Channel Probing in Communication Systems: Myopic Policies Are Not Always Optimal

Matthew Johnston, Eytan Modiano
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA
Email: {mrj, modiano}@mit.edu

Isaac Keslassy
Dep. of Electrical Engineering
Technion
Haifa, Israel
Email: isaac@ee.technion.ac.il

***Abstract*—We consider a multi-channel communication system in which a transmitter has access to a large number of channels, but does not know the state of these channels. We model channel state using an ON/OFF Markovian model, and allow the transmitter to probe one of the channels at predetermined probing intervals to decide over which channel to transmit. For models in which the transmitter must send over the probed channel, it has been shown that a myopic policy that probes the channel most likely to be ON is optimal. In this work, we allow the transmitter to select a channel over which to transmit that is not necessarily the one it probed. We show that the myopic policy is not optimal, and propose a simple alternative probing policy, which achieves a higher per-slot expected throughput. Finally, we consider the case where there is a fixed cost associated with probing and derive optimal probing intervals.**

I. Introduction

Consider a communication system in which a transmitter has access to multiple channels over which to communicate. The state of each channel evolves independently from all other channels, and the transmitter does not know the channel states *a priori*. The transmitter is allowed to probe a single channel after a predefined time interval to learn the channel state at the current time, which is either ON or OFF. Using the information obtained from the channel probes, the transmitter selects a channel in each time-slot over which to transmit, with the goal of maximizing throughput, or the number of successful transmissions.

This framework applies broadly to many opportunistic communication systems, in which there exists a tradeoff between overhead and performance. When there is a large number of channels over which to transmit, or a large number of users to transmit to, it is impractical to learn the channel state information (CSI) of every channel before scheduling a transmission; consequently, it is only practical for the transmitter to obtain partial channel state information, and use that partial CSI to make a decision. Therefore, the transmitter must decide *how much* information, and *which* information is needed in order to make efficient scheduling decisions.

Several works have studied channel probing policies in multichannel communication problems [1], [2], [3], [4], [5], [6], [7]. Of particular interest is the work in [8] and [9], in which the authors assume that after a channel is probed, the transmitter must transmit on that channel. They show that the optimal probing policy is a myopic policy, which probes the channel most likely to be ON.

This work was supported by NSF Grants CNS-0915988 and CNS-1217048, and ARO MURI Grant W911NF-08-1-0238.

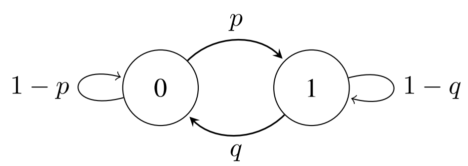

Fig. 1. Markov Chain describing the channel state evolution of each independent channel. State 0 corresponds to an OFF channel, while state 1 corresponds to an ON channel.

In this work, we show that when the transmitter is able to transmit over a channel other than that which was probed, the myopic probing policy in [8], [9] is no longer optimal. Specifically, we prove using renewal theory that a simple policy, namely the policy which probes the channel that is second most likely to be ON, has a higher per-slot expected throughput. We conjecture that this policy is in fact optimal for our setting. Additionally, we show that a round robin probing policy has the same expected throughput as the myopic policy in [8], [9]. Finally, we extend these results to the case where there is a fixed cost associated with a probe, and derive the optimal dynamic probing epochs.

The remainder of this paper is organized as follows. We describe the model and problem formulation in detail in Section II. In Section III, we prove that the myopic policy is suboptimal by analytically computing the expected per-slot throughput of two other policies. In Section IV, we solve for the optimal probing intervals when a fixed cost is associated with probing. Lastly, we conclude in Section V.

II. System Model

Consider a transmitter and a receiver that can communicate using a large number of independent channels. At every time slot, each channel is either in an OFF state or an ON state. Channels are i.i.d. with respect to each other, and evolve across time according to a discrete time Markov process described by Figure 1.

At each time slot, the transmitter chooses a single channel over which to transmit. If that channel is in the

978-1-4799-0446-4/13/$31.00

ON state, then the transmission is successful; otherwise, the transmission fails. We assume the transmitter does not receive feedback regarding previous transmissions[1]. The objective is to maximize the expected sum-rate throughput, equal to the number of successful transmissions over time.

Furthermore, at predefined epochs of T slots, the transmitter probes the receiver for the state of one of the channels at the current time, which is delivered instantaneously. The transmitter then uses the history of channel probes to make a scheduling decision.

