# Channel Probing for Opportunistic Access with Multi-channel Sensing

Keqin Liu,    Qing Zhao
University of California, Davis, CA 95616
kqliu@ucdavis.edu, qzhao@ece.ucdavis.edu

***Abstract*—We consider an opportunistic communication system consisting of multiple independent channels with time-varying states. We formulate the problem of optimal sequential channel selection as a restless multi-armed bandit process, for which a powerful policy—Whittle's index policy—can be implemented based on the indexability of the system. We obtain Whittle's index in closed-form under the average reward criterion, which leads to the direct implementation of Whittle's index policy. To evaluate the performance of Whittle's index policy, we provide simple algorithms to calculate an upper bound of the optimal performance. The tightness of the upper bound and the near-optimal performance of Whittle's index policy are illustrated with simulation examples. When channels are stochastically identical, we show that Whittle's index policy is equivalent to the myopic policy, which has a simple and robust structure. Based on this structure, we establish the approximation factors of the performance of Whittle's index policy. Furthermore, we show that Whittle's index policy is optimal under certain conditions.**

***Index Terms*—Multi-channel opportunistic access, restless multi-armed bandit, Whittle's index, indexability**

## I. Introduction

### A. Multichannel Opportunistic Access

Consider a system consisting of $N$ independent channels. We adopt the Gilbert-Elliot channel model [1], where the state of a channel—"good" or "bad"—evolves as a Markov chain from slot to slot. Due to limited sensing, a user can only sense and access $K$ of these $N$ channels in each slot and accrue rewards determined by the states of the chosen channels. The objective is to design an optimal sensing policy to maximize the long-run reward (*i.e.*, throughput). We formulate the problem as a Restless Multi-armed Bandit Process (RMBP) [2]. Unfortunately, a general RMBP has been shown to be PSPACE-hard [3]. By considering the Lagrangian relaxation of the problem, Whittle proposed a heuristic index policy for RMBP [2], which is optimal under the relaxed constraint on the average number of activated arms over the infinite horizon. Under the strict constraint that exactly $K$ arms are to be activated at each time, Whittle's index policy has been shown to be asymptotically $(N \to \infty)$ optimal under certain conditions [4]. In the finite regime, extensive empirical studies have demonstrated the near-optimal performance of Whittle's index policy, see, for example, [5], [6].

However, not every RMBP has a well-defined Whittle's index; those that admit Whittle's index policy are called *indexable* [2]. The indexability of an RMBP is often difficult to establish, and computing Whittle's index can be complex, often relying on numerical approximations.

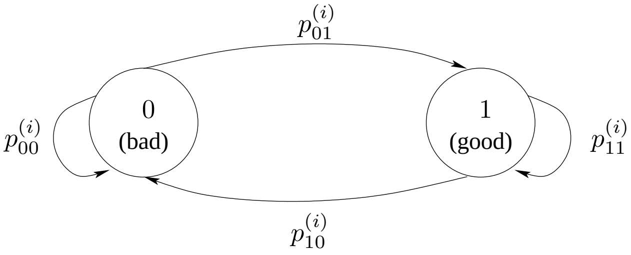

Fig. 1. The Gilber-Elliot channel model.

### B. Contribution

We formulate the design of the optimal sensing policy as an RMBP which has an uncountable state space. By exploiting the rich structure of the problem, we extend the indexability under the discounted reward criterion established in [7] to the average reward criterion. Furthermore, we obtain the closed-form Whittle's index which is the limit of that under the discounted reward criterion as the discount factor goes to 1. Whittle's index policy can then be implemented with simple evaluations of the closed-form expressions.

To develop the performance bound of Whittle's index policy, we consider the same RMBP but under the relaxed constraint on the average number of channels to sense. The optimal performance of this bandit process is thus an upper bound of the original one under the strict constraint. In this paper, we provide simple algorithms to evaluate this upper bound. The strong performance of Whittle's index policy can thus be demonstrated by comparing with this upper bound instead of the optimal performance of the original bandit process that requires the exponential complexity. The tightness of the upper bound and the powerful performance of Whittle's index policy are illustrated in simulation examples.

