that Proposition 4 does not give the closed form expression for $W_i(\pi_i)$. However, since the closed form expression of the value function $V_{i, W_i(\pi_i)}(p_i)$ and $V_{i, W_i(\pi_i)}(r_i)$ are derived in Lemma 2, closed form expressions of $W_i(\pi_i)$ can be easily calculated and is given in Appendix C. We now introduce Whittle’s index policy.

**Whittle’s Index Policy:** In each slot, with belief values $\pi_1, \ldots, \pi_N$, the user $I$ with the highest index value $W_i(\pi_i)$ is scheduled for transmission, i.e., $I = \arg \max_i W_i(\pi_i)$.

Note that, from the definition of indexability, the index value $W_i(\pi_i)$ monotonically increases with $\pi_i$. Therefore, when the Markovian channels have the same Markovian structure and vary independently across users (hence the state-index mappings are the same across users), Whittle’s index policy essentially becomes the greedy policy – schedule the user with the highest belief value.

Fig. 5 plots an example of the index value evolution for the case of positively correlated and negatively correlated channels when they stay idle, i.e., not scheduled for transmission. We see that, for the positively correlated channel, the index value behaves monotonically, while, for the negatively correlated channel, the index value shows oscillation. This resembles the evolution of the belief values, which, as proven in Lemma 3 in Appendix A, approaches steady state monotonically for the positively correlated channel, and with oscillation for the negatively correlated channel. This resemblance in Fig. 5 is expected since, from Proposition 3, we can infer that the index value monotonically increases with the belief value. Thus, in essence, from Proposition 3 and Fig. 5, we see that the index value captures the underlying dynamics of the Markovian channel.
Remark: The proposed Whittle’s Index Policy for downlink scheduling takes a simple form, and has linear complexity $O(N)$ over the number of users. This is in contrast with the PSPACE-hardness associated with directly solving the POMDP by jointly optimize over users and time [19].

VII. NUMERICAL PERFORMANCE ANALYSIS

A. Model for Simulation

In this section, we study, via numerical evaluations, the performance of Whittle’s index policy, henceforth simply the index policy, for joint estimation and scheduling in our downlink system. We consider the specific class of estimator and rate adapter structure, with pilot-aided training, discussed in Section II and illustrated in Fig. 3. We consider a fading channel with the fading coefficients quantized into two levels to reflect the two states of the Markov chain. Additive noise is assumed to be white Gaussian. The channel input-output model is given by $Y = hX + \epsilon$, where $X, Y$ correspond to transmitted and received signals, respectively, $h$ is the complex fading coefficient and $\epsilon$ is the complex Gaussian, unit variance additive noise. Conditioned on $h$, the Shannon capacity of the channel is given by $R = \log(1 + |h|^2)$. We quantize the fading coefficients such that the allowed rate at the lower state, $\delta = 0.2$ for all users. The channel state, represented by the fading coefficient, evolves as Markov chain with fading block length $T$.

We consider a class of Linear Minimum Mean Square Error (LMMSE) estimators [28]. LMMSE estimators are attractive because with additive white Gaussian noise, they can be characterized in closed form [28] and, hence, can be conveniently used in simulation. Let $\phi_\pi$ denote the optimal LMMSE estimator with prior $\{\pi, 1 - \pi\}$. We let $\Phi$ denote the set of LMMSE estimators optimized for various values of $\pi$.

B. Immediate Reward Structure

We now study the structure of the immediate reward $R(\pi)$. Note that $R(\pi)$ is optimized over the class of estimators $\Phi$. Fig. 6 illustrates $R(\pi)$, in comparison with the upper and lower bounds derived in Lemma 2, for two values of block length $T$. As established in Lemma 2, $R(\pi)$ shows a convex increasing structure and takes values within the bounds. Note that $R(\pi)$ also increases with $T$, since a larger $T$ provides more channel uses for channel probing and data transmission.

C. Near-Optimal Performance of Whittle’s Index Policy

We proceed to evaluate the performance of the index policy and compare it with the optimal policy. Given initial belief $\bar{\pi}[0]$ and policy $\Psi$, we let $V_\Psi(\bar{\pi}[0])$ and $V_\Psi^M(\bar{\pi}[0])$ be the infinite horizon and finite M-horizon reward, respectively, i.e.,

$$V_\Psi(\bar{\pi}[0]) = \sum_{t=0}^{\infty} \beta^t \mathbb{E}_{\bar{\pi}[t]}[R_{t|t} = \Psi(\bar{\pi}[t])|\pi_{t|t}[t]] \right)$$

$$V_\Psi^M(\bar{\pi}[0]) = \sum_{t=0}^{M} \beta^t \mathbb{E}_{\bar{\pi}[t]}[R_{t|t} = \Psi(\bar{\pi}[t])|\pi_{t|t}[t]] \right)$$

For notational convenience, we use $V_\Psi$ and $V_\Psi^M$ to denote $V_\Psi(\bar{\pi}[0])$ and $V_\Psi^M(\bar{\pi}[0])$ respectively, and indicate the corresponding initial belief value clearly in each numerical result.

In Fig. 7, we compare the expected rewards $V_{opt}^M$ and $V_{index}^M$ that, respectively, correspond to the optimal finite M-horizon policy and the index policy, for increasing horizon length $M$ and randomly generated system parameters. The value of $V_{opt}^M$ is obtained via brute-force search over the finite horizon. Fig. 7 illustrates the near optimal performance of the index policy. Also, as expected, the higher the value of $\beta$, the higher the expected reward.

Table I presents the performance of the index policy in a larger perspective. Here, with randomly generated system parameters, the infinite horizon reward under the index policy is compared with those of the optimal policy and a policy that ‘throws away’ the feedback from the scheduled user. Let $V_{nofb}$ denote the reward under this ‘no feedback’ policy. The infinite horizon rewards are obtained as limits of the finite horizon until 1% convergence is achieved. The high values of the quantity $\%gain = \frac{V_{opt} - V_{nofb}}{V_{opt}} \times 100\%$, in addition to underscoring the near-optimality of the index policy, also signifies the high system-level gains from exploiting the channel memory using the end-of-slot feedback.

Table I also shows the infinite horizon discounted reward of the greedy policy, denoted by $V_{greedy}$, in which the user with the highest immediate reward is selected in each slot. Note that, the greedy policy only exploits the instantaneous channel gain but does not explore the (possibly) outdated channel. On the other hand, the Whittle’s index policy captures the overall ‘exploitation vs exploration’ tradeoff. It can also be
observed from the table that the greedy algorithm provides
significant performance gain over the ‘no feedback’ policy,
and in many cases in the table (as well as in many other
numeric results of this section) provides performance close
to the Whittle’s index policy. In addition, the gains associated
with exploring the outdated channel can vary depending on
the system parameters (e.g., correlation degree, system asym-
metry, total number of users, etc.). While a comprehensive
analysis on the performance difference between the greedy
and index policy is beyond the scope of this work, we
compare their performance with respective to the asymmetry
of Markov transition statistics in Table II, with discount factor
\( \beta = 0.8 \) when \( N = 3 \). Notice that, from Table II, the larger the
asymmetry in the Markov channel statistics across users is, the
larger is the performance difference between the two policies.
This is also consistent with our observation in Section VI that
the Whittle’s index policy coincides with the greedy policy
when the channels have the same Markov statistics.

