Downlink Scheduling Over Markovian Fading Channels

Wenzhuo Ouyang, *Member, IEEE*, Atilla Eryilmaz, *Member, IEEE*, and Ness B. Shroff, *Fellow, IEEE*

***Abstract*—We consider the scheduling problem in downlink wireless networks with heterogeneous, Markov-modulated, ON/OFF channels. It is well known that the performance of scheduling over fading channels relies heavily on the accuracy of the available channel state information (CSI), which is costly to acquire. Thus, we consider the CSI acquisition via a practical ARQ-based feedback mechanism whereby channel states are revealed at the end of only scheduled users' transmissions. In the assumed presence of temporally correlated channel evolutions, the desired scheduler must optimally balance the *exploitation–exploration tradeoff*, whereby it schedules transmissions both to exploit those channels with up-to-date CSI and to explore the current state of those with outdated CSI. In earlier works, Whittle's Index Policy had been suggested as a low-complexity and high-performance solution to this problem. However, analyzing its performance in the typical scenario of statistically heterogeneous channel state processes has remained elusive and challenging, mainly because of the highly coupled and complex dynamics it possesses. In this work, we overcome these difficulties to rigorously establish the asymptotic optimality properties of Whittle's Index Policy in the limiting regime of many users. More specifically: 1) we prove the *local optimality* of Whittle's Index Policy, provided that the initial state of the system is within a certain neighborhood of a carefully selected state; (2) we then establish the *global optimality* of Whittle's Index Policy under a recurrence assumption that is verified numerically for our problem. These results establish that Whittle's Index Policy possesses analytically provable optimality characteristics for scheduling over heterogeneous and temporally correlated channels.**

***Index Terms*—Imperfect CSI, Markov channel model, Restless Multiarmed Bandit Problem, scheduling algorithm, Whittle's Index Policy.**

Manuscript received April 22, 2014; revised February 25, 2015; accepted April 24, 2015; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor M. Chen. This work was supported in part by the NSF under Grants CAREER-CNS-0953515, CNS-WiFiUS-1456806, CCSS-EARS-1444026, CCF-0916664, CNS-1012700, and CNS-1421576; the DTRA under Grants HDTRA 1-08-1-0016, HDTRA 1-14-1-0058, and HDTRA1-15-1-0003; and the Qatar National Research Fund (QNRF) under the National Priorities Research Program (NPRP) under Grants NPRP 09-1168-2-455 and NPRP 7-923-2-344. A preliminary version of this paper appeared in the IEEE INFOCOM 2012.

W. Ouyang is with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77096 USA (e-mail: wenzhuo.ouyang@rice.edu).

A. Eryilmaz is with the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210 USA (e-mail: eryilmaz@ece.osu.edu).

N. B. Shroff is with the Departments of Electrical and Computer Engineering and Computer Science and Engineering, The Ohio State University, Columbus, OH 43210 USA (e-mail: shroff@ece.osu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNET.2015.2438009

I. INTRODUCTION

CHANNEL fluctuation is an intrinsic characteristic of wireless communications. Such a variation calls for allocation of the wireless resources in a dynamic manner, leading to the classic *opportunistic scheduling principle* (e.g., [1] and [2]). Under the assumption that the instantaneous channel state information (CSI) is fully available to the scheduler, many efficient opportunistic scheduling algorithms (e.g., [4]–[6]) have been proposed and extensively studied.

More recent works have focused on designing scheduling algorithms under imperfect CSI, where the channel state is modeled as independent and identically distributed (*i.i.d.*) processes across time (e.g., [9]–[13]). On the other hand, although the *i.i.d.* channel model brings ease of analysis, it fails to capture the time-correlation of the fading channels [3]. Specifically, it fails to exploit the channel memory, which is a critical resource for making scheduling decisions. However, designing efficient scheduling schemes under time-correlated channels with imperfect CSI is a very challenging problem. The challenge is mainly because of the difficulty in making the classic "exploitation versus exploration" tradeoff (e.g., [7] and [8]), in which a scheduler needs to strike a balance between selecting the channels with up-to-date channel memory that guarantees high immediate gains, or to explore the channels with outdated CSI for more informed decisions and associated future throughput gains.

We consider the downlink scheduling problem where a base station transmits to the users within its transmission range, subject to scheduling constraints. To model the time correlations present over fading channels, we assume that wireless channels evolve as Markov-modulated ON/OFF processes. The channel state information is obtained from ARQ-based feedback, only *after* each scheduled transmission. Nevertheless, due to time correlation, the memory of the past channel state can be used to predict the current channel state *prior to* scheduling decision. Hence, channel memory should be intelligently exploited by the scheduler in order to achieve high throughput performance.

In a related work [14], a similar problem is considered under delayed CSI, where it is assumed that perfect CSI is available within a maximum delay, which is in turn smaller than the delay experienced by the ARQ feedback used for collision detection. These assumptions allow the scheduling decisions to be decoupled from CSI acquisition, which leads to the development of centralized as well as distributed schedulers. However, this approach does not use ARQ as a means of acquiring improved channel quality information. In contrast, in our setup the nature of ARQ feedback creates an implicit impact of scheduling decisions on the CSI feedback, which completely transforms the nature of the optimal scheduler design, and therefore requires a different approach. Under the scenario where all the channels have *identical Markov statistics*, round-robin-based algorithms (e.g., [15]–[18]) have been shown to possess optimality properties in throughput performance. However, the round-robin-based algorithms are no longer optimal in *asymmetric scenarios*, e.g., when different channels have different Markov transition statistics, as is naturally the case in typical heterogeneous conditions.

Under the asymmetric scenarios, our downlink scheduling problem is an example of the classic Restless Multiarmed Bandit Problem (RMBP) [19]. Low-complexity Whittle's Index Policies [19] for the downlink scheduling problem have been proposed in [20] and [21] based on RMBP theory. However, although Whittle's Index Policy can bring significant throughput gains by exploiting the channel memory [21], the analytical characterization of its performance under asymmetric scenarios is very challenging and prohibitively technical. This is because asymmetry leads to a sophisticated interplay of memory evolution among channels with heterogeneous characteristics, which brings a significant challenge to the analysis of Whittle's Index Policy not present in the perfectly symmetric scenario.

For RMBP problems under general scenarios, Whittle's Index Policy has been proven in [22] to be asymptotically optimal as the number of users grows, provided a nontrivial condition, known as Weber's condition, holds. Nonetheless, Weber's condition concerns the global convergence of a nonlinear differential equation, which is extremely difficult to verify even numerically in our downlink scheduling scenario. In [23], optimality properties of general RMBP are studied, where a suboptimal BALANCED-INDEX policy, as well as a THRESHOLD-WHITTLE policy, are proved to provide 2-approximation performance, i.e., achieves at least half of the optimal reward. Our work takes a different approach than [23] to specifically study the per-user throughput performance of the Whittle's Index Policy for downlink scheduling and consider the strict optimality metric in the asymptotic regime when the number of users scales.

In this paper, we take significant steps in analyzing the optimality properties of Whittle's Index Policy for the downlink scheduling problem in the presence of channel heterogeneity. Specifically, our contributions are as follows.

- We apply the Whittle's index framework to our downlink scheduling problem and identify the optimal policy for the problem with a relaxed constraint in Section III. This policy, with carefully selected randomization, provides a performance upper bound to Whittle's Index Policy.
- We establish the local optimality of Whittle's Index Policy in the asymptotic regime when the number of users scales in Section V. Specifically, we show that the performance of the index policy can get arbitrarily close to that of the relaxed-constraint optimal policy, provided that the initial state of the system is within a certain neighborhood of a carefully selected state.
- Based on the local optimality result, under a numerically verifiable recurrence assumption, we then establish the global optimality of Whittle's Index Policy in the limiting regime of many users in Section VI.

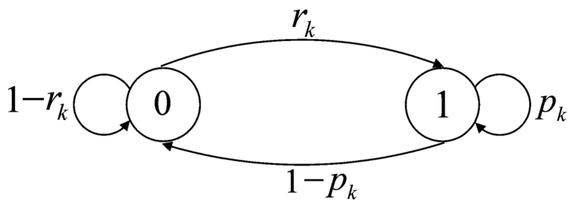

Fig. 1. Two-state Markov chain model for channels in class k.

II. System Model and Problem Formulation

A. Downlink Wireless Channel Model

We consider a time-slotted, wireless downlink system with one base station and N users. The wireless channel $C_i[t]$ between base station and user i remains static within each time-slot t and evolves stochastically across time-slots, independently across users. We adopt the simplest nontrivial model of time-correlated fading channels by considering two-state ON/OFF channels, where the state space of channel i is $\boldsymbol{\mathcal{S}}_i = \{0,1\}$, with the value of each state representing the transmission rate a channel can support at the state.

One important component of our model is the inclusion of channel heterogeneity that the users will typically experience in real systems. Such asymmetry creates a significant challenge to the design and analysis of optimal scheduling schemes compared to perfectly symmetric channels. To avoid cumbersome notation and unessential technical complications, in this work we model channel asymmetry by considering only *two classes* of channel statistics. Specifically, for all the channels in class k, $k=1,2$, their states evolve according to the same Markov statistics. However, these characteristics differ between classes. The state transition of channels in class k is depicted in Fig. 1, represented by a 2×2 probability transition matrix

$$\mathbb{P}_k = \begin{bmatrix} p_k & 1-p_k \\ r_k & 1-r_k \end{bmatrix}$$

where

$$\begin{aligned} p_k &:= \text{prob}(C_i[t]{=}1 \mid C_i[t{-}1]=1) \\ r_k &:= \text{prob}(C_i[t]{=}1 \mid C_i[t{-}1]{=}0). \end{aligned}$$

for channel i in class k. The number of class-k channels is $\gamma_k N$, $k \in \{1,2\}$ with γ_k being the *proportion* of channels in class k with respect to the total number N of channels.

We study the scenario where all the Markovian channels are positively correlated, i.e., $p_k > r_k$ for $k=1,2$. This assumption, which is commonly made in this domain (e.g., [17], [18], and [24]), means that the channel evolution has a positive auto-correlation. Hence, roughly speaking, the channel has a stronger potential to stay in its previous state than jumping to another, which is typical especially in slow fading environment. For ease of exposition, we shall exclude the trivial case when $r_k{=}0$ or $p_k=1$, $k=1,2$.