A. Notation

Let $S_i(t)$ be the state of channel i at time t. The transmitter has an estimate of this state based on previous probes and the channel state distribution. Define the *belief* of a channel to be the probability that a channel is ON given the history of channel probes. For any channel i that was last probed k slots ago and was in state s_i, the belief x_i is given by

$$\begin{aligned} x_i(t) &= \mathbf{P}\big(\text{Channel } i \text{ is ON}|\text{probing history}\big) \\ &= \mathbf{P}\big(S_i(t) = 1 | S_i(t-k) = s_i\big) \end{aligned} \tag{1}$$

where the second equality follows from the Markov property of the channel state process. The above probability is derived from the k-step transition probabilities of the Markov chain in Figure 1:

$$\begin{aligned} p_{00}^k &= \frac{q + p(1-p-q)^k}{p+q}, p_{01}^k = \frac{p - p(1-p-q)^k}{p+q} \\ p_{10}^k &= \frac{q - q(1-p-q)^k}{p+q}, p_{11}^k = \frac{p + q(1-p-q)^k}{p+q}. \end{aligned} \tag{2}$$

Throughout this work, we assume that $p \leq \frac{1}{2}$ and $q \leq \frac{1}{2}$, corresponding to channels with "positive memory." As the CSI of a channel grows stale, the probability π that the channel is ON is given by the stationary distribution of the chain in Figure 1.

$$\lim_{k\to\infty} p_{01}^k = \lim_{k\to\infty} p_{11}^k = \pi = \frac{p}{p+q}. \tag{3}$$

B. Optimal Scheduling

Since the objective is to maximize the expected sum-rate throughput, the optimal transmission decision at each time slot is given by the maximum likelihood (ML) rule, which is to transmit over the channel that is most likely to be ON, i.e. the channel with the highest belief. The expected throughput in a time slot is therefore given by

$$\max_i x_i(t). \tag{4}$$

Following the above assumptions, the optimal scheduling decision remains the same in between channel probes.

C. Infinite Channel Assumption

In the model used throughout this paper, there is an infinite number of channels available to the transmitter. This assumption affords various simplifications to our state space model. Whenever a probed channel is OFF, it is effectively removed from the system. This is because there always exists a channel which has not been probed for an infinitely long amount of time. Therefore, its belief equals the steady state probability π, and $p_{01}^k \leq \pi$ for all k.

[1]If such feedback exists in the form of higher layer acknowledgements, it arrives after a significant delay and is not useful for learning the channel state.

III. Average Throughput Analysis

To begin, consider the *probe best policy*, which probes the channel with the highest belief. This policy is also referred to as a myopic or greedy policy, as it maximizes the immediate reward without regard to future rewards. Such a policy is advantageous, as the channel with the highest belief is the most likely to be ON at the current time. Recall that this policy is shown to be optimal for the model in [8], [9]. For our model, we have the following result.

Lemma 1. *Assume the transmitter makes probing decisions every T slots according to the probe best policy. The expected per-slot throughput is given by*

$$\mathbb{E}[\mathit{Thpt}] = \pi + \frac{\pi p_{10}^T}{T(p+q)(p_{10}^T + \pi)} \tag{5}$$

Proof: We use renewal theory to compute the average throughput. Under the probe best policy, only one channel can have belief greater than π, since the policy won't probe other channels and therefore won't change its belief about these channels. Define a renewal to occur before a channel with belief π is probed. Therefore, if a channel is probed and if it is OFF, it is removed from the system and a renewal occurs. If the channel is ON, that channel is probed at all future probing instances until it is found to be OFF. The expected inter-renewal time $\bar{X}_B$ is given by

$$\bar{X}_B = T + T\pi\mathbb{E}(N) \tag{6}$$

where N is the number of times an ON channel is probed before it is OFF, and is geometrically distributed with parameter p_{10}^T. Equation (6) reduces to

$$\bar{X}_B = T + \frac{\pi T}{p_{10}^T}. \tag{7}$$

The expected reward $\bar{R}_B$ incurred over a renewal interval is πT for the interval immediately after the OFF probe, and $\sum_{i=0}^{T-1} p_{11}^i$ for each subsequent ON probe.

$$\bar{R}_B = \pi T + \pi\mathbb{E}\big[N\big] \sum_{i=0}^{T-1} p_{11}^i = \pi T + \frac{\pi \sum_{i=0}^{T-1} p_{11}^i}{p_{10}^T} \tag{8}$$

