Opportunistic Scheduling using ARQ feedback in Multi-Cell Downlink

Sugumar Murugesan, *Member, IEEE*, Philip Schniter, *Senior Member, IEEE*, Ness B. Shroff, *Fellow, IEEE*

***Abstract*—We study cooperative, opportunistic multiuser scheduling using ARQ feedback in multi-cell downlink systems. Adopting the *cell breathing* ICI control mechanism, we formulate the scheduling problem as an infinite horizon discounted reward partially observable Markov decision process and study two scenarios. When the cooperation between the cells is asymmetric, we show that the optimal scheduling policy has a greedy flavor and is simple to implement. Under symmetric cooperation, we link the scheduling problem with restless multiarmed bandit processes and propose a low complexity index scheduling policy. The proposed index policy is essentially Whittle's index policy, if the scheduling problem is Whittle indexable. Extensive numerical experiments suggest that the proposed policy is near-optimal.**

I. Introduction

Opportunistic multiuser scheduling [1], where system resources are allocated to user(s) based on instantaneous channel conditions, is well suited to the cellular environment thanks to the presence of the base station, a central coordinating authority. It is understandable that the availability of channel state information at the scheduler is crucial for the success of opportunistic scheduling schemes. Much of the literature on this topic make the unrealistic assumption of readily available channel state information at the scheduler. In reality, however, significant resources must be spent to obtain channel state information. This leads us to the following critical question: *Are there efficient joint channel acquisition - multiuser scheduling schemes for cellular systems?* For single-cell downlink systems, this issue was recently addressed in independent works [2] and [3]. In these works, the authors model the fading channels using *i.i.d* ON-OFF Markov channels (Gilbert Elliott (GE) channels [4]) and show that the memory in the fading channels can be exploited using the ARQ[1] feedback mechanism for efficient opportunistic scheduling. Rarely in realistic scenarios do we encounter a single-cell system, however. In a multi-cell system, transmission in one cell interferes with the transmissions in adjacent cells. It follows that the channel states of users in a cell is a function of transmissions and scheduling decisions in the adjacent cells, effectively imparting a convoluted dependence between the scheduling choices in neighboring cells. We now face the question: *How do we exploit the channel memory and the ARQ feedback mechanism for opportunistic scheduling in a multi-cell environment ?*

We address this problem by following a two layered approach: A well established inter-cell interference (ICI) control mechanism is adopted and assumed to be in place. On top of this layer, we optimize ARQ based multiuser scheduling across the cells. We restrict our focus to two-cell systems since our analysis can be readily extended to multi-cell systems with the use of six-directional antennae at the base stations [5].

By formulating the scheduling problem as an infinite horizon discounted reward Partially Observable Markov Decision Process (POMDP) [6], we study it under two scenarios: asymmetric and symmetric cooperation between cells. Under asymmetric cooperation, cell 1 makes scheduling decisions without any regard to the scheduling decisions in cell 2, while cell 2 adapts its decisions based on cell 1 behavior, so that the cell breathing protocol is not violated. Such asymmetric cooperation can model scenarios such as: cell 1 covers the heart of a city with higher data rate requirements compared to cell 2 that covers the suburbs. Under asymmetric cooperation, we explicitly characterize the optimal scheduling policy and show that it has a *greedy flavor* and that it can be implemented by a simple *round-robin* algorithm. In the symmetric cooperation case, cell 1 and cell 2 mutually cooperate, effectively emulating a centralized scheduling problem. We connect the symmetric case scheduling problem with Restless Multiarmed Bandit processes (RMAB) [7] and study it in a *Whittle's indexability* framework [7]. Based on this study, we propose an index policy which is, in fact, *Whittle's index policy* if the scheduling problem is Whittle indexable. The proposed policy is simple to implement and has near-optimal numerical performance.