In Fig. 8 we study the effect of the channel ‘memory’ on
the performances of various baseline policies. We consider
five users with statistically identical but independently varying
channels. Thus \( p_i = p, r_i = r, i \in \{1, \ldots, 5\} \). We define
the channel ‘memory’ as the difference \( p - r \) and increase
the memory by increasing \( p \) from 0.5 to 1 and maintaining
\( r = 1 - p \). Note that, with this approach, \( p + r = 1 \). Under this
condition, the steady state probability that a channel is in the
higher state \( h \) is kept constant under varying channel memory.
This, essentially, provides a degree of fairness between
systems with different channel memories. Fig. 8 compares the
rewards \( V_{opt} \), \( V_{index} \) and \( V_{nofb} \) that respectively correspond to
the rewards under the optimal policy, the index policy, and the
‘no feedback’ policy introduced earlier, for increasing channel
memory. Note that when \( p = r \), the channel of each user
evolves \( i.i.d. \) across time, with no information contained in
the channel state feedback. Thus the policy that throws away
this feedback achieves the same performance as the optimal
policy that optimally uses this feedback, i.e., \( V_{nofb} = V_{opt} \)
when \( p = r \). Also, since the channels are \( i.i.d. \) across
users, the index policy simplifies to a ‘randomized’ policy
that schedules randomly and uniformly across users, in effect
mirroring the ‘no feedback’ policy in this setting. This explains
\( V_{index} = V_{nofb} \) when \( p = r \). As the channel memory
increases, the significance of the channel state feedback
increases, resulting in an increasing gap between the policies
that use this feedback (optimal and Whittle’s index policies)
and the ‘no feedback’ policy.

Fig. 8, along with Table I, shows that exploiting channel
memory for opportunistic scheduling can result in significant

| \( N \) | \( \beta \) | \( (p_i, r_i)_{i=1 \ldots N} \) | \( \#[0] \) | \( V_{greedy} \) | \( V_{nofb} \) | \( V_{index} \) | \( V_{opt} \) | \%gain |
|---|---|---|---|---|---|---|---|---|
| 4 | 0.6793 | (0.0494, 0.4527), (0.1230, 0.8452) | \( 0.3338, 0.4058, 0.5993, 0.0895 \) | 1.7070 | 1.3203 | 1.8688 | 1.8688 | 100% |
| 4 | 0.5896 | (0.9309, 0.6005), (0.1332, 0.8304) | \( 0.2512, 0.9372, 0.1310, 0.9408 \) | 2.0512 | 0.9566 | 2.0506 | 2.0515 | 99.9178% |
| 4 | 0.6673 | (0.3693, 0.2948), (0.6089, 0.5611) | \( 0.075972, 0.25085, 0.57596, 0.64201 \) | 1.7976 | 1.4047 | 1.9792 | 1.9792 | 100% |
| 5 | 0.6640 | (0.2479, 0.7164), (0.6289, 0.5999), (0.0453, 0.9594) | \( 0.9994, 0.9616, 0.0589, 0.3603, 0.5485 \) | 2.1519 | 1.8368 | 2.2686 | 2.2686 | 100% |
| 5 | 0.6885 | (0.6045, 0.1107), (0.7603, 0.8322), (0.0573, 0.5980) | \( 0.5503, 0.3510, 0.6418, 0.6362, 0.1317 \) | 2.3731 | 1.7720 | 2.4963 | 2.4967 | 99.9448% |
| 5 | 0.6482 | (0.2213, 0.5709), (0.4779, 0.7323), (0.6358, 0.5993) | \( 0.6569, 0.6472, 0.2626, 0.5984, 0.2110 \) | 1.7396 | 1.4017 | 1.9055 | 1.9300 | 95.3625% |
| 5 | 0.6537 | (0.6154, 0.8936), (0.7382, 0.3528), (0.1762, 0.8131) | \( 0.1564, 0.1221, 0.7627, 0.7218, 0.6516 \) | 2.2806 | 1.2746 | 2.2819 | 2.2842 | 99.7721% |
due to high mobility of users – a possibility in reality. We let when the feedback channel environment changes drastically delay is of the order of the scheduling slot length, and results the feedback delay can be i.i.d.
state feedback is subject to a random delay which is where, once a user is scheduled, the corresponding channel performance of the index policy. We consider the scenario we consider random, i.i.d. (see [29]–[32]). While these works assume deterministic delay, allocation has been studied under various settings in the past scenarios. The effect of feedback delay on channel resource important consideration that cannot be overlooked in realistic across users. The delay in the feedback channel is an
mechanism and the feedback channel. In this section, we turn, resulting from imperfections in the feedback generating various imperfections such as random delays and errors, in
realistic scenarios, the channel state feedback is subject to various imperfections as such as random delays and errors, in turn, resulting from imperfections in the feedback generating mechanism and the feedback channel. In this section, we illustrate, via numerical experiments, that the index policy is robust against feedback imperfections, i.e., numerical results suggest that the index policy performs very close to the optimal policy.

We first investigate the impact of feedback delay on the performance of the index policy. We consider the scenario where, once a user is scheduled, the corresponding channel state feedback is subject to a random delay which is i.i.d. across users. The delay in the feedback channel is an important consideration that cannot be overlooked in realistic scenarios. The effect of feedback delay on channel resource allocation has been studied under various settings in the past (see [29]–[32]). While these works assume deterministic delay, we consider random, i.i.d. feedback delay. An instance when the feedback delay can be i.i.d. and non-negligible is when the delay is of the order of the scheduling slot length, and results from channel propagation time of the feedback signal, and when the feedback channel environment changes drastically due to high mobility of users – a possibility in reality. We let $P_D(d)$ denote the probability that a channel state feedback experience $d$ slot delays, where $d \in \{0, \cdots, d_{\text{max}}\}$. Here $d = 0$ indicates an end-of-slot feedback and $d_{\text{max}}$ indicates the maximum delay that the feedback can experience. We assume the channel state feedback is time-stamped. Thus the scheduler takes this information into account when it updates the belief values upon receipt of a (possibly delayed) feedback signal. Specifically, at time slot $t$, if the latest (possibly delayed) feedback from user $i$ corresponds to the channel state $\tau$ slots ago, then

$$\pi_i[\tau] = \begin{cases} Q^f(1) & \text{if } C_i[\tau-\tau] = h_i, \\ Q^f(0) & \text{if } C_i[\tau-\tau] = l_i \end{cases}$$

where, recall that, $Q^f$ stands for the $\tau$th iteration of the function $Q$.

Now, with delay taken into account in the belief value updates, the performance of the original index policy is compared with that of the optimal policy in Table III with $d_{\text{max}} \in \{1, 2\}$ and $N \in \{3, 4\}$ and randomly generated system parameters (i.e., $P_D(d), p_i, r_i$, and $\bar{\pi}[0]$). Note that the optimal policy takes into account the stochastic of the feedback delay and is implemented by exhaustive brute-force search, as before. The high value of the quantity %opt := $V_{\text{index}} / V_{\text{opt}} \times 100\%$ indicates that Whittles index policy has a performance very close to that of the optimal policy under the delayed feedback setup. In Table IV, the performance comparison is made under more controlled choice of delay, i.e., with $d_{\text{max}}$ fixed at 2, and the tail of the delay mass function is gradually made heavy. We observe that as the delay tail grows heavier, the performances of both the optimal and the index policy decrease. This is expected because, with the delay tail growing heavier, the received channel state feedback progressively tends to become outdated, and hence the value of information contained in the feedback decreases, essentially reducing the performances of both the optimal and the index policies that use this feedback. In summary, Tables III and IV illustrate that the index policy derived for the original system without feedback delay, performs very close to the optimal policy in the system with feedback delay, essentially indicating the robustness of the index policy.