When channels are stochastically identical, we show that Whittle's index policy coincides with the myopic policy. In this case, Whittle's index policy has a simple and robust structure that does not need the update of the belief values or the precise knowledge of the transition probabilities of the underlying Markovian channel model. This structure automatically tracks the model variations as long as the order of the transition

[0]This work was supported by the Army Research Laboratory CTA on Communication and Networks under Grant DAAD19-01-2-0011 and by the National Science Foundation under Grants ECS-0622200 and CCF-0830685.

probabilities is unchanged. Based on the structure, we build a lower bound of the performance by Whittle's index policy, leading to approximation factors of its performance. We show that Whittle's index policy achieves at least $\frac{K}{N}$ the optimal performance. Furthermore, Whittle's index policy is optimal when $K = N-1$ or $N$. The factor $\frac{K}{N}$ can be further improved when each channel is negatively correlated. In this case, Whittle's index policy achieves at least $\max\{\frac{1}{2}, \frac{K}{N}\}$ the optimal performance.

### C. Related Work

Multichannel opportunistic access in the context of cognitive radio systems has been studied in [8], [9] where the problem is formulated as a Partially Observable Markov Decision Process (POMDP) to take into account potential correlations among channels. For stochastically identical and independent channels and under the assumption of single-channel sensing ($K = 1$), the structure, optimality, and performance of the myopic policy have been investigated in [10], where the semi-universal structure of the myopic policy was established for all $N$ and the optimality of the myopic policy proved for $N = 2$. In a recent work [11], the optimality of the myopic policy was extended to $N > 2$ under the condition of $p_{11} \geq p_{01}$. In this paper, we establish the equivalence relationship between the myopic policy and Whittle's index policy when channels are stochastically identical. This equivalence relationship shows that the results obtained in [10], [11] for the myopic policy are directly applicable to Whittle's index policy. Furthermore, we extend these results to multichannel sensing ($K > 1$).

In [7], we have established the indexability and obtained closed-form Whittle's index of the RMBP for multichannel opportunistic access under the discounted reward criterion. These results are key to analyzing the indexability and solve for Whittle's index of the RMBP under the average discounted reward criterion. In [12], Le Ny *et al.* have considered the same class of RMBP motivated by the applications of target tracking. They have independently established the indexability and obtained the closed-form expressions for Whittle's index under the discounted reward criterion. However, the approach in [7] is different from that used in [12].

In the general context of RMBP, there is a rich literature on indexability. See [13] for the linear programming representation of conditions for indexability and [6] for examples of specific indexable restless bandit processes. Constant-factor approximation algorithms for RMBP have also been explored in the literature. For the same class of RMBP as considered in this paper, Guha and Munagala [14] have developed a constant-factor (1/68) approximation via LP relaxation under the condition that $p_{11} > \frac{1}{2} > p_{01}$ for each channel. In [15], Guha *et al.* have developed a factor-2 approximation policy via LP relaxation for the so-called monotone bandit processes.

## II. Multi-channel Opportunistic Access and Restless Bandit Formulation

Consider $N$ independent Gilbert-Elliot channels, each with transmission rate $B_i (i = 1, \cdots, N)$. The state of channel $i$—"good"(1) or "bad"(0)— evolves from slot to slot as a Markov chain with transition matrix $\mathbf{P}_i = \{p_{j,k}^{(i)}\}_{j,k\in\{0,1\}}$ as shown in Fig. 1. At the beginning of slot $t$, the user selects $K$ out of $N$ channels to sense. If the state $S_i(t)$ of the sensed channel $i$ is 1, the user transmits and collects $B_i$ units of reward in this channel. Otherwise, the user collects no reward in this channel. Let $U(t)$ denote the set of $K$ channels chosen in slot $t$. The reward obtained in slot $t$ is thus given by