B. Scheduling Model—Belief Value Evolution

We assume that the base station can simultaneously transmit to at most $\alpha N \in \mathbb{Z}^+$ users in a time-slot without interference, where $\alpha \in (0,1]$ stands for the maximum *fraction* of users

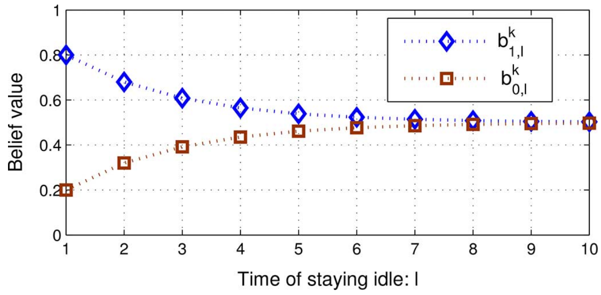

Fig. 2. Belief values update when staying idle, $p_k = 0.8, r_k = 0.2, b_s^k = 0.5$.

that can be activated. For example, in a multichannel communication model, α would correspond to the fraction of all users that can be simultaneously serviced in unit time. However, the scheduler does not know the exact channel state in the current slot when the scheduling decision is made. Instead, the scheduler maintains a *belief value* $\pi_i[t]$ for each channel i, which is defined as the probability of channel i being in the ON state at the beginning of slot t. The accurate channel state is revealed via ACK/NACK feedback from the scheduled users, only at the end of each time-slot after the data is transmitted. This accurate channel state feedback is in turn used by the scheduler to update the belief values.

For user i in class k, $k = 1, 2$, let $a_i[t] \in \{0, 1\}$ indicate whether the user is selected for transmission in slot t. Then, from the definition the belief values, $\pi_i[t]$ evolves as follows:

$$\pi_i[t+1] = \begin{cases} p_k, & \text{if } a_i[t] = 1, C_i[t] = 1 \\ r_k, & \text{if } a_i[t] = 1, C_i[t] = 0 \\ \pi_i[t]p_k + (1-\pi_i[t])r_k, & \text{if } a_i[t] = 0. \end{cases} \tag{1}$$

In our setup, belief values are known to be sufficient statistics to represent the past scheduling decisions and feedback (e.g., [16], [25]). In the meanwhile, in our ON/OFF channel model, $\pi_i[t]$ also equals the expected throughput contributed by channel i if it is scheduled in time-slot t.

For a user in class k, $k = 1, 2$, we use $b_{c,l}^k$ to denote its belief value when the most recent observed channel was $c \in \{0, 1\}$, and is l slots in the past. From the belief update rule (1), $b_{c,l}^k$ can be calculated as a function of $l \geq 1$ as

$$b_{0,l}^k = \frac{r_k - (p_k - r_k)^l r_k}{1 + r_k - p_k} \quad b_{1,l}^k = \frac{r_k + (1 - p_k)(p_k - r_k)^l}{1 + r_k - p_k}.$$

Fig. 2 illustrates the belief value update when a channel stays idle (i.e., $a_i = 0$). It is clear that if the scheduler is never updated of the state of channel i (in class k), the belief value will converge to its stationary probability of being ON, denoted by the stationary belief value $b_s^k := r_k/(1 + r_k - p_k)$.

The vector $\vec{\boldsymbol{\pi}}[t] = (\pi_1[t], \cdots, \pi_N[t])$ denotes the belief values of all channels at the beginning of slot t. We use $\mathcal{B}_k$ to represent the set of the belief values for class-k channels, where $\mathcal{B}_k = \{b_s^k, b_{c,l}^k, c \in \{0, 1\}, l \in \mathbb{Z}^+\}$. We assume that the system starts to operate from slot $t = 0$. At the beginning of slot 0, for each channel the scheduler has either observed its channel state before, or has never been updated of its channel state, i.e., with belief value b_s^k. It is then clear that, based on the belief update rule (1), $\pi_i[t] \in \mathcal{B}_k$ for all $t \geq 0$, i.e., each belief value $\pi_i[t]$ evolves over countably many states.

In the rest of the paper, we shall use "belief value" and "belief state" interchangeably.

C. *Downlink Scheduling Problem—POMDP Formulation*

We consider the broad class U of (possibly nonstationary) scheduling policies that makes a scheduling decision based on the history of observed channel states and scheduling actions. The downlink scheduling problem is then to identify a policy in U that maximizes the infinite horizon, *time average expected throughput*, subject to the constraint on the number of users selected at each time-slot. Given the initial state $\vec{\boldsymbol{\pi}}[0]$, the problem is formulated as

$$\max_{u \in U} \liminf_{T \to \infty} \frac{1}{T} E \left[\sum_{t=0}^{T-1} \sum_{i=1}^{N} \pi_i[t] \cdot a_i^u[t] \mid \vec{\boldsymbol{\pi}}[0] \right] \tag{2}$$

$$\text{s.t.} \sum_{i=1}^{N} a_i^u[t] \leq \alpha N \quad \forall t. \tag{3}$$

where the belief value $\pi_i[t]$ evolves according to rule (1) based on the scheduling decision $a_i^u[t]$ under policy u. Such an objective is standard in literature for Markov decision processes under the long-term average reward criteria (e.g., [26]). Noting that since the scheduling decisions are made based on incomplete knowledge of channel states, this problem is a partially observable Markov decision process (POMDP) [25].

This problem is in fact an example of Restless Multiarmed Bandit Problem [19]. For a general RMBP, finding an optimal solution is PSPACE-hard [27]. However, for the downlink scheduling problem at hand, a low-complexity Whittle's Index Policy was proposed in [20] and [21] based on the RMBP theory that inherently exploits the channel memory when making scheduling decisions. For detailed descriptions of general RMBP and Whittle's Index Policy for downlink scheduling, please refer to [19]–[21].

For the downlink scheduling problem, we note that there is only limited analytical characterization of Whittle's Index Policy, which is restricted in perfectly symmetric scenarios where Whittle's Index Policy takes a special round-robin form [20]. In asymmetric cases, however, the scheduling decision no longer takes the form of round-robin, bringing sophisticated complications in belief value evolutions that are tightly coupled among channels, which significantly complicates the analysis. The main focus of this paper is to analytically characterize the performance of Whittle's Index Policy in the asymmetric case with two classes of channels.

III. Upper Bound on Achievable Throughput

We begin our analysis by characterizing an upper bound to the throughput performance of all feasible downlink scheduling policies that satisfies the constraint (3). The upper bound is obtained from a fictitious policy that is optimal for the downlink scheduling problem under a *relaxed constraint*.

Note here that such relaxation is also a crucial step in the study of the general RMBP problem. Yet, our analysis, being specific to the downlink scheduling problem, has its novelties, as we shall remark on later.

A. Average-Constrained Relaxed Scheduling Problem

We consider an associated relaxed problem of (2)–(3) that only requires an *average number* of users to be activated in the long run, defined as follows:

$$\max_{u\in U}\ \liminf_{T\to\infty}\frac{1}{T}E\left[\sum_{t=0}^{T-1}\sum_{i=1}^{N}\pi_i[t]\cdot a_i^u[t]\mid\boldsymbol{\vec{\pi}}[0]\right] \tag{4}$$

$$\text{s.t. }\limsup_{T\to\infty}\frac{1}{T}E\left[\sum_{t=0}^{T-1}\sum_{i=1}^{N}a_i^u[t]\right]\le\alpha N. \tag{5}$$

Note that, contrary to the stringent constraint (3), the relaxed constraint (5) allows the activation of more than α fraction of users in each time-slot, provided the long-term average fraction does not exceed α. Hence, the optimal policy under this relaxed constraint, which we shall identify next, provides a throughput upper bound to any policy that satisfies the stringent constraint.

B. Optimal Policy for the Relaxed Problem

We remark that the relaxed problem is also an important component of Whittle's analysis of general RMBPs [19], in which an optimal policy for the relaxed problem is developed based on the *Whittle's index values*. Following the approach of classic RMBP framework [19], in our downlink scenario, we identify an optimal policy for the relaxed problem based on Whittle's indices.

Specifically, for channels in class k, the Whittle's index value $W_k(\pi)$ is assigned to each belief state $\pi\in\mathcal{B}_k$. These index values intuitively capture the exploitation and exploration value to be gained from scheduling the associated channel when its belief value is π. This characteristic of $W_k(\pi)$ is also illustrated in Section VII-B via numerical investigations. The index value function is expressed in closed form as

$$W_k(\pi)=\begin{cases}\frac{(b_{0,l}^k-b_{0,l+1}^k)(l+1)+b_{0,l+1}^k}{1-p_k+(b_{0,l}^k-b_{0,l+1}^k)l+b_{0,l+1}^k}, & \text{if } r_k\le\pi=b_{0,l}^k<b_s^k\\ \frac{r_k}{(1-p_k)(1+r_k-p_k)+r_k}, & \text{if } b_s^k\le\pi\le p_k.\end{cases} \tag{6}$$

Note that the above expression is a modified version of the expression in [20]. Details of the derivation can be found in [28].

The following two characteristics they possess are primarily significant for our analysis.

- $W_k(\pi)$ monotonically increases with $\pi\in\mathcal{B}_k$.
- $W_k(\pi)\in[0,1]$ for all $\pi\in\mathcal{B}_k$.

The next lemma identifies an index-based policy with *appropriate randomization* that is optimal for the relaxed constraint problem. This policy schedules each user based on its own belief value, independently from other users. The proof of the lemma can be found in [20].

Lemma 1: For the problem under relaxed constraint, there exists an optimal stationary policy ϕ^*, parameterized by the threshold ω^* and a randomization parameter $\rho^*\in(0,1]$, such that we have the following.

i) Channel i in class k is scheduled if $W_k(\pi_i[t])>\omega^*$, and stays idle if $W_k(\pi_i[t])<\omega^*$. If $W_k(\pi_i[t])=\omega^*$, it is scheduled with probability ρ^*.

ii) The parameters ω^* and ρ^* are such that, under policy ϕ^*, the relaxed constraint (5) is strictly satisfied with equality.