II. Problem Setup

Consider a two-cell system. Time is slotted and the base stations are the central controllers that control transmission to the users within their respective cells, in each slot. The fading coefficients between the base stations and the users are modeled with memory, i.e., with two-state Markov chains. The fading coefficients of the channels evolve synchronously across time slots based on the Markov chain statistics. We adopt the *cell breathing* protocol [8]- [10] as the ICI control mechanism. Consistent with [8], within each cell, we cluster users into near and far users, based on the geometric distance

This work was supported by the NSF CAREER grant 237037, the Office of Naval Research grant N00014-07-1-0209, NSF grants CNS-0721236, CNS-0626703, ARO W911NF-08-1-0238 and ARO W911NF-07-10376.

Murugesan was with the Dept. of ECE, The Ohio State University and is currently with the Dept. of ECEE, Arizona State University; Schniter is with the Department of ECE, The Ohio State University, and Shroff holds a joint appointment in both the Departments of ECE and CSE at The Ohio State University, Columbus, OH 43210, USA (E-mail: murugesan.4@buckeyemail.osu.edu, {schniter, shroff}@ece.osu.edu.)

[1] A standard in most upcoming wireless systems.

between the users and their respective base stations. Let N_j and F_j denote the number of near users and far users, respectively, in cell j. Only one user is scheduled per slot per cell. Also, in each slot, if, in any cell, a far user is scheduled, transmission takes place at power $P_1, P_1 > 0$ and the neighboring cell schedules a near user at the reduced power $P_2, P_2 < P_1$. This essentially aims to nullify the multi-cell variant of the *near-far* effect seen in single cell systems. We can show that [11], under cell breathing, the SINR channel (henceforth, simply the 'channel') of each user can also be modeled as *i.i.d* Markov chains with ON/OFF states. The two-state Markov channel is characterized by a 2×2 probability transition matrix

$$P = \begin{bmatrix} p & 1-p \\ r & 1-r \end{bmatrix}.$$

Here p denotes the probability that the channel evolves from ON state to ON state in adjacent slots and r denotes the probability that the channel evolves from OFF to ON state. Throughout this work, we assume the fading coefficients, and hence the SINR channels, are positively correlated, i.e., $p > r$.

A traditional ARQ based transmission is deployed in each cell. That is, at the beginning of a time slot, the head of line packet of the scheduled user is transmitted. If the packet does not go through, i.e., it is not successfully decoded by the user (occurs when the channel is in the OFF state), a NACK is reported by the user at the end of the slot, and the packet is retained at the head of the queue for retransmission at a later time. If the packet does go through (i.e., the ON state), an ACK is reported and the packet is removed from the queue. For each successful transmission, the corresponding base station accrues a reward 1 and no reward upon a failed transmission. This reward structure will be further explained shortly. At the end of the slot, the base stations of neighboring cells share their ARQ information. Thus each cell has all the channel information available to its neighbor, hence facilitating joint scheduling among the cells. The performance metric that the cells aim to maximize is the expected discounted reward over an infinite horizon, defined below.

Expected Discounted Reward under Policy $\mathfrak{A}$: Define the belief value of user i as the probability that the channel between user i and the corresponding base station is in the ON state in the current slot. A stationary scheduling policy $\mathfrak{A}$ is a stationary mapping between the belief values of the users and the near-far or far-near user pair scheduled in the adjacent cells. Let π^{n_j} indicate the vector of belief values of near users in cell j and π^{f_j} is similarly defined for far users in cell j. Thus with $\{\pi^{n_1}, \pi^{f_1}, \pi^{n_2}, \pi^{f_2}\}$ as the current system state, the expected discounted reward over an infinite horizon, under a policy $\mathfrak{A}$, is given by

$$\begin{aligned} & V(\pi^{n_1}, \pi^{f_1}, \pi^{n_2}, \pi^{f_2}, \mathfrak{A}) \\ = \; & R(\pi^{n_1}, \pi^{f_1}, \pi^{n_2}, \pi^{f_2}, (\mathsf{a}, a)) \\ & + \beta\, \mathrm{E}[V(T(\pi^{n_1}), T(\pi^{f_1}), T(\pi^{n_2}), T(\pi^{f_2}), \mathfrak{A})], \quad (1) \end{aligned}$$