| $N$ | $\beta$ | $[P_D(0), \cdots, P_D(d_{\text{max}})]$ | $\bar{\pi}[0]$ | $(p_i, r_i)$, $i = 1, \cdots, N$ | $V_{\text{greedy}}$ | $V_{\text{index}}$ | $V_{\text{opt}}$ | %opt |
|-----|---------|-------------------------------|---------------|-----------------|--------------|--------------|---------|------|
| 3   | 0.6308  | [0.8539, 0.1461]              | [0.4868, 0.4539, 0.4468] | (0.5896, 0.3478), (0.2703, 0.6754) | (0.6978, 0.1376) | 1.3585       | 1.3709  | 1.3710 | 99.9927 % |
| 4   | 0.5587  | [0.4317, 0.5683]              | (0.4526, 0.1983, 0.6838, 0.8625) | (0.8358, 0.4462), (0.6529, 0.5207) | (0.1792, 0.7268), (0.2877, 0.9321) | 1.8002       | 1.8124  | 1.8125 | 99.9944 % |
| 3   | 0.6536  | [0.3682, 0.5216, 0.1102]      | (0.8055, 0.5767, 0.1829) | (0.9138, 0.3075), (0.2298, 0.7946) | (0.3574, 0.7851) | 2.2522       | 2.2522  | 2.2531 | 99.9601 % |
| 4   | 0.5873  | [0.6239, 0.2589, 0.1541]      | (0.6620, 0.4162, 0.8419, 0.8329) | (0.2513, 0.7258), (0.6285, 0.3801) | (0.1676, 0.8245), (0.3058, 0.6822) | 1.7071       | 1.7093  | 1.7123 | 99.8247 % |

| $N$ | $\beta$ | $[P_D(0), \cdots, P_D(d_{\text{max}})]$ | $V_{\text{greedy}}$ | $V_{\text{index}}$ | $V_{\text{opt}}$ | %opt |
|-----|---------|-------------------------------|--------------|---------|----------|------|
| 3   | 0.6     | [1, 0]                        | 1.8938       | 1.8939  | 1.8941   | 99.9984 % |
| 3   | 0.6     | [2/3, 1/3, 0]                 | 1.8440       | 1.8440  | 1.8441   | 99.9946 % |
| 3   | 0.6     | [1/3, 1/3, 1/3]               | 1.7882       | 1.7882  | 1.7883   | 99.9944 % |
| 3   | 0.6     | [0, 1/3, 2/3]                 | 1.7274       | 1.7267  | 1.7274   | 99.9995 % |
| 3   | 0.6     | [0, 0, 1]                     | 1.7195       | 1.7177  | 1.7195   | 99.8953 % |

**D. Impact of Imperfections in Channel State Feedback**

In realistic scenarios, the channel state feedback is subject to various imperfections such as random delays and errors, in turn, resulting from imperfections in the feedback generating mechanism and the feedback channel. In this section, we illustrate, via numerical experiments, that the index policy is robust against feedback imperfections, i.e., numerical results suggest that the index policy performs very close to the optimal policy.
We now study the performance of the index policy in the presence of random errors in the channel state feedbacks. This error could have initiated at the feedback generating mechanism at the user or during propagation in the feedback channel. Let $F_i[t] \in \{l_i, h_i\}$ be the feedback received at the scheduler that corresponds to actual channel state $C_i[t]$. The channel state feedback error is characterized by the mismatch probability $\varepsilon$ defined as follows:

$$
\varepsilon := \Pr(F_i[t] = l_i | C_i[t] = h_i) = \Pr(F_i[t] = h_i | C_i[t] = l_i).
$$

We first assume that the error probability, $\varepsilon$, is known at the scheduler and compare the throughput performances in Table V. Observe that, for various values of error probabilities, the index policy still has performance very close to the optimal policy, essentially suggesting its robustness against feedback errors. As also observed from the table, the performances of both the optimal and Whittle’s index policies are symmetric around $\varepsilon = 0.5$ that corresponds to the worst rewards. This is expected since, when $\varepsilon = 0.5$, the feedback contains no information about the channel state, in turn, resulting in zero gain from exploiting channel memory.

We now consider the case when the scheduler is unaware that there is a (possible) error in the channel state feedback, and study its impact on the performances of the index and the optimal policies. In this scenario, the scheduler simply trusts the feedback to be accurate when making scheduling decisions. The performances of both policies under various values of $\varepsilon$ are recorded in Table VI. Once again, the high value of $\%opt$ suggests the robustness of the index policy against feedback errors even when the scheduler is unaware of the possible presence of errors. Also, as expected, when the error probability, $\varepsilon$, increases, the performances of both policies decrease monotonically. This phenomenon contrasts to the case in Table V, when the scheduler is aware of the presence of errors and its stochastic, i.e., the value $\varepsilon$.

### E. Impact of Incorrect Transmission Rate Knowledge

We now study the robustness of the index policy under mismatch between the supportable lower state transmission rate assumed at the scheduler and the actual lower state transmission rate. Recall that $\delta_i$ denotes the allowable transmission rate at the lower state $l_i$ of user $i$. Let $\delta'_i$ denote the lower state transmission rate assumed at the scheduler for this channel. We also assume that the scheduler is unaware of the presence of this mismatch which could have resulted from errors in the initial rate estimate or when the actual underlying rate has shifted from the initial value over time. We consider the case when the actual lower state transmission rate and the assumed rate at the scheduler are identical across users, i.e., $\delta_i = \delta_j$ and $\delta'_i = \delta'_j$ for all $i, j \in \{1, \ldots, N\}$. We assume $\delta = 0.5$ and compare the performance of the index policy with that of the optimal policy in Table VII, for various values of $\delta'$. Note that as before, the optimal policy is defined within the context of the imperfection ($\delta$ mismatch, in the present case).

### Table V
**Performance of the Index Policy With Known Probability of Error in Channel State Feedback**

| $N = 3, \beta = 0.6, \text{Delay} = 2[3/3], \sigma[0] = [0.8, 0.3, 0.65]$ | $N = 4, \beta = 0.6, \text{Delay} = 2[6.04], \sigma[0] = [0.45, 0.19, 0.68, 0.86]$ |
|---|---|
| $(p_i, r_i)_i = [(0.77, 0.25), (0.34, 0.90), (0.81, 0.30)]$ | $(p_i, r_i)_i = [(0.85, 0.25), (0.6, 0.35), (0.2, 0.7), (0.3, 0.8)]$ |
| $\varepsilon$ | $V_{\text{greedy}}$ | $V_{\text{index}}$ | $V_{\text{opt}}$ | $\%opt$ | $V_{\text{greedy}}$ | $V_{\text{index}}$ | $V_{\text{opt}}$ | $\%opt$ |
| 0 | 1.8440 | 1.8440 | 1.8441 | 99.9946 % | 0 | 1.8087 | 1.8085 | 1.8091 | 99.9968 % |
| 0.25 | 1.7714 | 1.7731 | 1.7737 | 99.8535 % | 0.25 | 1.7530 | 1.7528 | 1.7530 | 99.9886 % |
| 0.5 | 1.7304 | 1.7304 | 1.7304 | 100 % | 0.5 | 1.7333 | 1.7333 | 1.7333 | 100 % |
| 0.75 | 1.7412 | 1.7371 | 1.7757 | 99.8535 % | 0.75 | 1.7530 | 1.7528 | 1.7530 | 99.9886 % |
| 1 | 1.8440 | 1.8440 | 1.8441 | 99.9946 % | 1 | 1.8087 | 1.8085 | 1.8091 | 99.9968 % |