From now on, we shall denote ϕ^* as the "*Optimal Relaxed Policy*." For technical purposes, we henceforth assume α is such that $\rho^*\neq 1$. Since each α value maps to a unique (ω^*,ρ^*) pair [29], only countably many α values correspond to $\rho^*=1$, i.e., achieved by deterministic policies. Therefore, the set of $\alpha\in(0,1]$ for which $\rho^*\neq 1$ has Lebesgue measure one.

C. Steady-State Distribution of Belief Values

We next present the transition structure of the belief values under Optimal Relaxed Policy, captured in the following lemma. The structure will be critical in the development of our subsequent main results.

Lemma 2: For each channel in class k, under the Optimal Relaxed Policy, the structure of belief value evolution depends on the threshold ω^* of policy.

i) If $\omega^*<W_k(b_s^k)$, then the belief value evolution of each class-k channels is positive recurrent with a finite recurrent class.

ii) If $\omega^*\ge W_k(b_s^k)$, the belief value evolution is transient. With probability 1, ultimately no channel in class k will transmit.

Proof: The proof of this lemma follows from the monotonic structure of belief evolution, as shown in Fig. 2. Details are included in Appendix A. ■

Thus, if $\omega^*\ge\max\{W_1(b_s^1),W_2(b_s^2)\}$, the above analysis reveals that ultimately no user transits, corresponding to the trivial case of $\alpha N=0$. Also, if ω^* is between $W_1(b_s^1)$ and $W_2(b_s^2)$, the class with the smaller $W_k(b_s^k)$ will eventually transit into a passive mode, hence reducing the system to a well-understood scenario with a single class of channels [15], [16]. Thus, here we focus on the heterogeneous case of $\omega^*<W_k(b_s^k),k=1,2$, where the steady-state belief value distribution exists for both classes under the Optimal Relaxed Policy.

D. Upper Bound on Achievable Throughput

The throughput performance of Optimal Relaxed Policy provides an throughput upper bound for all policies under the stringent constraint. The value of such an upper bound clearly depends on the number of users in each class $\gamma_k N$, $k=1,2$, as well as the fraction α of users allowed for activation. Denoting $\boldsymbol{\gamma}=[\gamma_1,\gamma_2]$, we represent the time average expected throughput of the Optimal Relaxed Policy as $v^N(\boldsymbol{\gamma},\alpha)$. The following lemma states that, as long as $\boldsymbol{\gamma}$ and α are given, the *per-user* throughput (i.e., $v^N(\boldsymbol{\gamma},\alpha)/N$) is independent of N.

Lemma 3: Given $\boldsymbol{\gamma}$ and α, $\frac{v^N(\boldsymbol{\gamma},\alpha)}{N}$ is independent of N, denoted henceforth as $r(\boldsymbol{\gamma},\alpha)$.

Proof: The proof follows from showing that, when the number of users N grows, as long as the proportion of each class of channels stays the same and the fraction α of users activated does not change, the form of Optimal Relaxed Policy does not change. Since each user is scheduled independently, the throughput $v^N(\boldsymbol{\gamma},\alpha)$ is proportional to N, establishing the lemma. Details are provided in Appendix B. ■

We hence refer to the $(\boldsymbol{\gamma}, \alpha)$ pair as "*system parameters*." Therefore $Nr(\boldsymbol{\gamma}, \alpha)$ provides a throughput upper bound to any policy in the same system under the stringent constraint (3). Equivalently, $r(\boldsymbol{\gamma}, \alpha)$ provides a per-user throughput performance upper bound to all policies that satisfies the stringent constraint.

We next describe Whittle's Index Policy for the strictly constrained problem (2)–(3), and later study the closeness of its performance to the upper bound established here.

IV. Whittle's Index Policy Description

In this section, we formally introduce Whittle's Index Policy for solving the stringently constrained downlink scheduling problem (2)–(3).

A. Whittle's Index Policy

The Optimal Relaxed Policy, along with the Whittle's index values, gives consistent ordering of belief values with respective to the indices. For instance, under the Optimal Relaxed Policy, if it is optimal to schedule one channel, it is then optimal to transmit to other channels with higher index values. Thus, the Whittle's index value gives an intuitive order of how attractive the channel is for scheduling. This intuition leads to Whittle's Index Policy [20] under the stringent constraint on the maximum number of channels that can be scheduled.

Whittle's Index Policy: *At the beginning of each time-slot, the channel i in class k is scheduled if its Whittle's index value $W_k(\pi_i)$ is within the top αN index values of all channels in that slot, with arbitrary tie-breaking while assuring a total αN channels being scheduled.*

Whittle's Index Policy is attractive because it has very low complexity, and it was observed via numerical investigations to yield significant throughput performance gains over the scheduling strategies that does not utilize channel memory [21]. The main focus of our work is to analytically understand the approximate or asymptotic optimality of Whittle's Index Policy in asymmetric scenarios.

B. Whittle's Index Policy Over Truncated State Space

Recall from Section II that the belief values evolve over a countable state space, and also note that if a channel is not scheduled for a long time, its belief value will get arbitrarily close to its stationary belief value. This motivates us to consider a truncated version of the belief value evolution whereby the belief value is set to its steady state if the corresponding channel is not scheduled for a large number, say τ, slots. This mild assumption facilitates more tractable performance analysis of the policy. Thus, if a class-k user is not scheduled for τ time-slots, its channel state history is entirely forgotten and its belief value will transit to the stationary belief value b_s^k, where the truncation τ is assumed to be very large.

Whittle's Index Policy is then implemented over the truncated belief state, which differs from the nontruncated case merely in the truncated belief value evolution. We believe that the truncated scenario can provide arbitrarily close approximation to the original system when τ is large. More importantly, as we shall see in Sections V and VI, Whittle's Index Policy, implemented over the truncated belief state space, achieves asymptotically optimal performance as long as the truncation is sufficiently large.

V. Local Optimality of Whittle's Index Policy

In this section, we study the optimality properties of Whittle's Index Policy for downlink scheduling, over a large truncated belief space. This result forms the basis for the subsequent global optimality result in Section VI. We start by introducing a state space over which the local optimality will be established.

A. System State Vector

We define the *system state* $\boldsymbol{Z}^N$ as a vector that represents the proportion of channels in each belief value, over the truncated space when the total number of users is N, i.e., $\boldsymbol{Z}^N = [\boldsymbol{Z}^{1,N}, \boldsymbol{Z}^{2,N}]$, with

$$\boldsymbol{Z}^{k,N} = [Z_{0,1}^{k,N}, \ldots, Z_{0,\tau}^{k,N}, Z_s^{k,N}, Z_{1,\tau}^{k,N}, \ldots, Z_{1,1}^{k,N}], \quad k = 1, 2$$

where $Z_{c,l}^{k,N}$ and $Z_s^{k,N}$ respectively denote the *proportion* of channels in the corresponding belief state $b_{c,l}^k$ and b_s^k, with respect to the total number of users N. Hence, each element of $\boldsymbol{Z}^N$ is a multiple of $1/N$ so that $\boldsymbol{Z}^N$ takes values in a lattice with mesh size $1/N$. Noting that the total number of users in each class does not change over time, for any N the system state $\boldsymbol{Z}^N[t] \in \mathcal{Z}$ where

$$\mathcal{Z} := \left\{ \boldsymbol{Z}^N \geq 0 : Z_s^{k,N} + \sum_{c,l} Z_{c,l}^{k,N} = \gamma_k, k = 1, 2 \right\}. \quad (7)$$

The system state vector $\boldsymbol{Z}^N[t]$ does not distinguish users with the same belief state, thus its dimension will not scale with N. Therefore, compared to $\vec{\boldsymbol{\pi}}[t]$, it provides a more convenient representation of the system belief state. Furthermore, $\boldsymbol{Z}^N[t]$ fully determines the instantaneous throughput gain in slot t under both Whittle's Index Policy and the Optimal Relaxed Policy (introduced in Lemma 1) because the instantaneous throughput gains under both policies are only determined by the distribution of the channels with different belief values, not their identities.

From Lemma 2 and the subsequent remarks, under the operation of the Optimal Relaxed Policy, the belief state evolution of each channel is positive recurrent with a steady-state distribution. The following lemma also establishes the independence of this steady-state distribution from N and defines a useful parameter for future use.

Lemma 4: Given the system parameters $(\boldsymbol{\gamma}, \alpha)$, the system state vector $\boldsymbol{Z}^N[t]$ under the Optimal Relaxed Policy converges in distribution to a random vector, denoted as $\boldsymbol{Z}^N[\infty]$. The mean of $\boldsymbol{Z}^N[\infty]$ is independent of N and is denoted as

$$\vec{\boldsymbol{\zeta}}_{\boldsymbol{\gamma}}^{\alpha} := E[\boldsymbol{Z}^N[\infty]].$$

Proof: This lemma follows from a similar principle to the one we established in Lemma 3. For details, please refer Appendix C. ■

It is easy to see that $\vec{\boldsymbol{\zeta}}_{\boldsymbol{\gamma}}^{\alpha} \in \mathcal{Z}$ and the form of $\vec{\boldsymbol{\zeta}}_{\boldsymbol{\gamma}}^{\alpha}$ fully determines the time average throughput of the Optimal Relaxed Policy. Therefore, the vector $\vec{\boldsymbol{\zeta}}_{\boldsymbol{\gamma}}^{\alpha}$ provides an important benchmark for our asymptotic analysis. If, in the long run under Whittle's Index Policy, the system state $\boldsymbol{Z}^N[t]$ stays close to $\vec{\boldsymbol{\zeta}}_{\boldsymbol{\gamma}}^{\alpha}$, it indicates that

Whittle's Index Policy will have throughput performance close to that of the Optimal Relaxed Policy—the throughput upper bound. To capture the closeness, we define the δ neighborhood of $\vec{\boldsymbol{\zeta}}^{\alpha}_{\boldsymbol{\gamma}}$ as

$$\Omega_{\delta}(\vec{\boldsymbol{\zeta}}^{\alpha}_{\boldsymbol{\gamma}}) = \{\boldsymbol{Z} \in \mathcal{Z} : \|\boldsymbol{Z} - \vec{\boldsymbol{\zeta}}^{\alpha}_{\boldsymbol{\gamma}}\| \leq \delta\}, \tag{8}$$

for $\delta > 0$, where $\|\cdot\|$ stands for Euclidean distance. We are now ready to state and prove our first main result regarding a form of local optimality of Whittle's Index Policy.