where (a, a) is the user pair scheduled under policy $\mathfrak{A}$ in the current slot. The expectation is over the belief values $T(\pi^{n_1}), T(\pi^{f_1}), T(\pi^{n_2}), T(\pi^{f_2})$ and $T(.)$ is the Markovian belief evolution operator conditioned on the belief values and the ARQ feedbacks from the previous slot. The discount factor $\beta \in (0,1)$ gives greater weight to the immediate reward than the future reward. With n_j indicating the set of near users in cell j and f_j similarly defined for the set of far users, the expected immediate reward is given by

$$\begin{aligned} & R(\pi^{n_1}, \pi^{f_1}, \pi^{n_2}, \pi^{f_2}, (\mathsf{a}, a)) \\ = \; & \begin{cases} \pi^{n_1}(\mathsf{a}) + \pi^{f_2}(a), & \text{if } (\mathsf{a}, a) \in (n_1, f_2) \\ \pi^{f_1}(\mathsf{a}) + \pi^{n_2}(a), & \text{if } (\mathsf{a}, a) \in (f_1, n_2). \end{cases} \quad (2) \end{aligned}$$

Optimal Stationary Policy: A stationary policy that maximizes the expected discounted reward is optimal. Thus, for any state $\{\pi^{n_1}, \pi^{f_1}, \pi^{n_2}, \pi^{f_2}\}$, the optimal stationary policy is given by

$$\mathfrak{A}^* = \arg\max_{\mathfrak{A}} V(\pi^{n_1}, \pi^{f_1}, \pi^{n_2}, \pi^{f_2}, \mathfrak{A}) \quad (3)$$

III. Optimal Scheduling under Asymmetric Cooperation between Cells

Consider a system where cell breathing is deployed by the following asymmetric cooperation between the cells: cell 1 schedules transmission to its users without any regard to the decisions in cell 2, while cell 2 schedules based on the user group choice of cell 1, to conform with the cell breathing protocol. Cell 1 is aware of this compromise made by cell 2 and therefore adopts the two state Markov model for the channels of cell 1 users. Such an asymmetric cooperation can result from scenarios such as (1) Cell 1 covers the heart of a city with higher data rate requirements compared to cell 2, which covers the suburbs, (2) Sharing of ARQ feedback information between the adjacent base stations is not mutual due to, e.g., a partial link failure between the base stations, (3) In the context of game theory, cell 1 is a selfish player and cell 2 is a rule-abiding player.

Let π^j denote the vector of belief values of *all* users in cell j. Thus $\pi^j = \pi^{n_1} \cup \pi^{f_1}$. Let $\hat{\mathfrak{A}}$ denote the greedy policy within cell 1 with the mapping given by: $\pi^1 \to \mathsf{a} = \arg\max_i \pi^1(i)$. Let $\hat{\mathfrak{A}}^f$ denote the greedy policy within the far-user group, within cell 2. Thus $\hat{\mathfrak{A}}^f : \pi^{f_2} \to \arg\max_i \pi^{f_2}(i)$. Define the greedy policy $\hat{\mathfrak{A}}^n$ similarly for near users in cell 2. Define the joint, two-cell scheduling policy $\{\hat{\mathfrak{A}}, \hat{\mathfrak{A}}^f, \hat{\mathfrak{A}}^n\}$ as below: In cell 1, greedy policy $\hat{\mathfrak{A}}$ is implemented in each slot, without any regard to the cell breathing protocol. In cell 2, in each slot, if policy $\hat{\mathfrak{A}}$ in cell 1 scheduled a near user, cell 2 schedules a far user by implementing policy $\hat{\mathfrak{A}}^f$. Similarly, if cell 1 scheduled a far user, cell 2 schedules a near user by implementing the greedy policy $\hat{\mathfrak{A}}^n$ within the near user group. We now report our result on the optimal scheduling policy in the two-cell downlink, under asymmetric cooperation.

