Optimal, Suboptimal and Adaptive Threshold Policies for Power Efficiency of Wireless Networks*

P. T. Kabamba, S. M. Meerkov†, and C. Y. Tang
Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109-2122, USA

***Abstract*— In this paper, we prove that the optimal power-efficient transmission policy, which ensures a desired throughput in a wireless network, is necessarily of threshold nature. Although detailed properties of this policy might be quite complicated, we show that it can be approximated by a suboptimal one, which is both simple and practical. This suboptimal policy may lead to substantial improvement in power-efficiency (compared to the widely used constant SNR policy), but at the expense of location-fairness. To alleviate this deficiency, we introduce an adaptive threshold policy and show that it is both relatively power-efficient and location-fair.**

I. INTRODUCTION

Threshold policies in wireless networks, according to which transmissions are attempted only if channel conditions are sufficiently good, have been shown to be extremely power-efficient [1–4]. However, their deficiency, i.e., the lack of location-fairness, whereby a user in some location is more likely to transmit than in others, has not been exposed and compensated for. Along with proving optimality of threshold policies, this paper shows that these policies indeed lack location-fairness and designs an adaptive threshold policy where the threshold is adjusted so that location-fairness is guaranteed.

The outline of this paper is as follows: In Section II, the model of the network under consideration is described. Performance measures addressed are introduced in Section III. In Section IV, the optimal transmission policy is proved to be of threshold nature and shown to have limited practicality. Thus, in Section V we present a simple, practical threshold policy and show that it may result in up to $11\,\mathrm{dB}$ power-efficiency improvement without sacrificing throughput, or 90% throughput improvement without additional power consumption. In addition, Section V exposes the lack of location-fairness of this threshold policy, which may preclude its utilization in delay-sensitive applications. Therefore, in Section VI, an adaptive threshold policy is designed and its properties are investigated. We show that up to $3.6\,\mathrm{dB}$ power-efficiency or 30% throughput improvement can still be achieved, while maintaining location-fairness. Finally, the conclusions are formulated in Section VII. The proofs can be found in [5].

II. MODELING

The wireless network considered in this paper consists of a mobile user, a channel, and a base station, described below:

User: At each time slot $k \in \mathbb{Z}$, the user sends an information packet to the base station with transmit power $p(k) \geq 0$. If $p(k) = 0$, no transmission takes place.

*This research was supported by NSF under Grant No. ANI-0106716.

†Please address correspondence to Professor S. M. Meerkov, Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109-2122, USA, e-mail: smm@eecs.umich.edu, phone: (734) 763-6349, fax: (734) 763-8041.

Channel: The channel affects the transmissions so that the received SNR at time slot k, $r(k)$, is given by

$$r(k) = e^{x(k)}p(k), \quad k \in \mathbb{Z}, \tag{1}$$

where $e^{x(k)}$ is the *channel gain*, which combines path loss, shadowing, thermal noise power, and other radio-wave propagation effects, and $x(k) \in \mathbb{R}$ is the *log-channel gain*.

The sequence of log-channel gains, $\{x(k), k \in \mathbb{Z}\}$, is assumed to be a random process. Several additional assumptions will also be imposed. The first one is the most general and will be used in the derivation of the optimal transmission policy:

Assumption A1. Process $\{x(k), k \in \mathbb{Z}\}$ is such that each $x(k)$ is a continuous random variable with probability density function $f_{x(k)}$ satisfying $f_{x(k)}(v) > 0\ \forall v \in \mathbb{R}$. ■

The second assumption will be used to analyze the performance of transmission policies investigated in this work:

Assumption A2. Process $\{x(k), k \in \mathbb{Z}\}$ is a WSS Gaussian random process with mean $\mathrm{E}\{x(k)\} = \mu_x$ and autocovariance function $\mathrm{E}\{(x(k+\ell) - \mu_x)(x(k) - \mu_x)\} = \sigma_x^2\rho_x(\ell)$, $\ell \in \mathbb{Z}$, where $\mu_x \in \mathbb{R}$, $\sigma_x > 0$, $\rho_x(0) = 1$, $\rho_x(\ell) = \rho_x(-\ell)$, $|\rho_x(\ell)| < 1\ \forall \ell \neq 0$, and $\lim_{\ell\to\infty} \rho_x(\ell) = 0$, i.e., $\mathrm{E}\{(x(k) - \mu_x)^2\} = \sigma_x^2$ is the variance of $x(k)$ and $\rho_x(\ell)$ is the correlation coefficient of $x(k+\ell)$ and $x(k)$. ■

For performance comparison, we assume below a particular expression for ρ_x:

Assumption A3. Process $\{x(k), k \in \mathbb{Z}\}$ is as in Assumption A2, with $\rho_x(\ell) = \zeta_0\zeta_1^{|\ell|} + (1-\zeta_0)\zeta_2^{\ell^2}$, $\ell \in \mathbb{Z}$, where $0 < \zeta_0, \zeta_1, \zeta_2 < 1$. ■

Base Station: If a packet is sent at time slot k, i.e., if $p(k) > 0$, the base station attempts to decode the packet. Otherwise, i.e., if $p(k) = 0$, no attempt is made. The (normalized) throughput at time slot k, $t(k)$, is assumed to be a function of the SNR $r(k)$,

$$t(k) = \Phi(r(k)), \quad k \in \mathbb{Z}, \tag{2}$$

where $\Phi : [0,\infty) \to [0,1)$ depends on the modulation, demodulation, and coding schemes employed, as well as the channel.

Several assumptions on Φ will be introduced. The first one will be used to derive the optimal transmission policy:

Assumption B1. Function $\Phi : [0,\infty) \to [0,1)$ is strictly increasing, satisfies $\Phi(0) = 0$, and has a continuous, bounded derivative Φ'. ■

The second one, used in performance analysis, does not require Φ to be differentiable:

Assumption B2. Function $\Phi : [0,\infty) \to [0,1)$ is strictly increasing and satisfies $\Phi(0) = 0$. ■

The third one, used in performance comparison, assumes that Φ corresponds to a network operating in a Rayleigh fading channel using BFSK modulation, noncoherent demodulation, and Reed-Solomon codes (see [5,6] for more details):

Assumption B3. Function $\Phi : [0,\infty) \to [0,1)$ is defined by

$$\Phi(r) = \begin{cases} \max_{q\in\{2,4,\dots,32\}} \frac{q}{32}\varphi(q,r), & \text{if } r > 0,\\ 0, & \text{if } r = 0,\end{cases}$$

where

$$\varphi(q,r) = \sum_{j=0}^{\lfloor\frac{32-q}{2}\rfloor} \binom{32}{j}\left(1-\left(1-\tfrac{1}{2+r}\right)^5\right)^j\left(1-\tfrac{1}{2+r}\right)^{5(32-j)}. \qquad \blacksquare$$

III. PERFORMANCE MEASURES

Typically, performance measures considered in wireless networks are the average throughput and, perhaps, the average transmit power, defined on the *infinite* time interval. Unfortunately, these averages may be deficient in delay-sensitive applications. The reason is that, even if, for example, the average throughput is high, it does not imply that a reliable communication has taken place at every relatively short time interval. To account for this deficiency, in this work we consider averages defined on *finite* time intervals: the finite-time average transmit power,

$$\bar{p}(k_1,k_2) = \tfrac{1}{k_2-k_1+1}\sum_{k=k_1}^{k_2} p(k), \quad k_1,k_2\in\mathbb{Z},\ k_1\le k_2, \quad (3)$$

and the finite-time average throughput,

$$\bar{t}(k_1,k_2) = \tfrac{1}{k_2-k_1+1}\sum_{k=k_1}^{k_2} t(k), \quad k_1,k_2\in\mathbb{Z},\ k_1\le k_2. \quad (4)$$

The finite-time averages (3) and (4) are random variables. In this work, a number of their statistical properties are examined and treated as performance measures:

Performance Measure P1. Mean of $\bar{p}(k_1,k_2)$, $\mathrm{E}\{\bar{p}(k_1,k_2)\}$.