### Table VI
**Performance of the Index Policy With Unknown Probability of Error in Channel State Feedback**

| $N = 3, \beta = 0.6, \text{Delay} = 2[3/3], \sigma[0] = [0.8, 0.3, 0.65]$ | $N = 4, \beta = 0.6, \text{Delay} = 2[6.04], \sigma[0] = [0.45, 0.19, 0.68, 0.86]$ |
|---|---|
| $(p_i, r_i)_i = [(0.77, 0.25), (0.34, 0.90), (0.81, 0.30)]$ | $(p_i, r_i)_i = [(0.85, 0.25), (0.6, 0.35), (0.2, 0.7), (0.3, 0.8)]$ |
| $\varepsilon$ | $V_{\text{greedy}}$ | $V_{\text{index}}$ | $V_{\text{opt}}$ | $\%opt$ | $V_{\text{greedy}}$ | $V_{\text{index}}$ | $V_{\text{opt}}$ | $\%opt$ |
| 0 | 1.8440 | 1.8440 | 1.8441 | 99.9946 % | 0 | 1.8087 | 1.8085 | 1.8091 | 99.9968 % |
| 0.25 | 1.7734 | 1.7722 | 1.7746 | 99.8648 % | 0.25 | 1.7346 | 1.7348 | 1.7349 | 99.9942 % |
| 0.5 | 1.7034 | 1.7031 | 1.7056 | 99.8651 % | 0.5 | 1.6694 | 1.6697 | 1.6792 | 99.4334 % |
| 0.75 | 1.6309 | 1.6305 | 1.6326 | 99.8714 % | 0.75 | 1.6006 | 1.6011 | 1.6096 | 99.4719 % |
| 1 | 1.5695 | 1.5692 | 1.5711 | 99.8791 % | 1 | 1.5444 | 1.5450 | 1.5711 | 98.3387 % |

### Table VII
**Performance of the Index Policy Under Imperfect Knowledge of Lower State Transmission Rate. System Parameters Used:**

$\delta = 0.5$, $N = 3, \beta = 0.6$, $(p_i, r_i)_i = ((0.38, 0.05), (0.16, 0.95), (0.86, 0.12)), \sigma[0] = (0.8, 0.15, 0.4)$

| $\delta'$ | $V_{\text{greedy}}$ | $V_{\text{index}}$ | $V_{\text{opt}}$ | $\%opt$ |
|---|---|---|---|---|
| 0.1 | 1.8270 | 1.8270 | 1.8270 | 100 % |
| 0.3 | 1.8323 | 1.8342 | 1.8342 | 99.9982 % |
| 0.5 | 1.8514 | 1.8770 | 1.8775 | 99.9733 % |
| 0.7 | 1.6528 | 1.6621 | 1.6641 | 99.8798 % |
| 0.9 | 0.9802 | 1.0126 | 1.0151 | 99.7537 % |
It can be observed that for various values of $\delta'$, the index policy closely tracks the performance of the optimal policy, indicating its robustness against transmission rate mismatch. Also, it can be expected that when $\delta' < \delta$, both Whittle’s index and optimal policies under-utilize the available channel rate, resulting in reduced performances. On the other hand, when $\delta' > \delta$, both policies aggressively transmit, leading to outage, thus resulting in reduced performances. This can be observed in Table VI where the performances drop monotonically as $\delta'$ deviates from $\delta$. Also, the drop in performance appears to be more severe when $\delta' > \delta$, suggesting that aggressive transmission and the resulting outages can be more detrimental than conservative transmission and the associated channel under-utilization.

VIII. Future Work

In this work, we have assumed a two-state Markov model with known statistics. To facilitate practical implementation of the proposed scheme in real-world scenarios, future directions include designing efficient low-complexity algorithms for the system where the Markov statistics are unknown and need to be learned. For example, joint learning, scheduling and statistics learning can be considered where the transition probabilities are learned across time by studying the pattern of channel states realizations. The open challenge here is how to appropriately handle the non-trivial coupling between resource-consuming statistic learning, scheduling decisions, channel estimation, and the channel memory evolution.

Another open question is to exploit channel memory for downlink scheduling when the Markov channels have more than two states. This is challenging partly because, when there are more than two states, the system memory can no longer be sufficiently captured by the scalar belief value. Therefore, more advanced techniques are needed, like those in the work [33] which studied scheduling in time-correlated channels under three-state Markov channels.

IX. Conclusion

In this paper, we studied multiuser scheduling in downlink networks with Markov-modeled outage channels. We considered the scenario where the channel state information is not perfectly known at the scheduler, thereby requiring a joint design of user selection, channel estimation and rate adaptation. We performed a two-stage joint optimization: (1) Within each slot, conditioned on the user scheduling, the channel estimation and rate adaptation are optimized to maximize the transmission rate in that slot; (2) Across slots, users are selected to maximize the long-term sum-throughput of the downlink. We formulated the scheduling problem as a partially observable Markov decision problem characterized by the classic ‘exploitation vs exploration’ trade-off. We then linked the problem to restless multiarmed bandit processes and conducted a Whittle’s indexability analysis. By establishing structural properties of the optimal reward within the indexability setup, we showed that the downlink scheduling problem is Whittle indexable – a criterion traditionally known to be hard to verify. We then explicitly characterized the Whittle’s index policy and studied its performance using extensive numerical experiments. Numerical results suggest that the index policy has near optimal performance and that significant system-level gains can be realized by exploiting the channel memory for joint channel estimation and scheduling, as evidenced by the comparison between the greedy and ‘no feedback’ policies. Numerical results also suggest that the index policy is robust against various imperfections in channel state feedback that are likely to occur in realistic scenarios.

APPENDIX A

Proof of Lemma 2

We first establish structural properties of the belief update when a user stays idle. Suppose a user has the initial belief value $\pi[0]$ and stays idle at all times, the belief value at $t^{th}$ slot is then given by $\pi[t] = Q'(\pi, [0])$, where $Q'$ is the $t^{th}$ iteration of function $Q$, given by

$$Q'(\pi) = \frac{r - (p - r)\{r - (1 + r - p)\pi\}}{1 + r - p}. \quad (12)$$

We let $\pi^0$ be the steady state distribution of the two-state channel being at the higher state, i.e.,

$$\pi^0 = \frac{r}{1 + r - p}.$$

It is clear that $\pi^0 = \lim_{t \to \infty} Q'(\pi)$. An example of the belief evolution when a user stays idle is depicted in Fig. 9. This figure shows that, when staying idle, the belief value approaches steady state monotonically for positively correlated channel and approaches steady state with oscillation for negatively correlated channel. The structural properties of $Q'(\pi, [0])$ is critical to the rest of the proof and is recorded in the following lemma.

Lemma 3: (i) For positively correlated channel (i.e., $p > r$), $\pi[t]$ converges to steady state $\pi^0$ monotonically. For negatively correlated channel (i.e., $p \leq r$), $\pi[t]$ converges to steady state $\pi^0$ with oscillation and a monotonically converging envelope.

(ii) $\min\{p, r\} \leq Q'(\pi, [0]) \leq \max\{p, r\}$ for all $t = 1, 2, \ldots$ and $\pi[0] \in [0, 1]$.