Proposition 1. *Under asymmetric cooperation, the policy* $\{\hat{\mathfrak{A}}, \hat{\mathfrak{A}}^f, \hat{\mathfrak{A}}^n\}$ *is optimal.*

Proof outline: The proof proceeds by showing that, conditioned on the class (near/far) of the user scheduled in cell 1, the scheduling decisions in cell 1 and cell 2 are decoupled. Details are available in [11].

Note that the optimal policy $\{\hat{\mathfrak{A}}, \hat{\mathfrak{A}}^f, \hat{\mathfrak{A}}^n\}$ is a variant of the greedy policy and can be implemented using a simple round-robin algorithm. Also, it can be shown that the knowledge of the Markov channel parameters, i.e, p, r, is not required for the implementation of the optimal policy. The reader is referred to [11] for more details and discussion.

IV. Scheduling under Symmetric Cooperation between Cells - Index Policy

A direct optimality analysis of the ARQ based scheduling problem, with symmetric cooperation, appears very difficult due to the complex relationship between the scheduling decisions across space and time. We therefore establish a connection between the scheduling problem and the restless multiarmed bandit (RMAB) processes [7] and make use of the established theory behind RMAB processes in our analysis. A brief introduction to RMAB processes follows.

A. Restless Multiarmed Bandit Processes

RMAB processes are defined as a family of sequential dynamic resource allocation problems in the presence of several competing, *independently evolving* projects. They are characterized by a fundamental trade-off between decisions guaranteeing high immediate rewards versus those that sacrifice immediate rewards for better future rewards. In each slot, a single project must be allocated the available system resources. The state of the projects stochastically evolves from the current time slot to the next time slot, depending on the current state and the action taken.

Since a direct analysis of RMAB processes is traditionally known to be hard, Whittle developed the notion of *indexability* for RMAB processes and it is known (e.g., [12]) that some RMAB's satisfying this hard-to-establish property yield to low complexity, index-type optimal scheduling policies. We formally introduce Whittle's indexability framework next: *Consider only one project of the RMAB. The scheduler in each slot must either activate the project or let it stay idle. In the former case a reward dependent on the state of the project is accrued. This reward structure is the same as the one used in the original RMAB. In the case of the idle decision, a reward W for passivity is accrued. The goal of the scheduler is to maximize the total discounted reward over an infinite horizon. For a project state π, the value of W corresponding to equal net rewards for activate and idle decisions is defined as the index $I(\pi)$. Call the optimal activate/idle scheduling policy as the W-subsidy policy. Let $D(W)$ be the set of states under which a project would be kept idle under the W-subsidy policy. Let $\mathcal{S}$ denote the state space of the project. The project is indexable if $D(W)$ increases monotonically from $\emptyset$ to $\mathcal{S}$ as W increases from $-\infty$ to ∞. The RMAB is Whittle indexable if all the projects are indexable.*

It can be shown that the two-cell scheduling policy under symmetric cooperation is a general variant of the RMAB processes [11]. In the rest of this section we perform a Whittle's indexability study of the scheduling problem and using the structural results of this study, we derive an index scheduling policy.

B. Indexability Analysis

Consider a single project made up of a near-far or a far-near user pair. In each slot, the state of the project is given by the belief values of the channels of the users, i.e., (π_1, π_2). If the project is scheduled in a slot, i.e., if the users are scheduled, the belief value evolves into one of the following states: $\{(r,r),(r,p),(p,r),(p,p)\}$ corresponding, respectively, to ARQ feedbacks (NACK,NACK), (NACK,ACK), (ACK,NACK) and (ACK,ACK) respectively. Recall that p and r are the elements of the probability transition matrix of the Markov modeled channels with $p > r$.