$$R_{U(t)}(t) = \Sigma_{i\in U(t)} S_i(t) B_i.$$

The channel states $[S_1(t), ..., S_N(t)] \in \{0,1\}^N$ are not directly observable before the sensing action is made. The user can, however, infer the channel states from its decision and observation history. It has been shown that a sufficient statistic for optimal decision making is given by the conditional probability that each channel is in state 1 given all past decisions and observations [16]. Referred to as the belief vector, this sufficient statistic is denoted by $\Omega(t) \stackrel{\Delta}{=} [\omega_1(t), \cdots, \omega_N(t)]$, where $\omega_i(t)$ is the conditional probability that $S_i(t) = 1$. Given the sensing action $U(t)$ and the observation in slot $t$, the belief state in slot $t+1$ can be obtained recursively as follows:

$$\omega_i(t+1) = \begin{cases} p_{11}^{(i)}, & i \in U(t), S_i(t) = 1 \\ p_{01}^{(i)}, & i \in U(t), S_i(t) = 0 \\ \mathcal{T}(\omega_i(t)), & i \notin U(t) \end{cases}, \qquad (1)$$

where $\mathcal{T}(\omega_i(t)) \stackrel{\Delta}{=} \omega_i(t) p_{11}^{(i)} + (1-\omega_i(t)) p_{01}^{(i)}$ denotes the one-step belief update for unobserved channels.

We thus have an RMBP formulation, where each channel is considered as an arm and the state of arm $i$ in slot $t$ is the belief state $\omega_i(t)$. The user chooses an action $U(t)$ consisting of $K$ arms to activate (sense) in each slot, while other arms are made passive (unobserved). The states of both active and passive arms change as given in (1). A policy $\pi : \Omega(t) \rightarrow U(t)$ is a function that maps from the belief vector $\Omega(t)$ to the action $U(t)$ in slot $t$. Our objective is to design the optimal policy $\pi^*$ to maximize the expected average reward over the infinite horizon:

$$\max_{\pi} \{\mathbb{E}_{\pi}[\lim_{T\rightarrow\infty} \frac{1}{T} \Sigma_{t=1}^{T} R_{\pi(\Omega(t))}(t) | \Omega(1)]\}. \qquad (2)$$

## III. Whittle's Index Policy

To introduce indexability and Whittle's index policy, it suffices to consider a single arm. Assume a constant subsidy $m$ (*subsidy for passivity*) is obtained whenever the arm is made passive. In each slot, the user chooses one of two possible actions—$u \in \{0 \text{ (passive)}, 1 \text{ (active)}\}$—to make the arm passive or active. The objective is to decide whether to activate the arm in each slot to maximize the average reward. Let $u_m^*(\omega)$ denote the optimal action for belief state $\omega$ under subsidy $m$. The passive set $\mathcal{P}(m)$ under subsidy $m$ is given by

$$\mathcal{P}(m) \quad = \quad \{\omega : \ u_m^*(\omega) = 0\}. \qquad (3)$$

*Definition 1:* An arm is *indexable* if the passive set $\mathcal{P}(m)$ of the corresponding single-armed bandit process with subsidy $m$ monotonically increases from $\emptyset$ to the whole state space $[0,1]$ as $m$ increases from $-\infty$ to $+\infty$. An RMBP is indexable if every arm is indexable.

Under the indexability condition, Whittle's index is defined as follows.

*Definition 2:* If an arm is indexable, its *Whittle's index* $W(\omega)$ of the state $\omega$ is the infimum subsidy $m$ such that it is optimal to make the arm passive at $\omega$. Equivalently, Whittle's index $W(\omega)$ is the infimum subsidy $m$ that makes the passive and active actions equally rewarding.

$$W(\omega) \quad = \quad \inf_m \{m : \; u_m^*(\omega) = 0\}. \tag{4}$$

In each slot, Whittle's index policy senses the $K$ channels with the largest Whittle's indices.

### *A. Indexability and Closed-form Whittle's Index*

Our analysis on the indexability and Whittle's index hinges on the previous results under the discounted reward criterion [7]. First, we present a general result by Dutta [17] on the relationship between the value function and the optimal policy under the discounted reward criterion and those under the average reward criterion. This result allows us to study Whittle's index policy under the average reward criterion by examining its limiting behavior as the discount factor $\beta \to 1$.