Proof: (i) Since we have $0 < p - r < 1$ for positively correlated channel and $-1 \leq p - r \leq 0$ for negatively correlated channel, it is clear from the expression of (12) that $\pi[t]$ converges to steady state $\pi^0$ monotonically and approaches steady state $\pi^0$ with oscillation and a monotonically converging envelop.

(ii) Since we have established part (i), it suffices to check that the first step transition satisfies: $\min\{p, r\} \leq Q(\pi) \leq \max\{p, r\}$, for all $\pi$, as shown below.

$$Q(\pi) = \frac{r - (p - r)(1 + r - p)\pi}{1 + r - p}.$$

For positively correlated channel, since $p - r > 0$

$$Q(\pi) \geq \frac{r - (p - r)r}{1 + r - p} = r.$$

$$Q(\pi) \leq \frac{r - (p - r)(1 + r - p)}{1 + r - p} = \frac{p(1 - p + r)}{1 + r - p} = p.$$
Formally, \( \pi \) denotes the set of belief values for which the optimal decision calculated as follows.

\[
L(\pi) = \begin{cases} 
\frac{r - (p - r)r}{1 + r - p} & \text{if } \pi > \pi^* \\
\log \frac{r - (1 + r - p)\pi^*}{r - (1 + r - p)\pi} + 1 & \text{if } \pi \leq \pi^* < \pi_0 \\
\infty & \text{if } \pi \leq \pi^* \text{ and } \pi^* \geq \pi_0 
\end{cases}
\]

The lemma is thus proved. □

Recall that \( \pi^* \) is defined as the threshold in Proposition 2. We then define \( L(\pi, \pi^*) \) as the time needed for belief value of a user to exceed \( \pi^* \) from below, starting from initial value \( \pi \).

Formally,

\[
L(\pi, \pi^*) = \min \{Q'(\pi) > \pi^*\}
\]

Using Lemma 3 and expression (12), \( L(\pi, \pi^*) \) can be calculated as follows.

- Positive correlation \((p > r)\)

\[
L(\pi, \pi^*) = \begin{cases} 
0 & \text{if } \pi > \pi^* \\
\log \frac{r - (1 + r - p)\pi^*}{r - (1 + r - p)\pi} + 1 & \text{if } \pi \leq \pi^* < \pi_0 \\
\infty & \text{if } \pi \leq \pi^* \text{ and } \pi^* \geq \pi_0 
\end{cases}
\]

- Negative correlation \((p \leq r)\)

\[
L(\pi, \pi^*) = \begin{cases} 
0 & \text{if } \pi > \pi^* \\
1 & \text{if } \pi \leq \pi^* \text{ and } Q(\pi) > \pi^*, \\
\infty & \text{if } \pi \leq \pi^* \text{ and } Q(\pi) \leq \pi^*. 
\end{cases}
\]

We shall refer to the ‘active set’ as the set of belief values for which the optimal decision is to transmit. The ‘idle set’ denotes the set of belief values for which the optimal decision is to stay idle. We proceed to derive the value functions \( V_\omega(p) \) and \( V_\omega(r) \) based on the value of \( \pi^*(\omega) \).

1. Positive correlation \((p > r)\)

- When \( \pi^*(\omega) \geq p \), the belief value \( p \) is thus in the ‘idle set’. From Lemma 3(ii), if \( \pi[0] \geq p \), the system stays idle. Hence the reward function is expressed as

\[
V_\omega(p) = \omega + \beta \omega + \beta^2 \omega + \cdots = \frac{\omega}{1 - \beta}.
\]

- When \( \pi^*(\omega) < p \), the belief value \( p \) is then in the ‘active set’. Hence from the Bellman equation in (5),

\[
V_\omega(p) = R(p) + \beta(pV_\omega(p) + (1 - p)V_\omega(r)),
\]

where, recall that, \( R(\cdot) \) is the immediate reward function defined in equation (2). Rearranging the terms yields,

\[
V_\omega(p) = \frac{R(p) + \beta(1 - p)V_\omega(r)}{1 - \beta p}.
\]

- When \( \pi^*(\omega) < r \), the value \( r \) is then in ‘active set’. From Lemma 3(ii), regardless of the scheduling decision, the belief values \( \pi[t] \), starting from \( \pi[0] = r \), stays in the ‘active set’. Therefore

\[
V_\omega(r) = \sum_{t=0}^{\infty} \beta^t R(\pi^*(\omega)) = \sum_{t=0}^{\infty} \beta^t R(Q(\pi^*(\omega))(r)).
\]

where

\[
V_\omega^1(Q^{L(r, \pi^*(\omega))}(r)) = R(Q^{L(r, \pi^*(\omega))}(r)) + \beta(Q^{L(r, \pi^*(\omega))}(r)V_\omega(p) + (1 - Q^{L(r, \pi^*(\omega))}(r))V_\omega(r)).
\]

Substituting the above expression in (13), we obtain the expression of \( V_\omega(r) = \Theta \) as given in (14), as shown at the top of the next page.

Note that the above expressions of the value functions \( V_\omega(p) \) and \( V_\omega(r) \) are not in closed form. However, the closed form expressions for \( V_\omega(p) \) and \( V_\omega(r) \) can be easily calculated based on these expressions, and are given in Lemma 2.

2. Negative correlation \((p \leq r)\).

The derivation of \( V_\omega(p) \) and \( V_\omega(r) \) for negative correlation case follows an approach similar to that for the case of positive correlation. Details are, therefore, omitted here. □

**APPENDIX B**

**PROOF OF PROPOSITION 3**

We prove that the problem is Whittle indexable by showing that \( \pi^*(\omega) \) monotonically increases with \( \omega \). It is clear from Proposition 2 that \( \pi^*(\omega) = \kappa \) for \( \omega \in [0, \delta) \). So it suffices to show that \( \pi^*(\omega) \) is strictly increasing for \( \omega \in [\delta, 1] \). The proof technique follows along the lines of [20] and is presented next. We first proceed with the following lemma, where the right derivative of the reward function is compared.

![Graph](image-url)
Lemma 4: If for all \( \omega \in [\delta, 1] \), we have
\[
\frac{dV^1_\omega(\pi)}{d\omega} \bigg|_{\pi = \pi^*(\omega)} \geq \frac{dV_\omega^0(\pi)}{d\omega} \bigg|_{\pi = \pi^*(\omega)},
\]
then \( \pi^*(\omega) \) is strictly increasing with \( \omega \) for \( \omega \in [\delta, 1] \).

Proof: The lemma is proven by contradiction. Suppose there exists \( \omega_0 \in (\delta, 1] \), such that \( \pi^*(\omega) \) is decreasing (i.e., non-increasing) at \( \omega_0 \), hence it is decreasing in a neighborhood of \( \omega_0 \), say, \( [\omega_0, \omega_0 + \Delta_0] \). Since \( V_{\omega_0 + \Delta_0}(\pi^*(\omega_0 + \Delta_0)) = V_{\omega_0}^0(\pi^*(\omega_0 + \Delta_0)) \) and \( \pi^*(\omega) \) is decreasing at \( \omega_0 \), \( \pi^*(\omega) \) is within the ‘active set’ for the \((\omega_0 + \Delta_0)\)-subsidy problem. Therefore we have \( V_{\omega_0 + \Delta_0}^1(\pi^*(\omega_0)) \geq V_{\omega_0}^0(\pi^*(\omega_0)) \).