Define V^a and V^p as the total discounted rewards corresponding to activate and idle decisions in the current slot and optimal decisions in future slots. Let V denote the optimal total discounted reward. Thus, with $V_{xy} := V(x,y)$,

$$\begin{aligned} V^a(\pi_1,\pi_2) &= \pi_1 + \pi_2 + \beta\Big(\pi_1\pi_2 V_{pp} + \pi_1(1-\pi_2)V_{pr} \\ &\quad + (1-\pi_1)\pi_2 V_{rp} + (1-\pi_1)(1-\pi_2)V_{rr}\Big) \\ V^p(\pi_1,\pi_2) &= W + \beta V(T(\pi_1),T(\pi_2)) \\ V(\pi_1,\pi_2) &= \max(V^a(\pi_1,\pi_2), V^p(\pi_1,\pi_2)) \end{aligned} \tag{4}$$

where, recall, $T(\pi)$ is the one-step Markov channel evolution operator on the belief value π, conditioned on the current decision. Let R_I denote the region $\{(\pi_1,\pi_2); \pi_1 \in [\pi_{ss},1], \pi_2 \in [\pi_{ss},1]\}$, with $\pi_{ss} = \frac{r}{1-(p-r)}$ denoting the steady state ON-probability of the Markov channels. Let R_{II} denote the union of the regions $R_{II}^1 \doteq \{(\pi_1,\pi_2); \pi_1 \in [0,\pi_{ss}], \pi_2 \in [\pi_{ss}, 2\pi_{ss}-\pi_1]\}$ and $R_{II}^2 \doteq \{(\pi_1,\pi_2); \pi_2 \in [0,\pi_{ss}], \pi_1 \in [\pi_{ss}, 2\pi_{ss}-\pi_2]\}$. Let $\mathcal{A}$ be the set of states (π_1,π_2) in which it is optimal to activate. Let $\mathcal{P}$ be the set corresponding to optimal idle decision.

We first report the thresholdability properties of the W-subsidy scheduling policy when W is such that $V^a(\pi_{ss},\pi_{ss}) \geq V^p(\pi_{ss},\pi_{ss})$ and when $V^a(\pi_{ss},\pi_{ss}) < V^p(\pi_{ss},\pi_{ss})$. Proof details are available in [11].

Proposition 2. *If W is such that $V^a(\pi_{ss},\pi_{ss}) \geq V^p(\pi_{ss},\pi_{ss})$, then*

(1) $R_I \in \mathcal{A}$

(2) *Within region R_{II}, the threshold boundary is given by the upper segment of the hyperbola*

$$V^a(\pi_1,\pi_2) = W + \beta V^a(T(\pi_1),T(\pi_2))$$

where

$$\begin{aligned} &V^a(x_1,x_2) \\ &\quad = x_1 + x_2 + \beta\big[(1-x_1)(1-x_2)V(r,r) \\ &\quad\quad + (1-x_1)(x_2)V(r,p) \\ &\quad\quad + x_1(1-x_2)V(p,r) + x_1x_2V(p,p)\big], \end{aligned}$$

$T(x) = x(p-r)+r$, *and upper segment indicates the segment of the hyperbola that lies in the first quadrant around the asymptotes.*

Proposition 3. *If W is such that $V^a(\pi_{ss}, \pi_{ss}) < V^p(\pi_{ss}, \pi_{ss})$, then*

(1) $(\pi_1, \pi_2) \in \mathcal{P}$, $\forall \pi_1 + \pi_2 \leq 2\pi_{ss}$

(2) *Within region R_I, the threshold boundary is given by the upper segment of the hyperbola*

$$V^a(\pi_1, \pi_2) = \frac{W}{1-\beta}.$$

C. Index Policy

The thresholdability results of the W-subsidy policy and hence the threshold boundaries reported above were obtained using sufficient conditions that hold only in the shown regions. A tighter analysis needed to characterize the optimal W-subsidy policy in the whole state space appears intractable. We therefore make a set of assumptions (A) on the properties of the optimal W-subsidy policy and derive an index scheduling policy for the two-cell system. Our assumptions are stated next.