*Dutta's Theorem* [17]. Let $\mathcal{F}$ be the belief space of a POMDP and $V_\beta(\Omega)$ the value function (*i.e.,* the maximum expected total discounted reward starting from belief $\Omega \in \mathcal{F}$) with discount factor $\beta$. The POMDP satisfies the value boundedness condition if there exist a belief $\Omega'$, a real-valued function $c_1(\Omega) : \mathcal{F} \to \mathcal{R}$, and a constant $c_2 < \infty$ such that

$$c_1(\Omega) \leq V_\beta(\Omega) - V_\beta(\Omega') \leq c_2,$$

for any $\Omega \in \mathcal{F}$ and $\beta \in [0,1)$. Under the value-boundedness condition, if a series of optimal policies $\pi_{\beta_k}$ for a POMDP with discount factor $\beta_k$ pointwise converges to a limit $\pi^*$ as $\beta_k \to 1$, then $\pi^*$ is the optimal policy for the POMDP under the average reward criterion. Furthermore, let $J(\Omega)$ denote the maximum expected average reward over the infinite horizon starting from the initial belief $\Omega$, we have

$$J(\Omega) = \lim_{\beta_k \to 1} (1-\beta_k) V_{\beta_k}(\Omega)$$

and $J(\Omega) = J$ is independent of the initial belief $\Omega$.

*Lemma 1:* The single-armed bandit process with subsidy under the discounted reward criterion satisfies the value-boundedness condition.

*Proof:* Omitted due to space limit. See [18] for details. ■

In [7], we showed that the optimal policy under discounted reward criterion is a threshold policy with threshold $\omega_\beta^*(m)$. By Dutta's theorem and Lemma 1, we can show that the optimal policy for the single-armed bandit process with subsidy under the average reward criterion is also a threshold policy.

*Lemma 2:* Let $\omega_\beta^*(m)$ denote the threshold of the optimal policy for the single-armed bandit process with subsidy $m$ under the discounted reward criterion. Then $\lim_{\beta\to 1} \omega_\beta^*(m)$ exists for any $m$. Furthermore, the optimal policy for the single-armed bandit process with subsidy $m$ under the average reward criterion is also a threshold policy with threshold $\omega^*(m) = \lim_{\beta\to 1} \omega_\beta^*(m)$.

*Proof:* Omitted due to space limit. See [18] for details. ■

Based on Lemma 2, the restless multi-armed bandit process is indexable if the threshold $\omega^*(m)$ of the optimal policy is monotonically increasing with subsidy $m$. Based on the indexability property of the RMBP under the discounted reward criterion established in [7], we have $\omega_\beta^*(m)$ increases with $m$. Since $\omega^*(m) = \lim_{\beta\to 1} \omega_\beta^*(m)$, it is easy to see that $\omega^*(m)$ also increases with $m$. The bandit is thus indexable. Furthermore, Whittle's index is the limit of that under the discounted reward criterion obtained in [7] by letting the discount factor goes to $1$.

*Theorem 1:* The restless multi-armed bandit process is indexable with Whittle's index $W(\omega)$ given below.

- *Case 1: Positively correlated channel* $(p_{11}^{(i)} \geq p_{01}^{(i)})$.

$$W(\omega) = \begin{cases} \omega B_i, \quad \text{if } \omega \leq p_{01}^{(i)} \text{ or } \omega \geq p_{11}^{(i)} \\ \frac{(\omega - \mathcal{T}^1(\omega))(L(p_{01}^{(i)},\omega)+1) + \mathcal{T}^{L(p_{01}^{(i)},\omega)}(p_{01}^{(i)})}{1 - p_{11}^{(i)} + (\omega - \mathcal{T}^1(\omega))L(p_{01}^{(i)},\omega) + \mathcal{T}^{L(p_{01}^{(i)},\omega)}(p_{01}^{(i)})} B_i, \\ \text{if } p_{01}^{(i)} < \omega < \omega_o^{(i)} \\ \frac{\omega}{1 - p_{11}^{(i)} + \omega} B_i, \quad \text{if } \omega_o^{(i)} \leq \omega < p_{11}^{(i)} \end{cases}.$$