Performance Measure P2. Mean of $\bar{t}(k_1,k_2)$, $\mathrm{E}\{\bar{t}(k_1,k_2)\}$.

Under the assumption of ergodicity, P1 and P2 coincide with the infinite-time averages. Measure P2 reflects only the "average" behavior of $\bar{t}(k_1,k_2)$. It does not tell how $\bar{t}(k_1,k_2)$ would depend on the user's location relative to the base station. This shortcoming is alleviated by the following measure:

Performance Measure P3. Conditional mean of $\bar{t}(k_1,k_2)$ given $x(k_1)=x_o\in\mathbb{R}$, i.e., $\mathrm{E}\{\bar{t}(k_1,k_2)|x(k_1)=x_o\}$.

Since a large (small) $x(k_1)$ typically corresponds to the user being in a good (bad) location at time slot k_1, P3 expresses the dependency of $\bar{t}(k_1,k_2)$ on location and, thus, characterizes *location-fairness* of the network.

Another measure of interest is the number of consecutive time slots without a transmission, referred to as the *downtime*. To formalize, let $\mathbb{Z}_+$ denote the set of positive integers and let

$$d(k) = \begin{cases} \min\{\ell\in\mathbb{Z}_+ : p(k+\ell)>0\}, & \text{if } p(k-1)>0, p(k)=0,\\ 0, & \text{otherwise.}\end{cases}$$

Then, whenever $d(k)>0$, a period without a transmission begins at time slot k and lasts for $d(k)$ time slots, i.e., the downtime is $d(k)$. Here, we are interested in:

Performance Measure P4. Mean downtime, $\mathrm{E}\{d(k)|d(k)>0\}$.

IV. THRESHOLD NATURE OF OPTIMAL TRANSMISSION POLICY

In this Section, we prove that, to communicate information in the most power-efficient manner, the user must remain silent whenever the channel condition is worse than some threshold, and must transmit otherwise.

Definition 1. A *transmission policy* g is a function $g:\mathbb{R}\to[0,\infty)$ that, at each time slot $k\in\mathbb{Z}$, maps the log-channel gain $x(k)$ to the transmit power $p(k)$, i.e., $p(k)=g(x(k))$. $\blacksquare$

Thus, a transmission policy at each time slot k decides whether the user would send a packet, i.e., $p(k)>0$, or remain silent, i.e., $p(k)=0$, and, in the former case, with what power to transmit.

Definition 2. A transmission policy g is a *threshold policy* if there exists $\tau\in\mathbb{R}$ such that $g(x(k))>0$ if $x(k)>\tau$ and $g(x(k))=0$ if $x(k)\le\tau$. $\blacksquare$

Hence, a threshold policy at each time slot k instructs the user to send a packet if $x(k)>\tau$, i.e., the channel condition is better than some threshold τ, and remain silent otherwise.

Problem 1. Consider a network described by (1) and (2), with $x(k)$ specified by Assumption A1 and Φ by Assumption B1. Given $k_1,k_2\in\mathbb{Z}$, $k_1\le k_2$, and $0<c<\lim_{r\to\infty}\Phi(r)$, find a transmission policy g that minimizes the mean of the average transmit power from time slot k_1 to k_2, i.e., $\mathrm{E}\{\bar{p}(k_1,k_2)\}$, subject to the mean of the average throughput over the same time slots, i.e., $\mathrm{E}\{\bar{t}(k_1,k_2)\}$, being equal to c. $\blacksquare$

A network operating under the optimal transmission policy, i.e., the solution to Problem 1, may be regarded as having *the most power-efficient operation* since $\mathrm{E}\{\bar{p}(k_1,k_2)\}$ is minimized while $\mathrm{E}\{\bar{t}(k_1,k_2)\}=c$ is achieved.