(A0) The threshold boundaries reported in Propositions 2 and 3 in the restricted regions hold true in the entire state space.

(A1) The threshold boundaries progressively move to the right, i.e., the region $\mathcal{P}$ progressively expands, as W increases. This is essentially Whittle's indexability.

An illustration of the extrapolated boundaries in comparison with the established boundaries is available in [11].

We now classify the state space into four non-overlapping regions. Let W_0 be defined as the value of W at which $V^a(\pi_{ss}, \pi_{ss}) = V^p(\pi_{ss}, \pi_{ss})$. The state space is now classified as below:

- R_1: (π_1, π_2) such that $\pi_1 + \pi_2 \leq 2r$
- R_2: Region between the boundaries $\{(\pi_1, \pi_2) : \pi_1 + \pi_2 > 2r\}$ and $\{(\pi_1, \pi_2) : V^a(\pi_1, \pi_2)|_{W=W_0} = V^p(\pi_1, \pi_2)_{W=W_0}\}$. By definition of W_0, the second boundary passes through the steady state.
- R_3: (π_1, π_2) such that $(\pi_1, \pi_2) \notin R_1 \cup R_2$ and $\pi_1 + \pi_2 < 2p$
- R_4: (π_1, π_2) such that $\pi_1 + \pi_2 \geq 2p$

Under assumptions (A0) and (A1), as W increases from 0 to $2r$, the threshold boundary moves progressively outwards within region R_1, with the boundary given by (π_1, π_2) such that $\pi_1 + \pi_2 = W$. As W increases from $2r$ to W_0, the threshold boundary progressively moves outward within region R_2, with the boundary given by the extrapolation of the boundary derived in Proposition 2, i.e., (π_1, π_2) such that $V^a(\pi_1, \pi_2) = W + \beta V^a(T(\pi_1), T(\pi_2))$. When W increases from W_0 to $2p$, the threshold boundary progressively moves outward within R_3 and the boundary is given from Proposition 3 by the convex curve: (π_1, π_2) such that $V^a(\pi_1, \pi_2) = \frac{W}{1-\beta}$. When W increases from $2p$ to 2, the threshold boundary progressively moves outward within region R_4 with boundary given by (π_1, π_2) such that $\pi_1 + \pi_2 = W$.

The index we employ in our policy is defined as follows: For any state (π_1, π_2), the value of W for which the threshold boundary passes through (π_1, π_2) is the index of that state. Note that if assumptions (A) were true, the index we defined is the Whittle's index. Thus, when (A) is true, the index policy we propose is, in fact, the Whittle's index policy .

Initialization
This step involves evaluating the quantities $V(p,p), V(p,r), V(r,r), W_0$ and identifying the regions $R_1, \ldots, R_4$. Details of this step are available in [11].

Index policy on belief vector $\pi = (\pi_1, \ldots, \pi_N)$
(1) Within each user group (n_1, f_1, n_2, f_2), identify the users that have the highest belief values. Call them n_1^*, f_1^*, n_2^* and f_2^*, respectively. It can be shown [11] that, user pair (n_1^*, f_2^*) has an index higher than any other user pair from the composite group $n_1 \times f_2$. Likewise, the user pair (f_1^*, n_2^*) has higher index than any other pair from $f_1 \times n_2$. Thus it is sufficient to compare the indices of user pairs (n_1^*, f_2^*) and (f_1^*, n_2^*).
(2) Calculate the index of the states corresponding to user pairs (n_1^*, f_2^*) and (f_1^*, n_2^*). Index calculation is later explained as a separate step.
(3) Schedule the user pair with the higher index.
(4) Receive ARQ feedback from the scheduled user pair.
(5) Update the belief values of all the users based on the scheduling decisions and the ARQ feedback.
(6) Repeat the scheduling policy in the next time slot.