Using results from renewal-reward theory [10], the average per-slot reward is given by

$$\frac{\bar{R}_B}{\bar{X}_B} = \frac{\pi T p_{10}^T + \pi \sum_{i=0}^{T-1} p_{11}^i}{T p_{10}^T + \pi T} = \pi + \frac{\pi p_{10}^T}{T(p+q)(p_{10}^T + \pi)} \tag{9}$$

■

Observe that the per-slot throughput is always larger than π, and decreases toward π as T increases. The probe best policy is attractive because it maximizes the immediate reward. However, the drawback of this policy is that when the probed channel is OFF, the transmitter has no knowledge of the state of the other channels as it searches for an ON channel. Consequently, during these periods it has a low expected reward.

A. Probe Second Best Policy

Now, consider a simple alternative policy, the *probe second best* policy, which at each time slot probes the channel with the second highest belief, and transmits on the channel with the highest resulting belief. Consider channel state beliefs $x_1, x_2, x_3, \dots$ where $x_1 \geq x_2 \dots \geq x_i \dots \geq \pi$. The immediate reward from probing the best channel is given by

$$x_1 + (1 - x_1)x_2 = x_1 + x_2 - x_1 x_2, \tag{10}$$

and the immediate reward from probing the second best channel is given by

$$x_2 + (1 - x_2)x_1 = x_1 + x_2 - x_1 x_2. \tag{11}$$

The probe second best policy has the same immediate reward as the probe best policy; however, since two channels can have belief greater than π under the probe second best policy, when the probe second best policy probes an OFF channel, the transmitter can transmit over the channel with the highest belief, while looking for another ON channel, which results in a higher expected throughput over that interval. This is a strict improvement over the probe best policy, which transmits on a channel with belief equal to the steady state probability π. It is this intuition that leads us to consider the probe second best policy. The following theorem states that the probe second best policy yields a higher throughput than the probe best policy.

Theorem 1. *The average reward of the probe second best policy is greater than that of the probe best policy, for all fixed probing intervals T.*

Proof: Theorem 1 is proved using renewal theory to compute the average throughput of the probe second best policy, and comparing it to that of the probe best policy. The key to the proof is in the correct definition of the renewal interval. We define a renewal to occur when the best channel has belief p_{11}^{2T}, and the second best channel (and each other channels) has belief π. During a renewal interval, the transmitter probes new channels until it finds an ON channel, and then probes the second best channel until it finds an OFF channel. After this, a renewal occurs before the next probe. This definition ensures that each inter-renewal period is i.i.d. An example renewal interval is shown in Table I.

Time	0	T	$2T$	$3T$	$4T$	$5T$	$6T$
Best Channel Belief	p_{11}^{2T}	p_{11}^{3T}	p_{11}^{4T}	p_{11}^{T}	p_{11}^{T}	p_{11}^{T}	p_{11}^{2T}
Second Best Belief	π	π	π	p_{11}^{5T}	p_{11}^{2T}	p_{11}^{2T}	π
Probe Result	0	0	1	1	1	0	-

TABLE I
EXAMPLE RENEWAL INTERVAL STARTING AT TIME 0 AND RENEWING AT TIME $6T$. AT EACH PROBING INTERVAL, THE SECOND BEST CHANNEL IS PROBED.

The expected inter-renewal time is given by $T\mathbb{E}(N_1 + N_2)$, where N_1 is the number of probes it takes to find an ON channel, and is geometrically distributed with parameter π, and N_2 is the number of probes it takes until the next OFF probe. The distribution of N_2 is dependent on N_1, and has the following distribution function.

$$N_2 = \begin{cases} 1 \text{ w.p. } p_{10}^{(N_1+2)T} \\ i \text{ w.p. } p_{11}^{(N_1+2)T} p_{10}^{2T} (p_{11}^{2T})^{i-1} \quad i \geq 2 \end{cases} \tag{12}$$

Therefore,

$$\bar{X}_{SB} = T\mathbb{E}(N_1 + N_2) = T\left(\frac{1}{\pi} + 1 + \frac{\mathbb{E}[p_{11}^{(2+N_1)T}]}{p_{10}^{2T}}\right) \tag{13}$$