Theorem 1. *If g^* is a solution to Problem 1, then g^* is a threshold policy.*

Theorem 1 guarantees neither the existence of a solution to Problem 1 nor its uniqueness. Its value comes from its contrapositive and generality: If a transmission policy is not a threshold policy, it is not the most power-efficient, irrespective of the type of modulation, demodulation, coding, and channel, as long as they satisfy mild assumptions (Assumptions A1 and B1).

A result stronger than Theorem 1—the analytical solution to Problem 1—but valid only for a class of Φ is stated next:

Theorem 2. *If, in addition to Assumption B1, Φ is strictly concave, then Problem 1 has a unique solution, g^*, given by*

$$g^*(x(k)) = \begin{cases} e^{-x(k)}(\Phi')^{-1}\big(\Phi'(0)e^{-(x(k)-\tau^*)}\big), & \text{if } x(k)>\tau^*,\\ 0, & \text{otherwise,}\end{cases}$$

where $(\Phi')^{-1}$ is the inverse of Φ' and $\tau^*\in\mathbb{R}$ is the unique solution to

$$\int_{\tau^*}^{\infty}\Phi\big((\Phi')^{-1}\big(\Phi'(0)e^{-(v-\tau^*)}\big)\big)\big(\tfrac{1}{k_2-k_1+1}\sum_{k=k_1}^{k_2} f_{x(k)}(v)\big)\,dv = c.$$

Under the optimal transmission policy g^*, the throughput is

$$t(k) = \begin{cases} \Phi\big((\Phi')^{-1}\big(\Phi'(0)e^{-(x(k)-\tau^*)}\big)\big), & \text{if } x(k)>\tau^*,\\ 0, & \text{otherwise.}\end{cases}$$

Moreover, $t(k)$ is strictly increasing with respect to $x(k)$ for $x(k)>\tau^*$, and $\lim_{x(k)\to\infty} t(k) = \lim_{r\to\infty}\Phi(r)$.

In Theorem 2, the threshold τ^* can be evaluated numerically via the bisection method (see [5] for more details).

Utilization of the g^* of Theorem 2 requires the knowledge of Φ and $f_{x(k)}$, $k=k_1,\dots,k_2$, as well as the strict concavity of Φ. These requirements are seldom met in practice, for

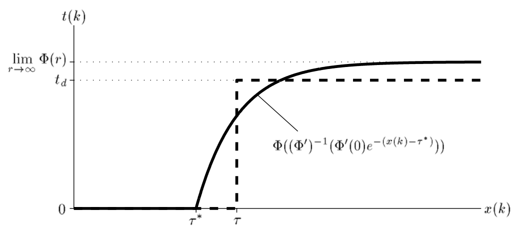

Fig. 1. Optimal throughput vs. log-channel gain characteristics (solid curve) and its approximation (dashed curve)

a variety of reasons. First, the channel, and hence Φ, is often uncertain. Second, even if the channel is known, analytical expression for Φ may still be unavailable for certain codes, such as convolutional or turbo codes, for which performance usually cannot be evaluated analytically. Consequently, Φ' may only be computed numerically with limited accuracy. Third, the $f_{x(k)}$'s depend on the user's behavior and, thus, are prone to modeling errors. Finally, Φ may not be concave, as was the case in Assumption B3. Due to these reasons, the g^* of Theorem 2 is of limited applicability. Nevertheless, this theorem leads to a suboptimal but practical transmission policy derived next.