- *Case 2: Negatively correlated channel* $(p_{11}^{(i)} < p_{01}^{(i)})$.

$$W(\omega) = \begin{cases} \omega B_i, \quad \text{if } \omega \leq p_{11}^{(i)} \text{ or } \omega \geq p_{01}^{(i)} \\ \frac{\omega + p_{01}^{(i)} - \mathcal{T}^1(\omega)}{1 + p_{01}^{(i)} - \mathcal{T}^1(p_{11}^{(i)}) + \mathcal{T}^1(\omega) - \omega} B_i \\ \text{if } p_{11}^{(i)} < \omega < \omega_o^{(i)} \\ \frac{p_{01}^{(i)}}{1 + p_{01}^{(i)} - \mathcal{T}^1(p_{11}^{(i)})} B_i, \quad \text{if } \omega_o^{(i)} \leq \omega < \mathcal{T}^1(p_{11}^{(i)}) \\ \frac{p_{01}^{(i)}}{1 + p_{01}^{(i)} - \omega} B_i, \quad \text{if } \mathcal{T}^1(p_{11}^{(i)}) \leq \omega < p_{01}^{(i)} \end{cases}.$$

*Proof:* Omitted due to the space limit. See [18] for details. ■

### *B. The Performance of Whittle's Index Policy*

Whittle's index policy is optimal when the constraint on the number of activated arms $K(t)$ $(t \geq 1)$ is relaxed to the following.

$$\mathbb{E}_\pi[\lim_{T\to\infty} \frac{1}{T} \Sigma_{t=1}^T K(t)] = K.$$

Let $\bar{J}(\Omega(1))$ denote the maximum expected average reward that can be obtained under this relaxed constraint. Based on the Lagrangian multiplier theorem, we have [2]

$$\bar{J} = \inf_m \{\Sigma_{i=1}^N J_m^{(i)} - m(N-K)\}, \tag{5}$$

where $J_m^{(i)}$ is the maximum expected average reward of the single-armed bandit process with subsidy $m$ that corresponds to the $i$-th channel.

Let $J(\Omega(1))$ denote the maximum expected average reward of the RMBP under the strict constraint that $K(t) = K$ for all $t$. Obviously, $J(\Omega(1)) \leq \bar{J}$. $\bar{J}$ thus provides a performance benchmark for Whittle's index policy under the strict constraint. To evaluate $\bar{J}$, we consider the single-armed bandit with subsidy $m$ under the average reward criterion (assume the bandwidth $B = 1$). The value function $J_m$ and the average passive time $D_m = \frac{d(J_m)}{dm}$ can be obtained in closed-form as shown in Lemma 3 below.

*Lemma 3:* Let $\omega^*(m)$ be the threshold of the optimal policy. We have

$$J_m = \begin{cases} \omega_o, \quad \text{if } \omega^*(m) < \min\{p_{01}, p_{11}\} \\ \frac{(1-p_{11})L(p_{01},\omega^*(m))m + \mathcal{T}^{L(p_{01},\omega^*(m))}(p_{01})}{(1-p_{11})(L(p_{01},\omega^*(m))+1) + \mathcal{T}^{L(p_{01},\omega^*(m))}(p_{01})}, \\ \text{if } p_{01} \leq \omega^*(m) < \omega_o \\ \frac{p_{01}m + p_{01}}{1+2p_{01}-\mathcal{T}^1(p_{11})}, \quad \text{if } p_{11} \leq \omega^*(m) < \mathcal{T}^1(p_{11}) \\ m, \quad \text{other cases} \end{cases}$$

and

$$D_m = \begin{cases} 0, \quad \text{if } \omega^*(m) < \min\{p_{01}, p_{11}\} \\ \frac{(1-p_{11})L(p_{01},\omega^*(m))}{(1-p_{11})(L(p_{01},\omega^*(m))+1) + \mathcal{T}^{L(p_{01},\omega^*(m))}(p_{01})}, \\ \text{if } p_{01} \leq \omega^*(m) < \omega_o \\ \frac{p_{01}}{1+2p_{01}-\mathcal{T}^1(p_{11})}, \quad \text{if } p_{11} \leq \omega^*(m) < \mathcal{T}^1(p_{11}) \\ 1, \quad \text{other cases} \end{cases}$$

Furthermore, $D_m$ is piecewise constant and increasing with $m$.