B. Local Optimality of Whittle's Index Policy

Under the system parameters $(\boldsymbol{\gamma}, \alpha)$, we let $R^N_T(\boldsymbol{\gamma}, \alpha, \boldsymbol{x})$ represent the time average throughput obtained over the time duration $0 \leq t < T$ under Whittle's Index Policy, conditioned on the initial system state $\boldsymbol{Z}^N[0] = \boldsymbol{x}$, i.e.,

$$R^N_T(\boldsymbol{\gamma}, \alpha, \boldsymbol{x}) := \frac{1}{T} E\left[\sum_{t=0}^{T-1} \sum_{i=1}^{N} \pi_i[t] a_i^{ind}[t] \mid \boldsymbol{Z}^N[0] = \boldsymbol{x}\right]$$

where $(a_i^{ind}[t])_i$ denotes the scheduling decision vector made by Whittle's Index Policy at time t.

Recall from Lemma 3 that $r(\boldsymbol{\gamma}, \alpha)$ denotes the per-user throughput under the Optimal Relaxed Policy, which serves as an upper bound on Whittle's Index Policy performance. The next proposition characterizes the local convergence property of Whittle's Index Policy performance to $r(\boldsymbol{\gamma}, \alpha)$.

Proposition 1: Under the system parameters $(\boldsymbol{\gamma}, \alpha)$, there exists a $\delta > 0$ neighborhood $\Omega_{\delta}(\vec{\boldsymbol{\zeta}}^{\alpha}_{\boldsymbol{\gamma}})$ of $\vec{\boldsymbol{\zeta}}^{\alpha}_{\boldsymbol{\gamma}}$ such that, if the initial system state $\boldsymbol{x}$ is within $\Omega_{\delta}(\vec{\boldsymbol{\zeta}}^{\alpha}_{\boldsymbol{\gamma}})$, then

$$\lim_{T\to\infty} \lim_{m\to\infty} \frac{R^{N_m}_T(\boldsymbol{\gamma}, \alpha, \boldsymbol{x})}{N_m} = r(\boldsymbol{\gamma}, \alpha).$$

where $\{N_m\}_m$ is any increasing sequence of positive integers with $\alpha N_m, \gamma_k N_m \in \mathbb{Z}^+$, for $k = 1, 2$ and all m.

Proof Outline: Here, we give a high-level description of the proof for an intuitive understanding, and refer the reader to [36] for the rigorous derivation.

- We start by defining a fluid approximation, whereby the discrete-time evolution of $\boldsymbol{Z}^N[t]$ under Whittle's Index Policy is modeled as a deterministic vector $\boldsymbol{z}[t] \in \mathcal{Z}$ that evolves in discrete time over $\mathcal{Z}$ and is independent of N. Under this fluid approximation, the users are no longer unsplittable entities so that the state space of $\boldsymbol{z}[t]$ is no longer restricted to a lattice as it was for $\boldsymbol{Z}^N[t]$. Also, the fluid approximation $\boldsymbol{z}[t]$ evolves in a deterministic manner, in contrast to the stochastic transition of $\boldsymbol{Z}^N[t]$. The evolution of $\boldsymbol{z}[t]$ is defined by a difference equation as a function of the *expected* state change of $\boldsymbol{Z}^N[t]$ under Whittle's Index Policy as follows:

$$\boldsymbol{z}[t+1] - \boldsymbol{z}[t] \,|_{\boldsymbol{z}[t]=\boldsymbol{z}} = E\left[\boldsymbol{Z}^N[t+1] - \boldsymbol{Z}^N[t] \mid \boldsymbol{Z}^N[t] = \boldsymbol{z}\right] \tag{9}$$

where N is any integer for which $\boldsymbol{z}$ is a feasible state.

- We then establish local convergence of the fluid approximation model when $\boldsymbol{z}[0]$ is within a small enough δ neighborhood $\Omega_{\delta}(\vec{\boldsymbol{\zeta}}^{\alpha}_{\boldsymbol{\gamma}})$ of $\vec{\boldsymbol{\zeta}}^{\alpha}_{\boldsymbol{\gamma}}$. We show the convergence by first noting that the differential equation (9) is linear within a wider convex region than $\Omega_{\delta}(\vec{\boldsymbol{\zeta}}^{\alpha}_{\boldsymbol{\gamma}})$. Within this region, we obtain a closed-form expression of the right-hand side of (9), which enables us to investigate the eigenvalue structure of the linear differential equation. We show that each eigenvalue λ satisfies $|\lambda| < 1$ and apply standard linear system theory to establish the local convergence.
- We then connect the fluid approximation model $\boldsymbol{z}[t]$ to the discrete-time stochastic system state $\boldsymbol{Z}^N[t]$ by using a discrete-time extension of Kurtz's Theorem, which can be interpreted as an extension of the strong law of large numbers to random processes (see [30]). Essentially, it states that, over any finite time duration $[0, T]$, the actual system evolution $\boldsymbol{Z}^N[t]$ can be made arbitrarily close to the above fluid approximation $\boldsymbol{z}[t]$ by increasing the number of channels N sufficiently, with exponential convergence rate.
- The previous convergence result, together with the local convergence result of the fluid evolution $\boldsymbol{z}[t]$ to $\vec{\boldsymbol{\zeta}}^{\alpha}_{\boldsymbol{\gamma}}$, enables us to establish the local convergence of the system state $\boldsymbol{Z}^N[t]$ to $\vec{\boldsymbol{\zeta}}^{\alpha}_{\boldsymbol{\gamma}}$ as the number of users N grows, provided that the initial state $\boldsymbol{Z}^N[0] \in \Omega_{\delta}(\vec{\boldsymbol{\zeta}}^{\alpha}_{\boldsymbol{\gamma}})$. Hence, the system state under Whittle's Index Policy will stay close (in a probabilistic sense) to the expectation $\vec{\boldsymbol{\zeta}}^{\alpha}_{\boldsymbol{\gamma}}$ of the system state under the Optimal Relaxed Policy, which, in turn, indicates that the throughput performance of Whittle's Index Policy will approach the throughput upper bound $r(\boldsymbol{\gamma}, \alpha)$, as expressed in the proposition.

We again emphasize that the technical details of the outlined steps are fairly intricate and are moved to [36]. ■

Proposition 1 illustrates an interesting local optimality property of Whittle's Index Policy as the number of users N and the time horizon T increases while the system parameters $(\boldsymbol{\gamma}, \alpha)$ stay the same. It indicates that, under Whittle's Index Policy, as long as the initial state $\boldsymbol{Z}^N[0]$ is close enough to $\vec{\boldsymbol{\zeta}}^{\alpha}_{\gamma}$, the average per-user throughput over any finite time duration will get arbitrarily close to the Optimal Relaxed Policy performance as the number of users scales.

Remark: We note that the sequence $\{N_m\}_m$ is used to guarantee that the number of channels in each class, as well as the number of scheduled users, take integer values. In fact, our result can be generalized to all N by slightly perturbing $\boldsymbol{\gamma}$ and α as a function of N but assuring their limits are well defined.

Note that we have assumed the model of two-classes of channels. Future research direction includes generalization to multiple-class scenarios or models where users have arbitrary transition probabilities. The main challenge in generalizing to such setup is to analyze the eigenvalue structure of the system state's transition matrix (e.g., [36, Lemma 11]) since analytically studying the form of the eigenvalues can be difficult when there are multiple classes of users.

VI. Global Optimality of Whittle's Index Policy

The above local optimality result heavily relies on the initial state $\boldsymbol{Z}^N[0]$ being close to $\vec{\boldsymbol{\zeta}}^{\alpha}_{\boldsymbol{\gamma}}$, which is difficult to guarantee. In this section, we study the global optimality of the infinite horizon throughput performance of Whittle's Index Policy starting from any initial state. We begin our analysis by presenting the recurrence structure of the system state.

Lemma 5: Under system parameters $(\boldsymbol{\gamma}, \alpha)$, for any $\epsilon > 0$, if the number of users N is large enough, we have the following.

i) The system state $\boldsymbol{Z}^N[t]$ evolves as an aperiodic Markov chain, in a state space that contains only one recurrent class.

ii) There exists at least one recurrent state within the ϵ neighborhood $\Omega_\epsilon(\vec{\boldsymbol{\zeta}}^\alpha_{\boldsymbol{\gamma}})$ of $\vec{\boldsymbol{\zeta}}^\alpha_{\boldsymbol{\gamma}}$.

Proof: We prove this lemma by constructing probability paths from any state to the neighborhood $\Omega_\epsilon(\vec{\boldsymbol{\zeta}}^\alpha_{\boldsymbol{\gamma}})$. Details of the proof are included in [36]. ■

This lemma states that $\boldsymbol{Z}^N[t]$ will ultimately enter any small neighborhood of $\vec{\boldsymbol{\zeta}}^\alpha_{\boldsymbol{\gamma}}$ when N is large enough. Together with Proposition 1, this result shows promise for establishing the global asymptotic optimality of Whittle's Index Policy. This is plausible because once $\boldsymbol{Z}^N[t]$ enters $\Omega_\delta(\vec{\boldsymbol{\zeta}}^\alpha_{\boldsymbol{\gamma}})$, the performance of Whittle's Index Policy *afterwards* can get very close to its upper bound as N scales, as established in Proposition 1. However, since we consider the infinite horizon time average throughput, this argument would break down if the time it takes for $\boldsymbol{Z}^N[t]$ to enter $\Omega_\delta(\vec{\boldsymbol{\zeta}}^\alpha_{\boldsymbol{\gamma}})$ also scales up with N. This observation motivates us to introduce a useful assumption, which will later be justified (in Section VII-A) via numerical studies.