Index calculation routine
(1) Determine the region (R_1, R_2, R_3, R_4) in which (π_1, π_2) belongs.
(2) Based on the identified region, identify the threshold boundary from the discussion in Subsections B and C. Table 1.
(3) Determine the value of W for which the identified threshold boundary passes through (π_1, π_2). This can be accomplished as follows: For discretized values of $W = 0 : \delta_W : 2$, find the value of W (call it W^*) for which the threshold boundary is closest to (π_1, π_2).
(4) W^* is the index of the user pair. Return W^* to the *index policy* routine.

V. Numerical Results and Discussion

We now proceed to report the numerical performance of the proposed index policy. Table I reports the quantity %subopt= $\frac{V_{\text{opt}} - V_{\text{index}}}{V_{\text{opt}}} \times 100\%$ where V_{opt} is the optimal infinite horizon discounted reward and V_{index} is the reward under the proposed index policy. The optimal policy is implemented by an exhaustive search over all possible $N_1 \times F_2 + F_1 \times N_2$ decisions in each time slot. The policies are repeated for increasing horizon lengths until convergence is reached. The quantity %subopt quantifies the degree of sub-optimality of the index policy. The very low values of %subopt suggests that the index policy is near optimal.

To illustrate the advantage of using ARQ feedback for scheduling, we compare the index policy performance with two baseline cases: V_{genie} and V_{rand} in Table II. V_{genie} is the optimal reward in the genie-aided system defined as follows: At the end of every time slot, the scheduler learns about the

$N1 = 2,\ F1 = 3,\ N2 = 2,\ F2 = 3$			
p	r	β	%subopt
0.7638	0.3663	0.8013	0.0008 %
0.9504	0.5462	0.8452	0.0002 %
0.8476	0.4230	0.5358	0.0001 %
0.7452	0.6356	0.8739	0.0031 %
0.7825	0.5010	0.4170	0.0000 %
0.5546	0.4580	0.3381	0.0001 %
0.8536	0.6670	0.2880	0.0001 %
0.6688	0.4065	0.7413	0.0002 %
0.8947	0.3289	0.2060	0.0000 %
0.5387	0.4922	0.7067	0.0007 %

TABLE I
ILLUSTRATION OF THE NEAR OPTIMAL PERFORMANCE OF THE PROPOSED INDEX POLICY. EACH ROW CORRESPONDS TO RANDOMLY GENERATED SYSTEM PARAMETERS (p, r, AND β) AND INITIAL BELIEF VALUES.

$N1 = 2,\ F1 = 3,\ N2 = 2,\ F2 = 3$			
p	r	β	%ARQgain
0.7397	0.5790	0.8436	93.5592 %
0.7600	0.4697	0.1551	98.1181 %
0.7058	0.5223	0.7954	89.0940 %
0.7801	0.6404	0.9649	93.6788 %
0.7994	0.3420	0.2678	97.9097 %
0.9919	0.2318	0.4784	94.2659 %
0.6064	0.5657	0.2762	99.3912 %
0.8446	0.5150	0.3688	98.6200 %
0.7051	0.5944	0.5793	97.2564 %
0.9631	0.7544	0.3493	98.7781 %

TABLE II
ILLUSTRATION OF THE SIGNIFICANCE OF USING ARQ FEEDBACK IN OPPORTUNISTIC SCHEDULING. EACH ROW CORRESPONDS TO RANDOMLY GENERATED SYSTEM PARAMETERS (p, r, AND β) AND INITIAL BELIEF VALUES.

channel state of *every* user in the system in that time slot. The optimal scheduling policy in the genie aided system is greedy, i.e., in each slot schedule the legitimate user pair that has the highest sum of belief values. V_{rand} is the reward accrued by a policy that ignores any channel feedback from the users and schedules randomly. The quantity $\%\text{ARQgain}=\frac{V_{\text{index}}-V_{\text{rand}}}{V_{\text{genie}}-V_{\text{rand}}}$ quantifies the gain in reward when the ARQ feedback is used in scheduling. The high values of $\%$ARQgain in Table II underlines the significance of exploiting ARQ feedback in our scheduling setup. More comprehensive numerical results illustrating the near optimality of the proposed index policy and the gains associated with ARQ based scheduling can be found in [11].