Besides, from the definition of threshold value \( \pi^*(\omega_0) \), \( V_{\omega_0}^1(\pi^*(\omega_0)) = V_{\omega_0}^0(\pi^*(\omega_0)) \). Therefore,
\[
\frac{dV^1_\omega(\pi)}{d\omega} \bigg|_{\pi = \pi^*(\omega)} = \lim_{\Delta_0 \to 0} \frac{V_{\omega_0 + \Delta_0}^1(\pi^*(\omega)) - V_{\omega_0}^1(\pi^*(\omega))}{\Delta_0} \geq \lim_{\Delta_0 \to 0} \frac{V_{\omega_0 + \Delta_0}^0(\pi^*(\omega_0)) - V_{\omega_0}^0(\pi^*(\omega))}{\Delta_0} = \frac{dV_\omega^0(\pi)}{d\omega} \bigg|_{\pi = \pi^*(\omega)},
\]
which contradicts with the assumption.

Therefore, to establish indexability, it suffices to prove the inequality (15), i.e.,
\[
\frac{dV^1_\omega(\pi)}{d\omega} \bigg|_{\pi = \pi^*(\omega)} < \frac{dV_\omega^0(\pi)}{d\omega} \bigg|_{\pi = \pi^*(\omega)}. \tag{15}
\]
Let \( D_\omega(\pi) \) be the discounted time the \( \omega \)-subsidy process, with initial belief \( \pi \), made passive, i.e.,
\[
D_\omega(\pi) = \sum_{t=0}^{\infty} \beta^t \mathbf{1}(a[t] = 0).
\]

Noting that giving the value of \( \pi^*(\omega) \), the studying the belief value evolution follows the same pattern as in ON/OFF channel case, hence the expression of \( D_\omega(\pi) \) takes the same form as given in [20]. It follows from [16] and [20] that \( D_\omega(\pi) = \frac{dV^1_\omega(\pi)}{d\omega} \). Taking derivative of the \( V^1_\omega(\pi) \) and \( V^2_\omega(\pi) \) expressions (8)-(9) with respect to \( \omega \), the objective (15) now becomes
\[
\beta(\pi^*(\omega))D_\omega(p) + (1 - \pi^*(\omega))D_\omega(r) < 1 + \beta D_\omega(Q(\pi^*(\omega))). \tag{16}
\]

Case (1): If \( 0 \leq \pi^*(\omega) < \min(p, r) \), from Lemma 3(ii), starting from the initial belief value \( \pi[0] = r \) or \( \pi[0] = p \), the belief value \( \pi[t] \) never evolves below \( \pi^*(\omega) \), hence the project is active at all times under optimal control. Therefore \( D_\omega(p) = D_\omega(r) = D_\omega(Q(\pi^*(\omega))) = 0 \). Equation (16) thus holds.

Case (2): If \( \pi_0 \leq \pi^*(\omega) \leq 1 \), starting from initial belief \( \pi[0] = Q(\pi^*(\omega)) \), the belief value \( \pi[t] \) always stays within the ‘idle set’, i.e., \( D_\omega(Q(\pi^*(\omega))) = \frac{1}{1 - \beta} \). Equation (16) holds since \( D_\omega(p) \leq 1 + \beta + \beta^2 + \cdots = \frac{1}{1 - \beta} \) and, similarly, \( D_\omega(r) \leq \frac{1}{1 - \beta} \).

Case (3): If \( \min(p, r) \leq \pi^*(\omega) \leq \pi_0 \), from Lemma 3(ii), \( Q(\pi^*(\omega)) \) is in ‘active set’. Since
\[
V_\omega(Q(\pi^*(\omega))) = R(Q(\pi^*(\omega))) + \beta Q(\pi^*(\omega))V_\omega(p) + (1 - Q(\pi^*(\omega)))V_\omega(r),
\]
we have
\[
D_\omega(Q(\pi^*(\omega))) = \beta[Q(\pi^*(\omega))D_\omega(p) + (1 - Q(\pi^*(\omega)))D_\omega(r)]. \tag{17}
\]
We then discuss (17) separately for negatively and positively correlated channels.

- Negatively correlated channel \((p < r)\). Since \( r > \pi^0 > \pi^*(\omega) \), the belief value \( r \) is in the ‘active set’, hence
\[
V_\omega(r) = R(r) + \beta(rV_\omega(p) + (1 - r)V_\omega(r)).
\]
Therefore, we have
\[
D_\omega(r) = \beta(rD_\omega(p) + (1 - r)D_\omega(r)). \tag{18}
\]
Substituting equation (17) and (18) in (16), we get
\[
\frac{\beta}{1 - \beta(1 - r)} D_\omega(p)(1 - \beta) \times (\beta r + \pi^*(\omega) - \beta Q(\pi^*(\omega))) < 1.
\]

Following the same technique as in [20], the above inequality can be verified by substituting \( \pi^*(\omega) \) by \( \pi^0 \) and \( D_\omega(p) \) by \( \frac{1}{1 - \beta} \).

- Positively correlated channel \((p > r)\). In this case, \( p \) is in the ‘active set’, hence
\[
V_\omega(p) = R(p) + \beta(pV_\omega(p) + (1 - p)V_\omega(r)).
\]
Taking derivative with respect to \( \omega \) we have,
\[
D_\omega(p) = \beta(pD_\omega(p) + (1 - p)D_\omega(r)). \tag{19}
\]
Substituting equations (17) and (19) in (16), we have
\[
\beta D_\omega(r)(1 - \beta) \times (1 - \pi^*(\omega) - \beta Q(\pi^*(\omega))) < 1.
\]

We note that the expression of \( D_\omega(r) \) takes the same form as in [20]. By applying the same technique as in [20], it can be checked that the above inequality indeed holds. Therefore the inequality (16) is justified and hence indexability holds.

\section*{Appendix C}
\textbf{Proof of Proposition 4}

For \( \omega \)-subsidy problem of user \( i \), from indexability, we know that \( \pi^*(\omega) \) strictly increases from 0 to 1 as \( \omega \) increases from \( \delta_i \) to 1. Hence the index value, from its definition in (10), is the subsidy value for which the active and idle decisions are equally attractive. We can hence derive index value \( W_i(\pi_i) \) by
equating $V_{i,\omega}^1(\pi_i)$ and $V_{i,\omega}^0(\pi_i)$ and solve for $\omega$ as a function of $\pi_i$, i.e.,

$$W_i(\pi_i) + \beta V_i, W_i(\pi_i)(Q_i(\pi_i)) = R(\pi_i) + \beta \left[ \pi_i V_i, W_i(p_i) + (1 - \pi_i)V_i, W_i(r_i) \right].$$

(20)

Note that the expressions of $V_{i,\omega}^1(p_i)$ and $V_{i,\omega}^0(r_i)$ have been given by Lemma 2. Substituting in (20) the values of $V_{i,\omega}^1(p_i)$ and $V_{i,\omega}^0(r_i)$, we obtain the index value expressions, explained in the following.

Case (1): Positively correlation ($p_i > r_i$).