Assumption Ψ: For each $\epsilon > 0$, let $\Gamma^N_{\boldsymbol{x}}(\epsilon)$ represent the first time of reaching $\Omega_\epsilon(\vec{\boldsymbol{\zeta}}^\alpha_{\boldsymbol{\gamma}})$ starting from $\boldsymbol{Z}^N[0] = \boldsymbol{x}$, i.e.,

$$\Gamma^N_{\boldsymbol{x}}(\epsilon) = \min\{t : \boldsymbol{Z}^N[t] \in \Omega_\epsilon(\vec{\boldsymbol{\zeta}}^\alpha_{\boldsymbol{\gamma}}) \mid \boldsymbol{Z}^N[0] = \boldsymbol{x}\}.$$

Then, we assume that the expected time of reaching $\Omega_\epsilon(\vec{\boldsymbol{\zeta}}^\alpha_{\boldsymbol{\gamma}})$ is bounded by a constant $M_\epsilon < \infty$, i.e.,

$$E\left[\Gamma^N_{\boldsymbol{x}}(\epsilon)\right] \le M_\epsilon$$

for all $\boldsymbol{x}$ and large enough N.

Since for each N, $\boldsymbol{Z}^N[t]$ under Whittle's Index Policy is recurrent and aperiodic with a finite state space, there exists a steady-state distribution associated with $\boldsymbol{Z}^N[t]$. As before, we use $\boldsymbol{Z}^N[\infty]$ to denote the associated limiting random vector. The next lemma establishes that, under Assumption Ψ, the distribution of $\boldsymbol{Z}^N[\infty]$ approaches a point-mass at $\vec{\boldsymbol{\zeta}}^\alpha_{\boldsymbol{\gamma}}$ as N scales. Here, again, the sequence $\{N_m\}_m$ is defined in the same way as in Proposition 1.

Lemma 6: Under Assumption Ψ and system parameters $(\boldsymbol{\gamma}, \alpha)$, for any $\epsilon > 0$, the steady-state probability of $\boldsymbol{Z}^N[t]$ under Whittle's Index Policy satisfies

$$\lim_{m\to\infty} P\left(\boldsymbol{Z}^{N_m}[\infty] \in \Omega_\epsilon(\vec{\boldsymbol{\zeta}}^\alpha_{\boldsymbol{\gamma}})\right) = 1.$$

Proof: The proof utilizes [30, Theorem 6.89], which builds on the following arguments.

Note that $\epsilon > 0$ can be selected to be small enough for the following argument. As depicted in Fig. 3, we let T_ϵ be a random variable denoting, in steady state, the time duration between *consecutive* hitting times into the neighborhood $\Omega_\epsilon(\vec{\boldsymbol{\zeta}}^\alpha_{\boldsymbol{\gamma}})$ from outside of the neighborhood. Let T^0_ϵ denote the time duration from the time $\boldsymbol{Z}^N[t]$ enters the neighborhood $\Omega_\epsilon(\vec{\boldsymbol{\zeta}}^\alpha_{\boldsymbol{\gamma}})$ from outside until the time it leaves. Hence, the expected proportion of time that $\boldsymbol{Z}^N[t]$ stays outside this neighborhood is $(E[T_\epsilon] - E[T^0_\epsilon])/E[T_\epsilon]$.

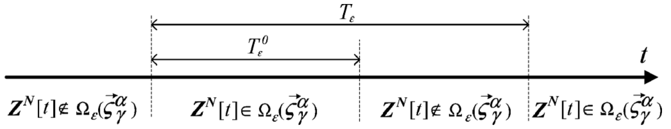
T_ε
T^0_ε
t
$\boldsymbol{Z}^N[t] \notin \Omega_\varepsilon(\vec{\varsigma}^\alpha_\gamma)$
$\boldsymbol{Z}^N[t] \in \Omega_\varepsilon(\vec{\varsigma}^\alpha_\gamma)$
$\boldsymbol{Z}^N[t] \notin \Omega_\varepsilon(\vec{\varsigma}^\alpha_\gamma)$
$\boldsymbol{Z}^N[t] \in \Omega_\varepsilon(\vec{\varsigma}^\alpha_\gamma)$

Fig. 3. Transition behavior of $\boldsymbol{Z}^N[t]$ in steady state.

We know that the numerator $E[T_\epsilon] - E[T^0_\epsilon]$ is uniformly bounded for all N due to Assumption Ψ. However, as N increases, it is more likely for $\boldsymbol{Z}^N[t]$ to stay within the neighborhood for a long time before exiting it (based on the convergence of fluid approximation model and Kurtz's Theorem in the proof of Proposition 1). Thus, $E[T^0_\epsilon]$, and hence the denominator $E[T_\epsilon]$, grow to infinity as N scales. Therefore, the expected proportion of time spent outside $\Omega_\epsilon(\vec{\boldsymbol{\zeta}}^\alpha_{\boldsymbol{\gamma}})$ vanishes as N scales up, which leads to the statement of the lemma. Details of the proof can be found in [36]. ■

Under Whittle's Index Policy with system parameters $(\boldsymbol{\gamma}, \alpha)$, we let $R^N_{\boldsymbol{x}}(\boldsymbol{\gamma}, \alpha)$ be the achieved infinite horizon, time average throughput, conditioned on the initial system state $\boldsymbol{Z}^N[0] = \boldsymbol{x}$, i.e.,

$$R^N_{\boldsymbol{x}}(\boldsymbol{\gamma}, \alpha) := \lim_{T\to\infty} \frac{1}{T} E\left[\sum_{t=0}^{T-1}\sum_{i=1}^{N} \pi_i[t] a_i^{ind}[t] \mid \boldsymbol{Z}^N[0] = \boldsymbol{x}\right].$$

From Lemma 6, we know that, in steady state, the system state $\boldsymbol{Z}^{N_m}[\infty]$ is increasingly concentrated around $\vec{\boldsymbol{\zeta}}^\alpha_{\boldsymbol{\gamma}}$ as m increases, regardless of the initial state $\boldsymbol{x}$. We build on this to establish the global asymptotical optimality of Whittle's Index Policy.

Proposition 2: Under Assumption $\boldsymbol{\Psi}$, for any initial system state $\boldsymbol{x}$, we have

$$\lim_{m\to\infty} \frac{R^{N_m}_{\boldsymbol{x}}(\boldsymbol{\gamma}, \alpha)}{N_m} = r(\boldsymbol{\gamma}, \alpha).$$

Since $r(\boldsymbol{\gamma}, \alpha)$ is an upper bound on the maximum achievable per-user throughput by any policy, this implies that Whittle's Index Policy is optimal in the many-user regime.

Proof: We prove this result by decomposing $R^N_{\boldsymbol{x}}(\boldsymbol{\gamma}, \alpha)$ as a summation of the expected throughput conditioned on whether the system state is within or outside an arbitrarily small ϵ neighborhood of $\vec{\boldsymbol{\zeta}}^\alpha_{\boldsymbol{\gamma}}$. Since the latter has diminishing probability according to Lemma 6, the expected throughput of Whittle's Index Policy can get arbitrarily close to that of Optimal Relaxed Policy. Details of the proof are provided in [36]. ■

Remarks:

1) We would like to emphasize that the global optimality result is not a straightforward extension of the local convergence result by contrasting Propositions 1 and 2. Note that in Proposition 1, the time limit is outside the limit of the number of users N, where each convergence (with N) is with respective to a *fixed time duration*. However, the order of limit is switched in the global optimality result of Proposition 2, as it states the convergence with N *the infinite horizon* average throughput, which is much stronger and hence is much more challenging to prove.

2) We would like to contrast Assumption Ψ with Weber's condition [22]. For general RMBP problem, Weber's

condition leads to the same global asymptotic optimality result. While confirming Weber's condition may be possible in very low-dimensional problems, in our downlink scheduling problem, this requires one to rule out the existence of both closed orbits and chaotic behavior of a high-dimensional nonlinear differential equation, which is extremely difficult to check—even numerically. Assumption Ψ, on the other hand, takes a much simpler form, as it is defined over the actual stochastic system and is amenable to easy numerical verification, as is performed in Section VII-A.

VII. Numerical Results

A. *Verification and Interpretation of Assumption* Ψ

We start by numerically verifying Assumption Ψ. We consider the asymmetric scenario with two classes of channels with system parameters $\gamma = [0.45, 0.55]$, $\alpha = 0.6$, with $p_1 = 0.9$, $r_1 = 0.45$, $p_2 = 0.8$, $r_2 = 0.3$.

We next examine the change of the average hitting time $\Gamma_{\boldsymbol{x}}^N(\epsilon)$, while maintaining α and $\boldsymbol{\gamma}$.

We let $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{Z}$ be initial values of $\boldsymbol{Z}^N[0]$ that are selected to be two extreme points in the state space to exhibit the uniformity of $\Gamma_{\boldsymbol{x}}^N(\epsilon)$ to the initial state. Specifically, state $\boldsymbol{x}$ corresponds to the case when all the users have just observed their channels to be in OFF state, i.e., with belief value $b_{0,1}^k$, $k = 1, 2$. And $\boldsymbol{y}$ corresponds to the case when all users have no initial observation of their channels state history, i.e., with belief value b_s^k, $k = 1, 2$.

We examine the average value of hitting time $\Gamma_{\boldsymbol{x}}^N(\epsilon)$ and $\Gamma_{\boldsymbol{y}}^N(\epsilon)$ with a very small neighborhood $\epsilon = 0.005$, when the number of users N grows from 10×10^3 to 500×10^3. As indicated in Fig. 4, for both cases, the average time of hitting the ϵ neighborhood first decreases with N, and then *converges* and stays almost the same as N scales up. This is especially intriguing. The rationale behind this phenomenon is as follows. Under Whittle's Index Policy, a total number of αN users are activated at each time-slot. Therefore, for relatively small number of users, the amount of probabilistic belief state transitions, as well as the amount of system states in the neighborhood, increases with N, leading to a higher chance of hitting the desired neighborhood $\Omega_\epsilon(\vec{\boldsymbol{\zeta}}_{\boldsymbol{\gamma}}^{\alpha})$ and smaller value of hitting time. However, the belief update of each user contributes to the $1/N$ change of the system state $\boldsymbol{Z}^N[t]$, which decreases with N. Therefore, as N further increases, the *total amount of transitions* of the system state $\boldsymbol{Z}^N[t]$ due to channel state feedback is roughly $\alpha N \cdot 1/N = \alpha$, which is invariant of N. This result, along with many other numerical experiments we have conducted that lead to the same observation [36], verifies Assumption Ψ.