*Proof:* By Dutta's Theorem and Lemma 1, $J_m$ can be obtained from the limit of the value function under the discounted reward criterion as the discount factor goes to 1. $D_m$ can be obtained directly from the closed form $J_m$. The monotonicity of $D_m$ follows from the convexity of $J_m$. See [18] for details. ■

Based on the closed-form $D_m$ given in Lemma 3, the subsidy $m^*$ that achieves the infimum in (5) is the supremum value of $m \in [0, 1]$ satisfying $\Sigma_{i=1}^{N} D_m^{(i)} \leq N - K$. After obtaining $m^*$, it is easy to calculate the infimum according to the closed-form $J_m$ given in Lemma 3. Based on the piecewise constant property of $D_m$, we can design an algorithm which runs in $O(N(\log N)^2)$ time to compute $\bar{J}$ within $\epsilon$-accuracy for any $\epsilon > 0$. Furthermore, when every channel satisfies $p_{11} < p_{01}$, we can compute $\bar{J}$ without error with complexity $O(N^2 \log N)$. See [18] for the detailed algorithms.

Figure 2 below shows an example of the performance of Whittle's index policy. We notice that the performance loss by Whittle's index policy is negligible and the upper bound of the optimal policy is tight.

## IV. Stochastically Identical Channels

Based on the monotonicity of Whittle's index with the belief state, Whittle's index policy is equivalent to the myopic policy for stochastically identical arms. A myopic policy ignores the impact of the current action on the future reward, focusing solely on maximizing the expected immediate reward. The myopic action $\hat{U}(t)$ under the belief state $\Omega(t) = [\omega_1(t), \cdots, \omega_N(t)]$ is given by

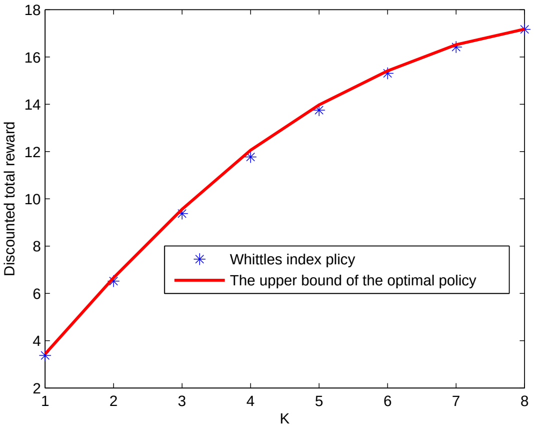

Fig. 2. The Performance of Whittle's index policy ($N = 8$).

$$\hat{U}(t) = \arg\max_{U(t)} \Sigma_{i \in U(t)}\, \omega_i(t) B_i. \tag{6}$$

We can analyze Whittle's index policy by focusing on the myopic policy which has a much simpler index form.

### A. The Structure of Whittle's Index Policy

The implementation of Whittle's index policy can be described with a queue structure. Specifically, all $N$ channels are ordered in a queue, and in each slot, those $K$ channels at the head of the queue are sensed. The initial channel ordering $\mathcal{K}(1)$ is determined by the initial belief vector as given below.

$$\omega_{n_1}(1) \geq \cdots \geq \omega_{n_N}(1) \implies \mathcal{K}(1) = (n_1, \cdots, n_N).$$

Based on the observations, channels are reordered at the end of each slot according to the following simple rules. When $p_{11} \geq p_{01}$, the channels observed in state $1$ will stay at the head of the queue while the channels observed in state $0$ will be moved to the end of the queue. When $p_{11} < p_{01}$, the channels observed in state $0$ will stay at the head of the queue while the channels observed in state $1$ will be moved to the end of the queue. The order of the unobserved channels are reversed.