- If $\pi_i \geq p_i$, the belief value $Q_i(\pi_i)$, $p_i$, $r_i$ are in the ‘idle set’ and, starting from initial belief $\pi_i[0] = Q_i(\pi_i)$ or $\pi_i[0] = p_i$ or $\pi_i[0] = r_i$, $\pi_i[i]$ will stay in the ‘idle set’. Hence

$$V_{i,\omega}(Q_i(\pi_i)) = V_{i,\omega}(p_i) = V_{i,\omega}(r_i) = \frac{\omega}{1 - \beta}.$$

Substituting the above expressions in (20) we obtain that

$$W_i(\pi_i) = R(\pi_i).$$

- If $\pi_i^0 \leq \pi_i < p_i$, then $p_i$ is in ‘active set’, and starting from initial belief $\pi_i[0] = r_i$ or $\pi_i[0] = Q_i(\pi_i)$, $\pi_i[i]$ stays within ‘idle set’ at all times. Hence

$$V_{i,\omega}(Q_i(\pi_i)) = V_{i,\omega}(r_i) = \frac{\omega}{1 - \beta}.$$

Substituting the above expressions and the expression of $V_{i,\omega}(p_i)$ (given in Lemma 2) in (20), we get

$$W_i(\pi_i) = \frac{\beta \pi_i R(p_i) + (1 - \beta p_i)R(r_i)}{1 + \beta \pi_i - \beta p_i}.$$

- If $\pi_i < \pi_i^0$, then the value $Q_i(\pi_i)$ is in the ‘active set’. Therefore,

$$V_{i,\omega}(Q_i(\pi_i)) = R(Q_i(\pi_i)) + \beta Q_i(\pi_i) V_{i,\omega}(p_i) + (1 - Q_i(\pi_i)) V_{i,\omega}(r_i).$$

Again, substituting the expression of $V_{i,\omega}(Q_i(\pi_i))$ in (20), we have

$$W_i(\pi_i) = \frac{R(\pi_i) - \beta R_i(Q_i(\pi_i)) + \beta (1 - \pi_i - \beta Q_i(\pi_i)) \sum_{t=0}^{\infty} \beta^t R \left( \frac{r_i + (p_i - r_i)^{t+1}(1 - p_i)}{1 + r_i - p_i} \right)}{1 - \beta (1 - p_i - \beta Q_i(\pi_i))}. \quad \text{(26)}$$

$$W_i(\pi_i) = \left[ R(\pi_i) - \beta R_i(Q_i(\pi_i)) \right] + \beta \left[ \pi_i - \beta Q_i(\pi_i) \right] \sum_{t=0}^{\infty} \beta^t R \left( \frac{r_i + (p_i - r_i)^{t+1}(1 - p_i)}{1 + r_i - p_i} \right)$$

$$+ \beta \left[ (1 - \pi_i) - \beta (1 - Q_i(\pi_i)) \right] \sum_{t=0}^{\infty} \beta^t R \left( \frac{r_i - (p_i - r_i)^{t+1}r_i}{1 + r_i - p_i} \right). \quad \text{(27)}$$
• If $Q_i(p_i) \leq r_i$, then
  \[ W_i(\pi_i) = \frac{(1-\beta)[1-\beta(1-r_i)]R(\pi_i)+\beta(1-\beta)(1-\pi_i)R(\pi_i)}{[1-\beta][1-\beta(1-r_i)]-\beta^2(1-\pi_i)r_i} . \]

• If $\pi_0^i < \pi_i < Q_i(p_i)$, the index value is expressed as
  \[ W_i(\pi_i) = \frac{(1-\beta)R(\pi_i)\Delta_i + \beta(1-\beta)\pi_i\Omega_i + \beta(1-\beta)(1-\pi_i)\Upsilon_i}{\Delta_i(1-\beta)(1-\beta(1-r_i))\pi_i(1-\beta^2)r_i(1-\pi_i)} , \]
  where
  \[ \Delta_i = (1-\beta(1-r_i))(1-\alpha^2Q_i(p_i)) = \beta^{-1}r_i(1-\alpha Q_i(p_i)) , \]
  \[ \Omega_i = \beta(1-\beta(1-r_i))R_i(Q_i(p_i)) + \beta^2(1-\alpha Q_i(p_i))R_i(\pi_i) , \]
  \[ \Upsilon_i = \beta^2 r_i R_i(Q_i(p_i)) + (1-\alpha^2 Q_i(p_i))R_i(\pi_i) . \]

• If $p_i \leq \pi_i < \pi_0^i$, $W_i(\pi_i)$ is given in (26), as shown at the bottom of the previous page, where $\Delta_i$, $\Omega_i$ and $\Upsilon_i$ are given by (23)-(25), respectively.

• If $\pi_i < p_i$, the index value $W_i(\pi_i)$ is given in equation (27), as shown at the bottom of the previous page.

REFERENCES

[1] R. Knopp and P. A. Humblet, “Information capacity and power control in single-cell multiuser communications,” in Proc. IEEE Int. Conf. Commun., Jun. 1995, pp. 331–335.

[2] D. Tse and P. Viswanath, Fundamentals of Wireless Communication. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[3] L. Tassiulas, “Scheduling and performance limits of networks with constantly changing topology,” IEEE Trans. Inf. Theory, vol. 43, no. 3, pp. 1067–1073, May 1997.

[4] M. J. Neely, E. Modiano, and C. E. Rohrs, “Dynamic power allocation and routing for time-varying wireless networks,” IEEE J. Sel. Areas Commun., vol. 23, no. 1, pp. 89–103, Jan. 2005.

[5] S. Shakkottai and A. L. Stolyar, “Scheduling for multiple flows sharing a time-varying channel: The exponential rule,” Amer. Math. Soc. Trans., vol. 2, no. 207, pp. 185–202, 2002.

[6] A. Eryilmaz and S. P. Stolyar, “Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control,” IEEE/ACM Trans. Netw., vol. 15, no. 6, pp. 1333–1344, Dec. 2007.

[7] M. J. Neely, “Max weight learning algorithms with application to scheduling in unknown environments.” [Online]. Available: http://arxiv.org/abs/0902.0630

[8] P. C. Thejasi, J. Zhang, M.-O. Pun, H. V. Poor, and D. Zheng, “Distributed opportunistic scheduling with two-level probing,” IEEE/ACM Trans. Netw., vol. 18, no. 5, pp. 1464–1477, Oct. 2009.

[9] L. A. Johnston and V. Krishnamurthy, “Opportunistic file transfer over a fading channel: A POMDP search theory formulation with optimal threshold policies,” IEEE Trans. Wireless Commun., vol. 5, no. 2, pp. 394–405, Feb. 2006.

[10] S. Murugesan, P. Schniter, and N. B. Shroff, “Multiuser scheduling in a Markov-modeled downlink using randomly delayed ARQ feedback,” IEEE Trans. Inf. Theory, vol. 58, no. 2, pp. 1025–1042, Feb. 2012.

[11] S. H. A. 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, Sep. 2009.

[12] C.-P. Li and M. J. Neely, “Exploiting channel memory for multiuser wireless scheduling without channel measurement: Capacity regions and algorithms,” Perform. Eval., vol. 68, no. 8, pp. 631–657, Aug. 2011.

[13] C.-P. Li and M. J. Neely, “Network utility maximization over partially observable Markovian channels,” in Proc. IEEE WiOpt, May 2011, pp. 17–24.

[14] C. Safran and C. G. Chute, “Exploration and exploitation of clinical databases,” Int. J. Bio-Med. Comput., vol. 39, no. 1, pp. 151–156, Apr. 1995.

[15] L. P. Kaelbling, M. L. Littman, and A. W. Moore, “Reinforcement learning: A survey,” J. Artif. Intell. Res., vol. 4, pp. 273–285, May 1996.

[16] P. Whittle, “Restless bandits: Activity allocation in a changing world,” J. Appl. Probab., vol. 25, pp. 287–298, Jan. 1988.

[17] K. D. Glazebrook, H. M. Mitchell, and P. S. Ansell, “Index policies for the maintenance of a collection of machines by a set of repairmen,” Eur. J. Oper. Res., vol. 165, no. 1, pp. 267–284, Aug. 2005.