While the transmitter searches for an ON channel, a reward of

$$\bar{R}_{SB}^1 = \mathbb{E}\left[\sum_{i=0}^{(N_1-1)T-1} p_{11}^{i+2T} + \sum_{i=0}^{T-1} p_{11}^i\right] \tag{14}$$

is earned. The first term is the throughput from transmitting over the best channel while looking for an ON channel, which starts with belief p_{11}^{2T} and decays until an ON channel is found. The second term represents the reward earned over the interval once the ON channel is found. While the transmitter waits for the second-best channel to be OFF, a reward of

$$\bar{R}_{SB}^2 = \mathbb{E}\left[(N_2 - 1)\sum_{i=0}^{T-1} p_{11}^i + \sum_{i=0}^{T-1} p_{11}^{T+i}\right] \tag{15}$$

is accumulated. The average reward per time slot is given by

$$\frac{\bar{R}_{SB}^1 + \bar{R}_{SB}^2}{\bar{X}_{SB}} = \pi + \frac{\pi p_{10}^T(\pi + p_{10}^{2T})}{(p+q)T[\pi^2 + p_{10}^{2T}(1 - (1-p-q)^T + \pi)]} \tag{16}$$

We can compute the difference between (16) and (5) from Lemma 1 as

$$\frac{\bar{R}_{SB}^1 + \bar{R}_{SB}^2}{\bar{X}_{SB}} - \frac{\bar{R}_B}{\bar{X}_B} = \frac{((1-p-q)^T \pi p_{10}^T)^2}{T(p+q)(\pi + p_{10}^T)(\pi^2 + p_{10}^{2T}(\pi + 1 - (1-p-q)^T))} \tag{17}$$

Since $p \leq \frac{1}{2}$ and $q \leq \frac{1}{2}$, we have $0 \leq (1-p-q)^T \leq 1$ for all T. Therefore, the expression in (17) is positive, completing the proof. ■

Theorem 1 states that probing the channel with the second highest belief is a better policy than probing the channel with the highest belief. A numerical comparison between these two policies is shown in Figure 2. This result is in sharp contrast to the result in [8] that shows that probing the channel with the highest belief is optimal. In our model, when a probed channel is OFF, the transmitter uses its knowledge of the system to transmit over another

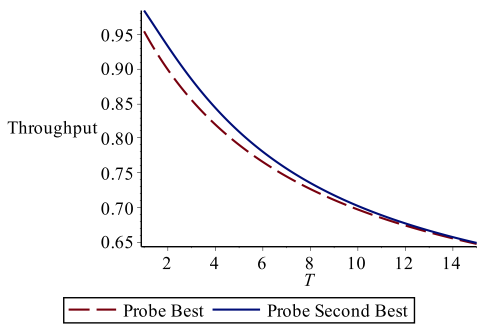

Fig. 2. Comparison of the probe best policy and the probe 2nd best policy for varying probing intervals T. In this example, $p = q = 0.05$.

channel believed to be ON. In the model of [8], when an OFF channel is probed, the transmitter can not schedule a packet in that slot. This difference in reward after probing leads to significantly different probing policies.

B. Round Robin Policy

It is of additional interest to consider a min-max policy, the *round robin policy*, which probes the channel for which the transmitter has the least knowledge. In a system with finitely many channels, the round robin policy probes all of the channels sequentially, but since there are infinitely many channels, the policy always probes a channel that has previously never been probed. Consider channel state beliefs $x_1, x_2, x_3, \ldots$ where $x_1 \geq x_2 \ldots \geq x_i \ldots \geq \pi$. The immediate reward of round robin is given by:

$$\pi + (1-\pi)x_1 = \pi + x_1 - \pi x_1. \tag{18}$$

By comparing (18) to (10), it is clear the immediate reward of the round robin policy is less than the probe best policy. Interestingly, the following Theorem shows that the average per-slot throughput is the same for the round robin policy as the myopic probe best policy.

Theorem 2. *For all fixed T, the round robin policy has a per-slot average throughput of*

$$\mathbb{E}[\mathit{Thpt}] = \pi + \frac{\pi p_{10}^T}{T(p+q)(p_{10}^T + \pi)}, \tag{19}$$

the same as the probe best policy.