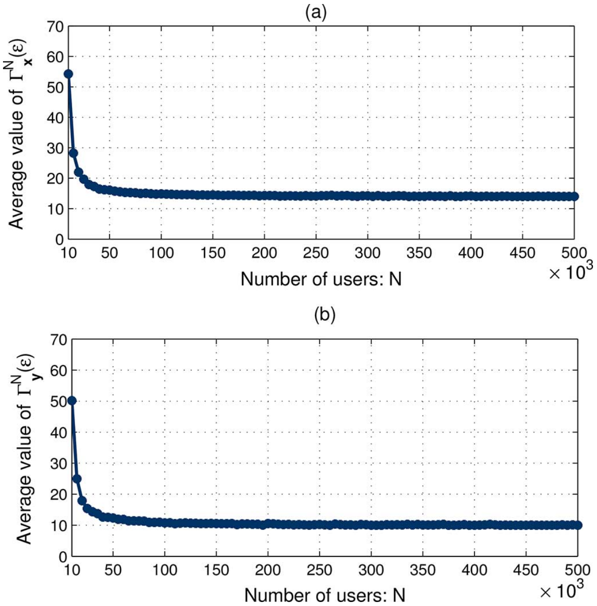

Fig. 4. Average time of hitting $\Omega_\epsilon(\vec{\boldsymbol{\zeta}}_{\boldsymbol{\gamma}}^{\alpha})$. (a) $\boldsymbol{Z}^N[0] = \boldsymbol{x}$. (b) $\boldsymbol{Z}^N[0] = \boldsymbol{y}$.

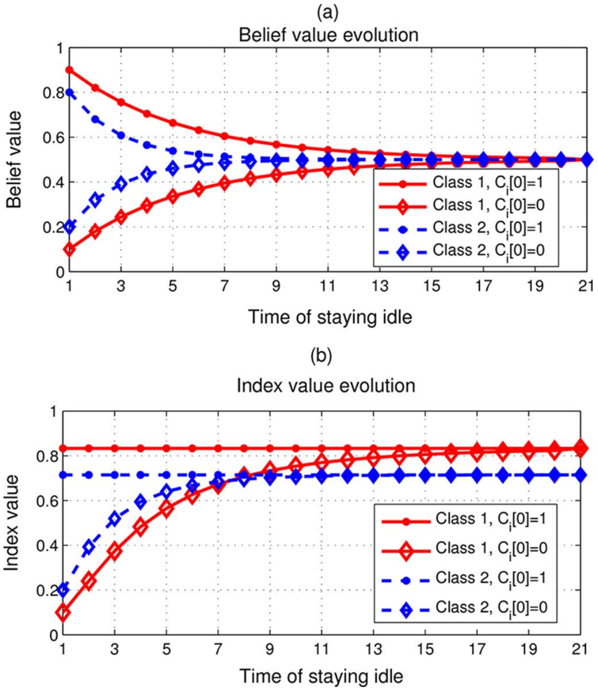

Fig. 5. Evolution of (a) belief value and (b) Whittle's index value.

B. *"Exploitation Versus Exploration" Tradeoff*

In this section, we demonstrate how the Whittle's index value captures the "exploitation versus exploration" tradeoff for our *asymmetric downlink scheduling problem*.

Consider two classes of ON/OFF fading channels with belief value evolutions plotted in Fig. 5(a). Note that both classes have the same stationary distribution $b_s^k = 0.5$, $k \in \{1, 2\}$ of being at ON state, but channels in class 1 have a higher degree of time correlation, i.e., fade slower, than channels in class 2 since $p_1 > p_2$ and $r_1 < r_2$. The corresponding Whittle index values of the two classes of channels are depicted in Fig. 5(b) as functions of the updated belief value starting from different initial states.

To understand the nature of Whittle's index value, we first consider the case when the channels in both classes are observed to be ON at time 0 and stay passive since then. As indicated in Fig. 5(a), the class-1 channel has a higher belief value than the class-2 channel, hence scheduling the class-1 channel gives a higher immediate throughput than scheduling the class-2 channel. Moreover, once a class-1 channel is scheduled, it is more likely to stay in ON state again, bringing high future gains.

Accordingly, the index values in Fig. 5(b) when both state evolutions start from ON states capture that it is more attractive to schedule the class-1 channel because of the advantage in both exploitation and exploration.

On the other hand, when the scheduler has observed channels in both classes to be OFF at time 0, Fig. 5(a) shows that the class 2 channel has a higher belief value than the class-1 channel. However, although the Whittle's index value in Fig. 5(b) of class-2 channel is initially smaller than that of class-1 channel, after a certain amount of delay (around 8 slots in the figure) this order is switched, which is interpreted as follows: Initially, since the class-1 channel has smaller belief value than that of the class-2 channel, it is more attractive to exploit the immediate gain brought by the class-2 channel. However, as the passive time grows, as indicated in Fig. 5(a), the difference between immediate gain of both classes diminishes. Then, it becomes more attractive to explore the class-1 channel because its longer memory can bring higher future gains if it turns out to be in ON state.

This investigation reveals the intricate nature of Whittle's index value in capturing the fundamental "exploration versus exploitation" tradeoff. In our scheduling problem with asymmetric channel statistics, such a property of Whittle's Index Policy turns out to be crucial in *achieving asymptotically optimal performance*.

C. Performance Evaluation and Comparison

Note that our results focus on asymptotic regime when the number of users scales up. We next numerically evaluate the performance of the Whittle's Index Policy under finite number of users. We next consider a system where $\boldsymbol{\gamma} = [0.6, 0.4]$, $\alpha = 0.3$, $(p_1, r_1) = (0.75, 0.2)$, and $(p_2, r_2) = (0.8, 0.3)$, and evaluate the value $R_{\boldsymbol{x}}^N(\boldsymbol{\gamma}, \alpha)/N$ when N increases as multiples of 5, i.e., $N = 5m, m = 1, 2, \ldots$. Fig. 6(a) and (b) respectively corresponds to the aforementioned extreme points. As observed in Fig. 6, the per-user throughput value $R_{\boldsymbol{x}}^N(\boldsymbol{\gamma}, \alpha)/N$ of Whittle's Index Policy quickly converges to the upper bound value $r(\boldsymbol{\gamma}, \alpha)$. This result indicates that, in realistic scenarios with finite N, the global convergence result in Proposition 2 holds under moderate number of users (under $N = 50$ as shown in Fig. 6).

Fig. 6 also plots the per-user throughput performance of the BALANCEDINDEX policy, which is proposed in [23] and proved to achieve throughput half of the optimal throughput, i.e., 2-approximation performance. As observed in Fig. 6, the asymptotic per-user throughput performance of BALANCEDINDEX is strictly lower than the Whittle's Index Policy. This is because although BALANCEDINDEX policy guarantees 2-approximation to the optimal throughput performance, it does not provide strictly optimal per-user throughput performance in the asymptotic regime of large number of users, as compared with Whittle's Index Policy. Fig. 6 also evaluates the performance of a slight modification of Whittle's Index Policy, namely the THRESHOLD-WHITTLE policy, proposed in [23] by slightly adjusting the Whittles index value at belief values $p_i, i = 1, 2$. It can be observed from the figure that the per-user throughput performance of THRESHOLD-WHITTLE policy is very close to that of the Whittle's Index Policy, indicating that the modification of the Whittle's indices in THRESHOLD-WHITTLE policy does not bring significantly change the throughput performance for the plotted example. It was proven in [23] that the THRESHOLD-WHITTLE policy achieves at least half of the optimal throughput. However, analytically proving the asymptotic optimality of THRESHOLD-WHITTLE policy remains an open question.

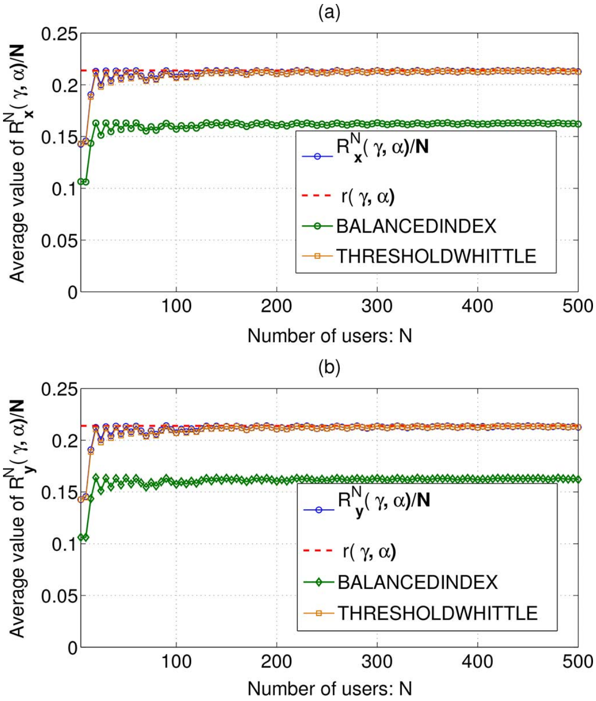

Fig. 6. Performance evaluation and comparison of per-user throughput of Whittle's Index Policy. (a) $\boldsymbol{Z}^N[0] = \boldsymbol{x}$. (b) $\boldsymbol{Z}^N[0] = \boldsymbol{y}$.