Based on the structure, Whittle's index policy can be implemented without knowing the channel transition probabilities except the order of $p_{11}$ and $p_{01}$. As a result, Whittle's index policy is robust against model mismatch and automatically tracks variations in the channel model provided that the order of $p_{11}$ and $p_{01}$ remains unchanged. As show in Fig. 3, the transition probabilities change abruptly in the fifth slot, which corresponds to an increase in the occurrence of good channel state in the system. From this figure, we can observe, from the change in the throughput increasing rate, that Whittle's index policy effectively tracks the model variations.

### B. Optimality and Approximation Factor

Based on the simple structure of Whittle's index policy for stochastically identical channels, we can obtain a lower

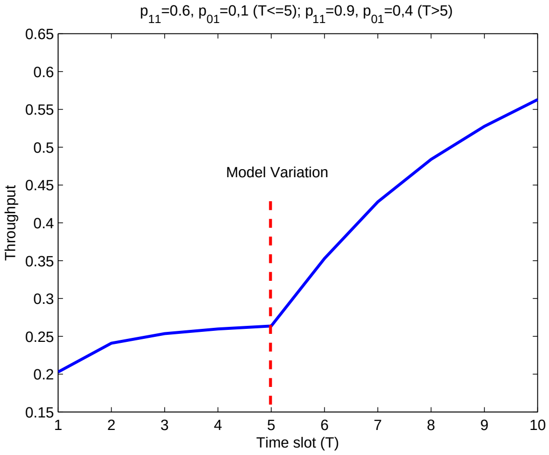

Fig. 3. Tracking the change in channel transition probabilities occurred at $t = 6$.

bound of its performance. Combining this lower bound and the upper bound shown in Sec. III-B, we further obtain the approximation factor of the performance by Whittle's index policy, which are independent of channel parameters. Recall that $J$ denote the average reward achieved by the optimal policy. Let $J_w$ denote the average reward achieved by Whittle's index policy,

*Theorem 2: Lower and Upper Bounds of The Performance of Whittle's Index Policy*

- *Case 1:* $p_{11} \geq p_{01}$

$$\frac{K\mathcal{T}^{\lfloor \frac{N}{K} \rfloor -1}(p_{01})}{1-p_{11}+\mathcal{T}^{\lfloor \frac{N}{K} \rfloor -1}(p_{01})} \leq J_w \leq J \leq \min\{\frac{K\omega_o}{1-p_{11}+\omega_o},\ \omega_o N\}$$

- *Case 2:* $p_{11} < p_{01}$

$$\frac{Kp_{01}}{1-\mathcal{T}^{2\lfloor \frac{N}{K} \rfloor -2}(p_{11})+p_{01}} \leq J_w \leq J \leq \min\{\frac{Kp_{01}}{1-\mathcal{T}^{1}(p_{11})+p_{01}},\ \omega_o N\}$$

*Proof:* The upper bound of $J$ is obtained from the upper bound of the optimal performance for generally non-identical channels as given in (5). The lower bound of $J_w$ is obtained from the structure of Whittle's index policy. See [18] for details. ■

*Corollary 1:* Let $\eta = \frac{J_w}{J}$ be the approximation factor defined as the ratio of the performance by Whittle's index policy to the optimal performance. We have

$$p_{11} \geq p_{01} \qquad\qquad p_{11} < p_{01}$$

$$\begin{cases} \eta = 1, & \text{for } K = 1, N-1, N \\ \eta \geq \frac{K}{N}, & \text{o.w.} \end{cases} \qquad \begin{cases} \eta = 1, & \text{for } K = N-1, N \\ \eta \geq \max\{\frac{1}{2}, \frac{K}{N}\}, & \text{o.w.} \end{cases}$$

*Proof:* Omitted due to the space limit. See [18] for details. ■

From Corollary 1, Whittle's index policy is optimal when $K = 1$ (for positively correlated channels) and $K = N - 1$. The optimality for $K = N$ is trivial. We point out that for a general $K$, numerical examples have shown that actions given by Whittle's index policy match with the optimal actions for randomly generated sample paths, suggesting the optimality of Whittle's index policy.