Proof: Let a renewal occur every time a new channel is probed and found to be ON. Since the result of each probe is an i.i.d. random variable with parameter π, the inter-renewal intervals are i.i.d. The inter-renewal time $X_{RR} = T \cdot N$, where T is the time between probes, and N is a geometric random variable with parameter π, as defined in (3). Over that interval, the transmitter transmits over the last channel known to be ON, until a new ON channel is found. The expected reward earned over each renewal period is given by

$$\bar{R}_{RR} = \mathbb{E}\left[\sum_{i=0}^{N*T-1} p_{11}^i\right] = \mathbb{E}\left[\pi NT + \frac{p_{10}^{NT}}{p+q}\right] \tag{20}$$

$$= T + \frac{p_{10}^T}{p+q-q(1-p-q)^T}. \tag{21}$$

Policy	Theory	Simulation
Probe Best	0.7659	0.7657
Probe Second Best	0.7806	0.7806
Round Robin	0.7659	0.7662

TABLE II
THROUGHPUT COMPARISON FOR DIFFERENT PROBING POLICIES WITH $p = q = 0.05$, $T = 6$. SIMULATION ASSUMES 500 CHANNELS AND A TIME HORIZON OF 1,000,000 PROBES.

Thus, the time-average reward is given by

$$\frac{\bar{R}_{RR}}{\bar{X}_{RR}} = \pi + \frac{\pi p_{10}^T}{T(p+q)(\pi + p_{10}^T)}, \tag{22}$$

which is the same as the reward of the probe best policy in Lemma 1. ■

This result is surprising, since the round robin policy trades off immediate reward for increasing knowledge of the channel states, but yields the same average throughput as a myopic policy.

C. Optimal Probing Policy

In order to confirm the results in the previous section, we simulate a system of 500 channels assuming a time horizon of 1,000,000 slots, and apply different probing policies at a fixed probing interval of 6 slots. We compute the average throughput obtained over the total horizon. The results are shown in Table II. In this simulation, the probe second best policy is optimal over all policies considered, while the probe best policy and round robin policies have the same average throughput. These results lead to the following conjecture.

Conjecture 1. *The probe second best policy is optimal among all channel probing policies for fixed probing intervals T.*

Intuition on Potential Proof: We believe this statement can be proven using reverse induction over a finite horizon of N probing times. As a terminal reward, the probe second best policy and probe best policy are optimal, as they maximize the immediate reward. Conditioned on the optimality of probe second best for probes $n+1, \ldots, N$, we prove that probe second best has a higher expected reward than probe k^{th} best at probe n for all k. The difficulty in this proof lies in showing that the gain in expected future reward of probe k^{th} best over probe second best is less than the gain in immediate reward from probing the 2^{nd} best at the current time. Note that the simple coupling argument used to prove the optimality of the probe best policy in [8] cannot be used here, as the probe second best policy does not induce a cyclic ordering of probed channels like the probe best policy.

IV. DYNAMIC OPTIMIZATION OF PROBING EPOCHS

Consider the policy where at each time, the transmitter either probes the channel with the highest belief for a fixed cost c, or does not probe. The optimal decision as to whether to probe is a function of the state, and is described by the following Theorem.

Theorem 3. *Given that the transmitter only probes the channel with the highest belief, the optimal probing decision is to probe immediately after probing an OFF channel, and to probe k^* slots after probing an ON channel, where k^* is given by*

$$k^* = \arg\max_k \frac{1}{k\pi + p_{10}^k}\left(\frac{\pi p_{10}^k}{(p+q)} - c(\pi + p_{10}^k)\right) \quad (23)$$

Proof: Under the probe best policy, the belief of the best channel x_1 at every slot satisfies $x_1 \geq \pi$, and the belief of every other channel equals π. When a probed channel is OFF, the belief of every channel is π, and the system remains in that state until an ON channel is probed. Therefore, the optimal decision at this state is either to never probe, or to always probe[2]. When a probed channel is ON, the highest belief is always $1-q$ in the next slot, and so there exists an optimal time k^* to wait before probing again.

Assume a probe occurs in the slot immediatly after probing an OFF channel, and k slots after probing an ON channel. Define a renewal to occur when the transmitter probes an OFF channel. It follows that the expected inter-renewal time is given by

$$\bar{X}_B = 1 + \pi k \mathbb{E}[N], \quad (24)$$

where N is the number of ON probes before an OFF probe, and is geometrically distributed with parameter p_{10}^k. The reward accumulated over this interval is given by

$$\bar{R}_B = (\pi - c) + \pi\mathbb{E}[N]\left(\sum_{i=0}^{k-1} p_{11}^i - c\right). \quad (25)$$

Therefore, the average per-time slot reward is given by

$$\frac{\bar{R}_B}{\bar{X}_B} = \frac{p_{10}^k(\pi - c) + \pi\left(\sum_{i=0}^{k-1} p_{11}^i - c\right)}{p_{10}^k + k\pi} \quad (26)$$

$$= \pi - c\left(\frac{\pi + p_{10}^k}{k\pi + p_{10}^k}\right) + \frac{\pi p_{10}^k}{(p+q)(k\pi + p_{10}^k)} \quad (27)$$