D. Evaluation of Fairness Among Users

In this section, we evaluate the fairness performance of Whittle's Index Policy. We examine the throughput difference between the two types of users, under different sets of Markov transition statistics. To facilitate better evaluation, we define the throughput $r_{\boldsymbol{x}}^N(k, \gamma, \alpha)$ to be the per-user throughput *within each class* k *of users*, i.e.,

$$r_{\boldsymbol{x}}^N(k, \gamma, \alpha) = \frac{\lim_{T \to \infty} \frac{1}{T} E\left[\sum_{t=0}^{T-1} \sum_{i \in \mathcal{N}_k} \pi_i[t] a_i^{ind}[t] \mid \boldsymbol{Z}^N[0] = \boldsymbol{x}\right]}{\gamma_k N}$$

where $\mathcal{N}_k$ represents the set of users in class k. We consider the scenario where $(p_1, r_1) = (0.9, 0.1)$ and $(p_2, r_2) = (0.6, 0.4)$ with $\boldsymbol{\gamma} = [0.5, 0.5], \alpha = 0.3$. Therefore, the channels in class 1 have a much higher degree of correlation than the channels in class 2, i.e., it is more likely for the channels in class 1 to stay in its previous-slot state than change to a different state compared to channels in class 2. However, channels in both classes have the same steady-state probability in state "1," i.e., $b_s^1 = b_s^2 = 0.5$. Fig. 7 plots the per-user throughput within each class under Whittle's Index Policy. It can be observed that users in class 1 achieve higher throughput than users in class 2. The

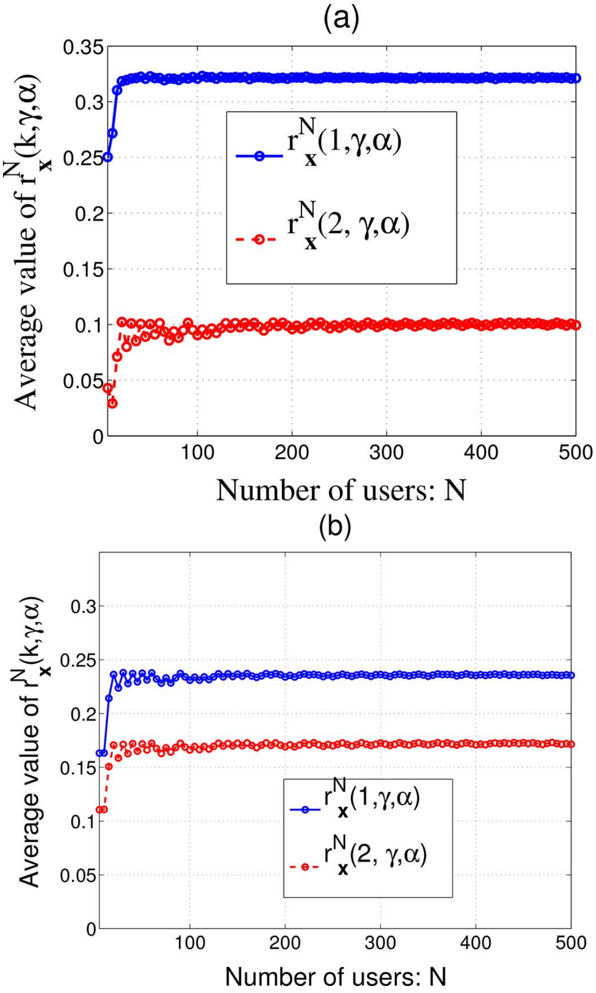

Fig. 7. Evaluation of $r^N_{\boldsymbol{x}}(k, \boldsymbol{\gamma}, \alpha)$ with N. (a) Whittle's Index Policy. (b) Policy Ξ.

higher throughput gain of class 1 is brought by the higher degree of temporal correlation and also the aforementioned "Exploitation versus Exploration" tradeoff. Since the class-1 channels have higher degree of time-correlation, if a class-1 channel is previously observed in state 1, the scheduler tends to continue to serve it for longer time to obtain high immediate gains. It is also more attractive to explore a channel in class 1 because, as previously discussed, higher future gains can be obtained if it turns out to be in state "1." Therefore, channels in class 1 have higher overall throughput than channels in class 2, resulting in the big gap in throughput between the two classes of users in Fig. 7.

To facilitate better performance in terms of fairness, we evaluate the performance of the following heuristic policy Ξ based on the Whittle's index values. In policy Ξ, instead of directly using Whittle's index values, the algorithm schedules the αN users with the largest

$$\frac{W_k(\pi_i[t])}{\overline{R}_i[t]}$$

at slot t, where $\overline{R}_i[t]$ is user i's achieved throughput up to slot t, i.e., $\overline{R}_i[t] = \sum_{\tau=1}^{t-1} \pi_i[\tau] \cdot a_i^{\Xi}[\tau] \mid \vec{\boldsymbol{\pi}}[0]$. Hence, a user's priority for scheduling is determined by its Whittle's index value relative to its own actual achieved throughput. Therefore, policy Ξ mimics the proportional fair scheduling algorithms (e.g., [3]) commonly used in communication networks. Fig. 7(b) evaluates the performance of policy Ξ. As we can see, under the algorithm Ξ, the throughput gap between the two classes of channels is closer than Whittle's index policy, indicating improved fairness performance. Finally, we believe that combining Whittle's index and the frame-based scheduling [18] can lead to low-complexity algorithms that optimally meet the fairness constraints among different users.

VIII. Conclusion

In this paper, we studied the problem of downlink scheduling over ON/OFF Markovian fading channels in the presence of channel heterogeneity. We consider the scenario where instantaneous channel state information is not perfectly known at the scheduler, but is acquired via a practical ARQ-styled feedback after each scheduled transmission. We analytically characterized the performance of Whittle's Index Policy for downlink scheduling and proved its local and global asymptotic optimality properties as the number of users scales. Specifically, provided that the initial system state is within a certain region, we established the local optimality of Whittle's Index Policy by investigating the evolution of the system belief state with a fluid approximation. We then established the global asymptotic optimality of Whittle's Index Policy under a recurrence condition, which is suitable for numerical verification. Our results establish that Whittle's Index Policy, which is attractive due to its low-complexity operation, also processes strong asymptotic optimality properties for scheduling over heterogeneous Markovian fading channels. Future research directions include design of scheduling algorithms that not only maximizes the sum throughput, but also provides fairness among heterogeneous users using Whittle's index.

Appendix A
Proof of Lemma 2

(i) First consider the scenario where $\omega^* < W_k(b_s^k)$ and suppose $\omega^* = W_k(b_{0,h_k^*}^k)$ for the belief state $b_{0,h_k^*}^k$. If the belief value of a channel is above $b_{0,h_k^*}^k$ at the beginning of a slot, the channel will be activated. According to the belief value evolution rule (1), in the next slot its belief value will either be p_k or r_k, depending on the underlying channel state revealed at the end of a slot. Clearly, the belief evolution in this case is positive recurrent within a finite state space, i.e., the belief state can only take the values $p_k, r_k, b_{0,2}^k, \ldots, b_{0,h_k^*+1}^k$. On the other hand, if the belief value is below $b_{0,h_k^*}^k$, the channel remains idle and will activate once its belief value exceeds $b_{0,h_k^*}^k$. Fig. 8 illustrates the belief evolution in steady state under this scenario.

(ii) Consider the scenario where $\omega^* \geq W_k(b_s^k)$. In this case, a channel is activated if its index value is above ω^*. After transmission, if the channel is observed to be in OFF state, its belief value will transit to r_k and stays idle until its index value crosses

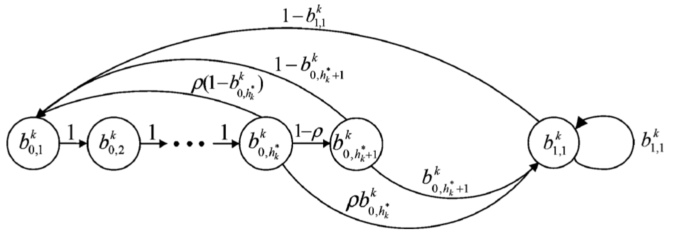

Fig. 8. Belief value transition in steady state when $\omega^* = W_k(b^k_{0,h^*_k})$.

ω^*. Since $\omega^* \geq W_k(b^k_s)$, it is clear from the belief value evolution (see Fig. 2) that, starting from r_k, the belief value will always be smaller than b^k_s. Hence the channel will stay idle at all times. On the other hand, if the channel is observed to be in ON state after transmission, the belief value will transit to p_k and the channel will keep on transmitting until the underlying channel turns out to be in OFF state. Since we assumed $p_k < 1$, the channel will ultimately be in OFF state and its belief value will transit to r_k and stays in idle mode ever since. Therefore, eventually no channel in class k will be scheduled, and the belief values will keep transit toward, but never reach, the steady-state belief value b^k_s.

Appendix B
Proof of Lemma 3

Consider two systems with different total number of users but identical α and $\boldsymbol{\gamma}$. Suppose the first system has N_1 total number of users while the second system has N_2 number of users. For the first system with N_1 total number of users, suppose the policy ϕ^*, specified in Lemma 1, is optimal for the relaxed-constraint problem. For each channel i in class k, we let $A^k_{\phi^*}$ denote the expected fraction of time of activation, i.e.,

$$A^k_{\phi^*} = \limsup_{T\to\infty} \frac{1}{T} E\left[\sum_{t=0}^{T-1} a_i^{\phi^*}[t]\right].$$

Then, according to Lemma 1(ii), the expected number of activated users satisfies

$$\gamma_1 N_1 \cdot A^1_{\phi^*} + \gamma_2 N_1 \cdot A^2_{\phi^*} = \alpha N_1.$$

Now apply the same policy ϕ^* when the total number of users is N_2. Since ϕ^* schedules each channel independently, $A^1_{\phi^*}$ and $A^2_{\phi^*}$ does not change in this scenario. Therefore, the expected number of activated users is expressed as

$$\gamma_1 N_2 \cdot A^1_{\phi^*} + \gamma_2 N_2 \cdot A^2_{\phi^*} = \frac{N_2}{N_1}\left[\gamma_1 N_1 \cdot A^1_{\phi^*} + \gamma_2 N_1 \cdot A^2_{\phi^*}\right] = \alpha N_2$$

hence the complementary slackness condition [i.e., Lemma 1(ii)] for the relaxed-constraint problem is also satisfied under ϕ^*, when the total number of users is N_2. Hence, the policy ϕ^* satisfies both Lemma 1(i) and (ii) under the total number of users N_2 and is an optimal policy for that scenario.

Therefore, fixing system parameters $(\boldsymbol{\gamma}, \alpha)$, for different number N of users, the policy ϕ^* is always optimal. Since the policy ϕ^* schedules each channel independently, we let $v_k(\boldsymbol{\gamma}, \alpha)$ denote the expected reward contributed by each channel in class k. Hence, we have

$$v^N(\boldsymbol{\gamma}, \alpha) = N\gamma_1 v_1(\boldsymbol{\gamma}, \alpha) + N\gamma_2 v_2(\boldsymbol{\gamma}, \alpha).$$

Therefore, the per-user throughput is

$$\frac{v^N(\boldsymbol{\gamma}, \alpha)}{N} = \gamma_1 v_1(\boldsymbol{\gamma}, \alpha) + \gamma_2 v_2(\boldsymbol{\gamma}, \alpha)$$

which is independent of N. Hence, the lemma is proven.