## V. Conclusion

In this paper, we extended the indexability and Whittle's index in [7] to the average reward criteria. We provided simple algorithms to evaluate an upper bound of the optimal performance. When channels are stochastically identical, we have shown that Whittle's index policy coincides with the myopic policy. Based on this equivalency, we have established the semi-universal structure and the optimality of Whittle index policy under certain conditions.

## References

[1] E. N. Gilbert, "Capacity of burst-noise channels," Bell Syst. Tech. J., vol. 39, pp. 1253-1265, Sept. 1960.

[2] P. Whittle, "Restless bandits: Activity allocation in a changing world", in *Journal of Applied Probability*, Vol. 25, 1988.

[3] C. H. Papadimitriou and J. N. Tsitsiklis, "The Complexity of Optimal Queueing Network Control," in *Mathematics of Operations Research*, Vol. 24, No. 2, May 1999, pp. 293-305.

[4] R. R. Weber and G. Weiss, "On an Index Policy for Restless Bandits," in *Journal of Applied Probability*, Vol.27, No.3, pp. 637-648, Sep 1990.

[5] P. S. Ansell, K. D. Glazebrook, J.E. Nio-Mora, and M. O'Keeffe, "Whittle's index policy for a multi-class queueing system with convex holding costs," in *Math. Meth. Operat. Res.* 57, 21–39, 2003.

[6] K. D. Glazebrook, D. Ruiz-Hernandez, and C. Kirkbride, "Some Indexable Families of Restless Bandit Problems ," in *Advances in Applied Probability*, 38:643-672, 2006.

[7] K. Liu and Q. Zhao, "A Restless Bandit Formulation of Opportunistic Access: Indexablity and Index Policy," in *Proc. of the 5th IEEE Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON) Workshops*, June, 2008.

[8] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized Cognitive MAC for Opportunistic Spectrum Access in Ad Hoc Networks: A POMDP Framework," in *IEEE Journal on Selected Areas in Communications (JSAC): Special Issue on Adaptive, Spectrum Agile and Cognitive Wireles Networks* , April 2007.

[9] Y. Chen, Q. Zhao, and A. Swami, "Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors," in *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 2053-2071, May, 2008

[10] Q. Zhao, B. Krishnamachari, and K. Liu, "On Myopic Sensing for Multi-Channel Opportunistic Access: Structure, Optimality, and Performance," *to appear in IEEE Trans. Wireless Communications*, Dec., 2008.

[11] S. H. Ahmad, M. Liu, T. Javadi, Q. Zhao and B. Krishnamachari, "Optimality of Myopic Sensing in Multi-Channel Opportunistic Access," submitted to IEEE Transactions on Information Theory, May, 2008, available at http://arxiv.org/abs/0811.0637

[12] J. Le Ny, M. Dahleh, E. Feron, "Multi-UAV Dynamic Routing with Partial Observations using Restless Bandit Allocation Indices,", in *Proceedings of the 2008 American Control Conference*, Seattle, WA, June 2008.

[13] J. E. Nio-Mora, "Restless bandits, partial conservation laws and indexability," in *Advances in Applied Probability*, 33:7698, 2001.

[14] S. Guha and K. Munagala, "Approximation algorithms for partial-information based stochastic control with Markovian rewards," in *Proc. 48th IEEE Symposium on Foundations of Computer Science (FOCS)*, 2007.

[15] S. Guha, K. Munagala, "Approximation Algorithms for Restless Bandit Problems," http://arxiv.org/abs/0711.3861.

[16] E. J. Sondik, " The Optimal Control at Partially Observable Markov Processes Over the Infinite Horizon: Discounted Costs," in *Operations Research*, Vol.26, No.2 (Mar. - Apr,. 1978), 282 - 304.

[17] P. K. Dutta, "What do discounted optima converge to? A theory of discount rate asymptotics in economic models," in Journal of Economic Theory 55, pp. 6494, 1991.

[18] K. Liu and Q. Zhao, "Indexability of Restless Bandit Problems and Optimality of Whittle's Index for Dynamic Multichannel Access," submitted to IEEE Transactions on Information Theory, November, 2008. Available at http://arxiv.org/abs/0810.4658.