The maximizing value of k in equation (27) is the optimal time k^* to wait after an ON probe. ∎

While Theorem 3 is specific to the probe best policy, a similar result holds for the round robin policy in Section III-B, in that it is optimal to probe in the next slot after an OFF channel is probed, and it is optimal to wait k' after an ON channel is probed to probe again. However, the k' here is different than k^* in Theorem 3. To illustrate, Figure 3 plots the average reward of round robin and probe best for different values of k. While under fixed probing intervals, Theorem 2 states that both policies have the same average reward, the probe best policy outperforms the round robin policy after dynamically optimizing over the probing intervals. For probe second best, the optimal time until the next probe depends on the belief of the best channel after an ON channel is probed, and consequently, probe second best does not have a single solution for the optimal probing interval after an ON channel has been probed.

[2]We assume the policy that never probes is suboptimal. One can characterize the value of c for which probing is necessary, but it is omitted here for brevity.

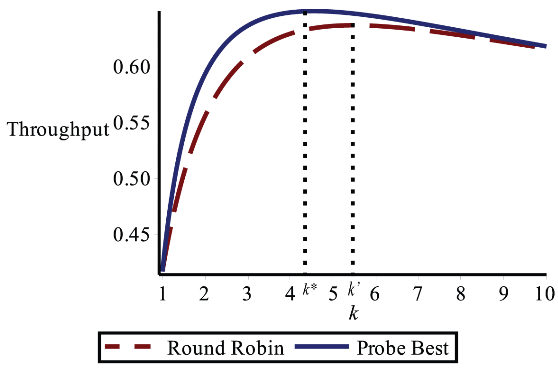

Fig. 3. Comparison of the probe best policy and round robin for varying values of k, the minimum interval between probes. In this example, $p = q = 0.1$, and $c = 0.5$.

V. Conclusion

In contrast to the work in [8], [9] that established the optimality of the myopic probe best policy, we showed that for a slightly modified model, probing the best channel is no longer optimal. Using renewal theory, we proved that a simple alternative, the probe second best policy, outperforms the probe best policy in terms of average throughput. As noted in Section III, we conjecture that probing the second best channel is the optimal decision at all probing instances. Proving this conjecture is interesting, and remains an open problem.

References

[1] K. Jagannathan, S. Mannor, I. Menache, and E. Modiano, "A state action frequency approach to throughput maximization over uncertain wireless channels," in *INFOCOM, 2011 Proceedings IEEE*. IEEE, 2011.

[2] S. Guha, K. Munagala, and S. Sarkar, "Jointly optimal transmission and probing strategies for multichannel wireless systems," in *Information Sciences and Systems, 2006 40th Annual Conference on*. IEEE, 2006.

[3] ——, "Optimizing transmission rate in wireless channels using adaptive probes," in *ACM SIGMETRICS Performance Evaluation Review*. ACM, 2006.

[4] P. Chaporkar and A. Proutiere, "Optimal joint probing and transmission strategy for maximizing throughput in wireless systems," *Selected Areas in Communications, IEEE Journal on*, 2008.

[5] N. Chang and M. Liu, "Optimal channel probing and transmission scheduling for opportunistic spectrum access," in *International Conference on Mobile Computing and Networking: Proceedings of the 13 th annual ACM international conference on Mobile computing and networking*, 2007.

[6] A. Gopalan, C. Caramanis, and S. Shakkottai, "On wireless scheduling with partial channel-state information," in *Proc. Ann. Allerton Conf. Communication, Control and Computing*, 2007.

[7] K. Kar, X. Luo, and S. Sarkar, "Throughput-optimal scheduling in multichannel access point networks under infrequent channel measurements," *Wireless Communications, IEEE Transactions on*, 2008.

[8] S. Ahmad, M. Liu, T. Javidi, Q. Zhao, and B. Krishnamachari, "Optimality of myopic sensing in multichannel opportunistic access," *Information Theory, IEEE Transactions on*, 2009.

[9] Q. Zhao, B. Krishnamachari, and K. Liu, "On myopic sensing for multi-channel opportunistic access: Structure, optimality, and performance," *Wireless Communications, IEEE Transactions on*, 2008.

[10] R. Gallager and R. Gallager, *Discrete stochastic processes*. Kluwer Academic Publishers, 1996.