V. SIMPLIFIED AND SUBOPTIMAL THRESHOLD POLICIES

A. Policy Formulation

According to Theorem 2, $t(k)$ under g^* is related to $x(k)$ qualitatively as shown in Figure 1 by the solid curve. The exact nature of this curve depends on the term $\Phi((\Phi')^{-1}(\Phi'(0)e^{-(x(k)-\tau^*)}))$, which is fairly complicated and may be sensitive to modeling errors, as mentioned above. To avoid these problems, we consider a simple approximation of this curve by a step function, as shown in Figure 1 by the dashed curve, i.e.,

$$t(k) = \begin{cases} t_d, & \text{if } x(k) > \tau, \\ 0, & \text{otherwise,} \end{cases} \quad k \in \mathbb{Z}, \tag{5}$$

where $0 < t_d < \lim_{r\to\infty}\Phi(r)$ is the desired throughput and $\tau \in \mathbb{R}$ is the threshold. If $t(k)$ is defined by (5), it follows from (1) and (2) that

$$p(k) = \begin{cases} r_d e^{-x(k)}, & \text{if } x(k) > \tau, \\ 0, & \text{otherwise,} \end{cases} \quad k \in \mathbb{Z}, \tag{6}$$

where $r_d > 0$ is the desired SNR defined by $r_d = \Phi^{-1}(t_d)$.

Policy (6), which instructs the user to regulate $r(k)$ at r_d if $x(k) > \tau$ and remain silent otherwise, is referred to as the *simplified threshold policy*. It is rather practical since it does not depend on Φ and the $f_{x(k)}$'s and contains only two free parameters, r_d and τ. When r_d and τ are chosen to minimize $\mathrm{E}\{\bar{p}(k_1,k_2)\}$ subject to some desired $\mathrm{E}\{\bar{t}(k_1,k_2)\}$, we refer to (6) as the *suboptimal threshold policy*.

Policy (6) is not new; it has been studied in recent literature [1–4]. The novelty here is in the analysis of its performance (Section V-B), which reveals not only its strength (Section V-C), but also its weaknesses (Sections V-D and V-E) that motivate the design of an adaptive threshold policy (Section VI).

B. Performance Analysis

Theorem 3. *Consider a network described by* (1) *and* (2)*, with $x(k)$ specified by Assumption A2 and Φ by Assumption B2. Suppose it operates under the simplified threshold policy* (6)*. Then, for any $k_1, k_2 \in \mathbb{Z}$, $k_1 \le k_2$, and any $k \in \mathbb{Z}$, Performance Measures P1–P4 are given by*

$$\mathrm{E}\{\bar{p}(k_1,k_2)\} = r_d e^{\frac{\sigma_x^2}{2}-\mu_x} Q_1(\tau_x+\sigma_x) \triangleq \mathrm{E}\{\bar{p}\},$$
$$\mathrm{E}\{\bar{t}(k_1,k_2)\} = \Phi(r_d) Q_1(\tau_x) \triangleq \mathrm{E}\{\bar{t}\},$$
$$\mathrm{E}\{\bar{t}(k_1,k_2)|x(k_1)=x_o\} = \frac{\Phi(r_d)}{K}\Big[\frac{1+\operatorname{sgn}(\frac{x_o-\mu_x}{\sigma_x}-\tau_x)}{2} + \sum_{\ell=1}^{K-1} Q_1\Big(\frac{\tau_x-\rho_x(\ell)\frac{x_o-\mu_x}{\sigma_x}}{\sqrt{1-\rho_x^2(\ell)}}\Big)\Big] \triangleq \mathrm{E}\{\bar{t}_K|x_o\},$$
$$\mathrm{E}\{d(k)|d(k)>0\} = \frac{1-Q_1(\tau_x)}{Q_1(\tau_x)-Q_2(\tau_x;\rho_x(1))} \triangleq \mathrm{E}\{d|d>0\},$$

where $\tau_x = \frac{\tau-\mu_x}{\sigma_x}$, $K = k_2-k_1+1$, $Q_1(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{v^2}{2}}\,dv$, and $Q_2(x;\rho) = \frac{1}{2\pi\sqrt{1-\rho^2}}\int_x^\infty\int_x^\infty e^{-\frac{v^2+w^2-2\rho vw}{2(1-\rho^2)}}\,dw\,dv$.