[18] P. S. Ansell, K. D. Glazebrook, J. Nino-Mora, and M. O’Keeffe, “Whittle’s index policy for a multi-class queueing system with convex holding costs,” Math. Methods Oper. Res., vol. 57, no. 1, pp. 21–39, Apr. 2003.

[19] C. H. Papadimitriou and J. N. Tsitsiklis, “The complexity of optimal queuing network control,” Math. Oper. Res., vol. 24, no. 2, pp. 293–305, 1999.

[20] K. Liu and Q. Zhao, “Indexability of restless bandit problems and optimality of Whittle index for dynamic multichannel access,” IEEE Trans. Inf. Theory, vol. 56, no. 11, pp. 5547–5567, Nov. 2010.

[21] K. Jagannathan, S. Mannor, I. Menache, and E. Modiano, “A state action frequency approach to throughput maximization over uncertain wireless channels,” in Proc. IEEE INFOCOM, Shanghai, China, Apr. 2011, pp. 491–495.

[22] W. Ouyang, A. Eryilmaz, and N. B. Shroff, “Asymptotically optimal downlink scheduling over Markovian fading channels,” in Proc. IEEE INFOCOM, Orlando, FL, USA, Mar. 2012, pp. 1224–1232.

[23] W. Ouyang, A. Eryilmaz, and N. B. Shroff, “Low-complexity optimal scheduling over correlated fading channels with ARQ feedback,” in Proc. IEEE WoPi, Paderborn, Germany, May 2012, pp. 270–277.

[24] E. J. Sondik, “The optimal control of partially observable Markov decision processes,” Ph.D. dissertation, Dept. Elect. Eng., Stanford Univ., Stanford, CA, USA, 1971.

[25] D. P. Bertsekas, Dynamic Programming and Optimal Control, vols. 1–2. Belmont, MA, USA: Athena Scientific, 2005.

[26] E. J. Sondik, “The optimal control of partially observable Markov processes over the infinite horizon: Discounted costs,” Oper. Res., vol. 26, no. 2, pp. 282–304, Mar. 1978.

[27] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[28] T. Kailath, A. H. Sayed, and B. Hassibi, Linear Estimation. Englewood Cliffs, NJ, USA: Prentice-Hall, 2000.

[29] H. Viswanathan, “Capacity of Markov channels with receiver CSI and delayed feedback,” IEEE Trans. Inf. Theory, vol. 45, no. 2, pp. 761–771, Mar. 1999.

[30] L. Ying and S. Shakkottai, “On throughput optimality with delayed network-state information,” in Proc. Inf. Theory Appl. Workshop, Jan./Feb. 2008, pp. 339–344.

[31] K. Kar, X. Luo, and S. Sarkar, “Throughput-optimal scheduling in multichannel access point networks under infrequent channel measurements,” IEEE Trans. Wireless Commun., vol. 7, no. 7, pp. 2619–2629, Jul. 2008.

[32] V. Annadureddy, D. Marathe, T. R. Ramya, and S. Bhashyam, “Outage probability of multiple-input single-output (MISO) systems with delayed feedback,” IEEE Trans. Commun., vol. 57, no. 2, pp. 319–326, Feb. 2009.

[33] S. Murugesan and P. Schniter, “Opportunistic multiuser scheduling in a three state Markov-modeled downlink.” [Online]. Available: http://arxiv.org/abs/0904.1754

Wenzhuo Ouyang (S’07–M’14) received his Ph.D. degree in Electrical and Computer Engineering at the Ohio State University in 2014, and his B.E. degree in Electronic Engineering at Shanghai Jiao Tong University in 2008. During his Bachelor’s study, he was an exchanged student in system and control group in department of electrical and computer engineering at the University of New Mexico in 2007. He received best student paper award at IEEE WiOPT 2012 and student travel award at INFOCOM 2011. He served the TPC of S Workshop 2011, IEEE WiOPT 2015, and IEEE WASA 2015. He is currently a postdoctoral researcher at Rice University. His research interests span the areas of communication systems, wireless networks, Markov decision process and stochastic systems.
Sugumar Murugesan (S’10–M’11) received the B.E. degree in Electronics and Communication Engineering from the College of Engineering, Anna University, India in 2004 and the M.S. and Ph.D. degrees in Electrical and Computer Engineering from The Ohio State University, USA, in 2006 and 2010, respectively. Between 2010 and 2012, Dr. Murugesan was affiliated with the research faculty of ECEE department at Arizona State University. Since 2012, he has been involved in applied research and development in the telecommunication industry. He is a recipient of the OSU University Fellowship (2004-2005) and served on the TPC of IEEE INFOCOM 2012 and 2013. His research interests include communication networks, sequential decision processes and sustainable systems.

Atilla Eryilmaz (S’00–M’06) received his B.S. degree in Electrical and Electronics Engineering from Boğaziçi University, Istanbul, in 1999, and the M.S. and Ph.D. degrees in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign in 2001 and 2005, respectively. Between 2005 and 2007, he worked as a Postdoctoral Associate at the Laboratory for Information and Decision Systems at the Massachusetts Institute of Technology. He is currently an Associate Professor of Electrical and Computer Engineering at The Ohio State University. Dr. Eryilmaz’s research interests include design and analysis for communication networks, optimal control of stochastic networks, optimization theory, distributed algorithms, pricing in networked systems, and information theory. He received the NSF-CAREER and the Lumley Research Awards in 2010.

Ness B. Shroff (S’91–M’93–SM’01–F’07) received his Ph.D. degree in Electrical Engineering from Columbia University in 1994. He joined Purdue university immediately thereafter as an Assistant Professor in the school of Electrical and Computer Engineering. At Purdue, he became Full Professor of ECE in 2003 and director of CWSA in 2004, a university wide center on wireless systems and applications. In July 2007, he joined The Ohio State University, where he holds the Ohio Eminent Scholar endowed chair in Networking and Communications, in the departments of ECE and CSE. From 2009-2012, he served as a Guest Chaired professor of Wireless Communications at Tsinghua University, Beijing, China, and currently holds an honorary Guest professor at Shanghai Jiaotong University in China. His research interests span the areas of communication, social, and cyberphysical networks. He is especially interested in fundamental problems in the design, control, performance, pricing, and security of these networks. Dr. Shroff is a past editor for IEEE TRANSACTION ON NETWORKING/ACM Transaction on Networking and the IEEE COMMUNICATION LETTERS. He currently serves on the editorial board of the Computer Networks Journal, IEEE Network Magazine, and the Networking Science journal. He has chaired various conferences and workshops, and co-organized workshops for the NSF to chart the future of communication networks. Dr. Shroff is a Fellow of the IEEE and an NSF CAREER awardee. He has received numerous best paper awards for his research, e.g., at IEEE INFOCOM 2008, IEEE INFOCOM 2006, Journal of Communication and Networking 2005, Computer Networks 2003 (two of his papers also received runner-up awards at IEEE INFOCOM 2005 and INFOCOM 2013), and also student best paper awards (from all papers whose first author is a student) at IEEE WiOPT 2013, IEEE WiOPT 2012 and IEEE IWQoS 2006. Dr. Shroff is among the list of highly cited researchers from Thomson Reuters (formerly ISI web of Knowledge) and in Thomson Reuters Book on The World’s Most Influential Scientific Minds in 2014. In 2014, he received the IEEE INFOCOM achievement award for seminal contributions to scheduling and resource allocation in wireless networks.