VI. CONCLUSION

We studied ARQ based cooperative scheduling in multi-cell downlink system. When the cooperation between the cells is asymmetric, the optimal scheduling policy has a greedy flavor and is simple to implement. Under symmetric cooperation, however, a direct optimality analysis appears difficult. We formulated the scheduling problem as a more general variant of the restless multiarmed bandit processes and studied it from the perspective of Whittle indexability. Whittle indexability is an important condition that is known to predispose the Whittle's index policy towards optimality in various RMAB processes. Founded on the indexability analysis of the two-cell scheduling problem, we proposed an easy-to-implement index policy. Upon Whittle indexability of the scheduling problem, the proposed policy is essentially the Whittle's index policy. Extensive numerical experiments suggest that the proposed policy is near optimal and that significant gains can be realized by exploiting the memory in the fading channels.

REFERENCES

[1] R. Knopp and P. A. Humblet, "Information capacity and power control in single cell multiuser communications," *Proc. IEEE International Conference on Communications*, (Seattle, WA), pp. 331-335, June 1995.

[2] S. H. A. Ahmad, M. Liu, T. Javadi, Q. Zhao, and B. Krishnamachari, "Optimality of Myopic Sensing in Multi-Channel Opportunistic Access," *IEEE Trans. on Information Theory*, vol. 55, no. 9, pp. 4040-4050, September, 2009

[3] S. Murugesan, P. Schniter, and N. B. Shroff, "Multiuser Scheduling in a Markov-modeled Downlink using Randomly Delayed ARQ Feedback," *IEEE Trans. on Information Theory*, submitted Feb 2010, revised November 2010.

[4] E. Gilbert, "Capacity of a burst-noise channel," *Bell Systems Technical Journal*, vol. 39, pp. 1253-1266, 1960.

[5] V. H. Mac Donald, "The Cellular Concept,"*The Bell System Technical Journal*, vol. 58, no. 1, pp. 15-41, Jan 1979.

[6] E. J. Sondik, "The optimal control of partially observable Markov processes," *PhD Thesis*, Stanford University, 1971.

[7] P. Whittle, "Restless bandits: Activity allocation in a changing world," *Journal of Applied Probability*, 25A:287 - 298, 1988.

[8] J. Li and N. B. Shroff and E. K. P. Chong, "A Reduced-Power Channel Reuse Scheme for Wireless Packet Cellular Networks," *IEEE/ACM Trans. on Networking*, vol. 7, no. 6, pp. 818-832, Dec 1999.

[9] X. Wu, A. Das, J. Li and R. Laroia, "Fractional Power Reuse In Cellular Networks," *Proc. Allerton Conf. on Communication, Control, and Computing*, (Monticello, IL), Oct. 2006.

[10] S. Jing, D. N. C. Tse, J. Hou, J. B. Soriaga, J. E. Smee, R. Padovani, "Multi-Cell Downlink Capacity with Coordinated Processing," *Proc. Information Theory and Applications Workshop (ITA),*" San Diego, CA, Jan 2007.

[11] S. Murugesan, "Opportunistic Scheduling Using Channel Memory in Markov-modeled Wireless Networks," *PhD Thesis*, The Ohio State University, 2010.

[12] K. D. Glazebrook, J. Nino Mora, and P. S. Ansell, "Index policies for a class of discounted restless bandits," *Advances in Applied Probability*, 34(4):754774, 2002.