Appendix C
Proof of Lemma 4

Given system parameters $(\boldsymbol{\gamma}, \alpha)$, we know from the proof of Lemma 3 that the form of the Optimal Relaxed Policy, denoted by ϕ^*, does not change with the number N of users. Since ϕ^* schedules each channel independently, we let vector $\boldsymbol{\varepsilon}^k = [\varepsilon^k_{0,1}, \ldots, \varepsilon^k_{0,\tau}, \varepsilon^k_s, \varepsilon^k_{1,\tau}, \ldots, \varepsilon^k_{1,1}]$ denote the steady-state distribution of the belief value of a user in class k under ϕ^*, with $\varepsilon^k_s + \sum_{c,h} \varepsilon^k_{c,h} = 1$. Therefore

$$E[\boldsymbol{Z}^N(\infty)] = \frac{1}{N}[\gamma_1 N\boldsymbol{\varepsilon}^1, \gamma_2 N\boldsymbol{\varepsilon}^2] = [\gamma_1\boldsymbol{\varepsilon}^1, \gamma_2\boldsymbol{\varepsilon}^2].$$

Since ϕ^* is independent of N, $\boldsymbol{\varepsilon}^k$ is independent of N for $k = 1, 2$. Therefore, $E[\boldsymbol{Z}^N(\infty)]$ is independent of the user number N, which proves the lemma.

References

[1] R. Knopp and P. A. Humblet, "Information capacity and power control in single cell multiuser communications," in *Proc. IEEE ICC*, 1995, pp. 331–335.

[2] X. Liu, E. K. P. Chong, and N. B. Shroff, "Opportunistic transmission scheduling with resource-sharing constraints in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 10, pp. 2053–2064, Oct. 2001.

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

[4] 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.

[5] X. Lin and N. B. Shroff, "The impact of imperfect scheduling on cross-layer congestion control in wireless networks," *IEEE/ACM Trans. Netw.*, vol. 14, no. 2, pp. 302–315, Apr. 2006.

[6] A. Eryilmaz and R. Srikant, "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] C. Safran and C. G. Chute, "Exploration and exploitation of clinical databases," *Int. J. Bio-Med. Comput.*, vol. 39, pp. 151–156, 1995.

[8] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *J. Artif. Intell. Res.*, vol. cs.AI/9605, pp. 237–285, 1996.

[9] M. J. Neely, S. T. Rager, and T. F. La Porta, "Max weight learning algorithms for scheduling in unknown environments," *IEEE Trans. Autom. Control*, vol. 57, no. 5, pp. 1179–1191, May 2012.

[10] J. Huang, R. A. Berry, and M. L. Honig, "Wireless scheduling with hybrid ARQ," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2801–2810, Nov. 2005.

[11] R. Aggarwal, M. Assaad, C. E. Koksal, and P. Schniter, "Joint scheduling and resource allocation in the ofdma downlink: Utility maximization under imperfect channel-state information," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5589–5604, Nov. 2011.

[12] C. Thejaswi, J. Zhang, S. Pun, and V. H. Poor, "Distributed opportunistic scheduling with two-level channel probing," *IEEE/ACM Trans. Netw.*, vol. 18, no. 5, pp. 1464–1477, Oct. 2009.

[13] W. Ouyang, S. Murugesan, A. Eryilmaz, and N. B. Shroff, "Scheduling with rate adaptation under incomplete knowledge of channel/estimator statistics," in *Proc. Allerton Conf.*, 2010, pp. 670–677.

[14] L. Ying and S. Shakkottai, "On throughput optimality with delayed network-state information," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5116–5132, Aug. 2011.

[15] Q. Zhao, B. Krishnamachari, and K. Liu, "On myopic sensing for multichannel opportunistic access: Structure, optimality, and performance," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5431–5440, Dec. 2008.

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

[17] C. Li and M. J. Neely, "Exploiting channel memory for multi-user wireless scheduling without channel measurement: Capacity regions and algorithms," *Perform. Eval.*, vol. 68, no. 8, pp. 631–657, 2011.

[18] C. Li and M. J. Neely, "Network utility maximization over partially observable Markovian channels," in *Proc. IEEE WiOpt*, May 2011, pp. 17–24.

[19] P. Whittle, "Restless bandits: Activity allocation in a changing world," *J. Appl. Probab.*, vol. 25, pp. 287–298, 1988.

[20] K. Liu and Q. Zhao, "Indexability of restless bandit problems and optimality of Whittle's index for dynamic multi-channel access," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5547–5567, Nov. 2010.

[21] W. Ouyang, S. Murugesan, A. Eryilmaz, and N. Shroff, "Exploiting channel memory for joint estimation and scheduling in downlink networks," in *Proc. IEEE INFOCOM*, 2011, pp. 3056–3064.

[22] R. Weber and G. Weiss, "On an index policy for restless bandits," *J. Appl. Probab.*, vol. 27, no. 3, pp. 637–648, 1990.

[23] S. Guha, K. Munagala, and P. Shi, "Approximation algorithms for restless bandit problems," *J. ACM*, vol. 58, no. 1, 2010, Art. no. 3.

[24] S. Murugesan, P. Schniter, and N. B. Shroff, "Opportunistic scheduling using ARQ feedback in multi-cell downlink," in *Proc. Asilomar*, 2010, pp. 1733–1737.

[25] E. J. Sondik, "The optimal control of partially observable Markov decision processes," Ph.D. dissertation, Stanford University, Stanford, CA, USA, 1971.

[26] E. Altman, *Constrained Markov Decision Processes*. London, U.K.: Chapman & Hall, 1999.

[27] C. Papadimitriou and J. N. Tsitsiklis, "The complexity of optimal queueing network control," *Math. Oper. Res.*, vol. 24, no. 2, pp. 293–305, 1999.

[28] W. Ouyang, A. Erilmaz, and N. B. Shroff, "Asymptotically optimal downlink scheduling over Markovian fading channels," in *Proc. IEEE INFOCOM*, Orlando, FL, USA, 2012, pp. 1224–1232.

[29] W. Ouyang, A. Eryilmaz, and N. B. Shroff, "Low-complexity optimal scheduling over correlated fading channels with ARQ feedback," in *Proc. IEEE WiOpt*, Paderborn, Germany, 2012, pp. 270–277.

[30] A. Shwartz and A. Weiss, *Large Deviation for Performance Analysis*. London, U.K.: Chapman & Hall, 1994.

[31] P. K. Dutta, "What do discounted optima converge to? A theory of discount rate asymptotics in economic models," *J. Econ. Theory*, vol. 55, pp. 64–94, 1991.

[32] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA, USA: Athena Scientific, 1999.

[33] T. G. Kurtz, "Strong approximation theorems for density dependent Markov chains," *Stochastic Processes Their Appl.*, vol. 6, no. 3, pp. 223–240, 1978.

[34] R. A. Horn, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1999.

[35] W. J. Rugh, *Linear System Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1996.

[36] W. Ouyang, A. Eryilmaz, and N. B. Shroff, "Asymptotically optimal downlink scheduling over Markovian fading channels," arXiv:1108.3768.

Wenzhuo Ouyang (S'07–M'14) received the B.E. degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2008, and the Ph.D. degree in electrical and computer engineering from The Ohio State University, Columbus, OH, USA, in 2014.

During his bachelor's study, he was an exchange student with System and Control Group, Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM, USA, in 2007. He is currently a Postdocoral Researcher with Rice University, Houston, TX, USA. His research interests span the areas of communication systems, wireless networks, Markov decision process, and stochastic systems.

Dr. Ouyang served the TPC of the S3 Workshop 2011, IEEE WiOPT 2015, and IEEE WASA 2015. He received the Best Student Paper Award at IEEE WiOPT 2012 and the Student Travel Award at IEEE INFOCOM 2011.

Atilla Eryilmaz (S'00–M'06) received the B.S. degree in electrical and electronics engineering from Boğaziçi University, Istanbul, Turkey, in 1999, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Illinois at Urbana-Champaign, Urbana, IL, USA, in 2001 and 2005, respectively.

Between 2005 and 2007, he worked as a Postdoctoral Associate with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, USA. He is currently an Associate Professor of electrical and computer engineering with The Ohio State University, Columbus, OH, USA. His 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.

Dr. Eryilmaz received the NSF CAREER and the Lumley Research awards in 2010.

Ness B. Shroff (S'91–M'93–SM'01–F'07) received the Ph.D. degree in electrical engineering from Columbia University, New York, NY, USA, in 1994.

He joined Purdue University, West Lafayette, IN, USA, immediately thereafter as an Assistant Professor with the School of Electrical and Computer Engineering. At Purdue, he became Full Professor of electrical and computer engineering (ECE) in 2003 and Director of the Center for Wireless Systems and Applications (CWSA) in 2004. In 2007, he joined The Ohio State University, Columbus, OH, USA, where he holds the Ohio Eminent Scholar Endowed Chair in Networking and Communications in the departments of ECE and Computer Science and Engineering. From 2009 to 2012, he served as a Guest Chaired Professor of wireless communications with Tsinghua University, Beijing, China, and currently holds an honorary Guest Professor with Shanghai Jiaotong University, Shanghai, 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 the IEEE/ACM TRANSACTIONS ON NETWORKING and the IEEE COMMUNICATION LETTERS. He currently serves on the Editorial Board of *Computer Networks*, *IEEE Network*, and *Networking Science*. He has chaired various conferences and workshops, and co-organized workshops for the NSF to chart the future of communication networks. He is an NSF CAREER awardee. He has received numerous Best Paper awards for his research, e.g., at IEEE INFOCOM 2008, at IEEE INFOCOM 2006, in *Communication and Networking* in 2005, and in *Computer Networks* in 2003 (two of his papers also received runner-up awards at IEEE INFOCOM 2005 and 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. He is among the list of highly cited researchers from Thomson Reuters (formerly ISI Web of Knowledge) and in Thomson Reuters' book *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.