In Theorem 3, τ_x is the *normalized threshold* since the user would transmit if and only if $\frac{x(k)-\mu_x}{\sigma_x} > \tau_x$, and $\frac{x(k)-\mu_x}{\sigma_x}$ is a standard Gaussian random variable. Parameter K is the number of time slots between k_1 and k_2. We write $\mathrm{E}\{\bar{p}(k_1,k_2)\}$ and $\mathrm{E}\{\bar{t}(k_1,k_2)\}$ as $\mathrm{E}\{\bar{p}\}$ and $\mathrm{E}\{\bar{t}\}$ to stress their independence with respect to k_1 and k_2. We also write $\mathrm{E}\{\bar{t}(k_1,k_2)|x(k_1)=x_o\}$ as $\mathrm{E}\{\bar{t}_K|x_o\}$ to stress its dependence on K, and $\mathrm{E}\{d(k)|d(k)>0\}$ as $\mathrm{E}\{d|d>0\}$ to stress its independence with respect to k. Note that Q_1 is the standard Q-function and Q_2 is its two-dimensional counterpart.

The performance of the simplified threshold policy (6) will be compared to that of the widely adopted *constant SNR policy* [7, 8], defined as

$$p(k) = r_d e^{-x(k)}, \quad k \in \mathbb{Z}, \tag{7}$$

where $r_d > 0$ is the desired SNR. This policy, as its name suggests, instructs the user to always transmit and maintain a constant SNR, $r(k) = r_d$, despite the channel conditions.

Theorem 4. *Consider a network described by* (1) *and* (2)*, with $x(k)$ specified by Assumption A2 and Φ by Assumption B2. Suppose it operates under the constant SNR policy* (7)*. Then, for any $k_1, k_2 \in \mathbb{Z}$, $k_1 \le k_2$, Performance Measures P1–P3 are given by*

$$\mathrm{E}\{\bar{p}(k_1,k_2)\} = r_d e^{\frac{\sigma_x^2}{2}-\mu_x} \triangleq \mathrm{E}\{\bar{p}\},$$
$$\mathrm{E}\{\bar{t}(k_1,k_2)\} = \Phi(r_d) \triangleq \mathrm{E}\{\bar{t}\},$$
$$\mathrm{E}\{\bar{t}(k_1,k_2)|x(k_1)=x_o\} = \Phi(r_d).$$

Similar to Theorem 3, $\mathrm{E}\{\bar{p}(k_1,k_2)\}$ and $\mathrm{E}\{\bar{t}(k_1,k_2)\}$ in Theorem 4 are written as $\mathrm{E}\{\bar{p}\}$ and $\mathrm{E}\{\bar{t}\}$ to emphasize their independence with respect to k_1 and k_2. Yet, unlike Theorem 3, $\mathrm{E}\{\bar{t}(k_1,k_2)|x(k_1)=x_o\}$ here is independent of k_1 and k_2. Furthermore, $\mathrm{E}\{d(k)|d(k)>0\}$ is not computed since the constant SNR policy (7) has zero downtime.

Theorems 3 and 4 are used next to compare the performance of policies (6) and (7). For comparison purpose, we adopt Assumptions A3 and B3 and let $\mu_x = 2$, $\sigma_x = 2$, $\zeta_0 = 0.6$, $\zeta_1 = 0.99999$, and $\zeta_2 = 0.98$.

C. Power-Efficiency and Throughput Comparisons

Figure 2 illustrates the performance of policies (6) and (7) in terms of $\mathrm{E}\{\bar{p}\}$ and $\mathrm{E}\{\bar{t}\}$, i.e., Performance Measures P1 and P2. Each point in the gray region is feasible in the sense that it corresponds to a specific r_d and τ_x of the simplified threshold policy, referred to as a *realization* of the policy for that point. For

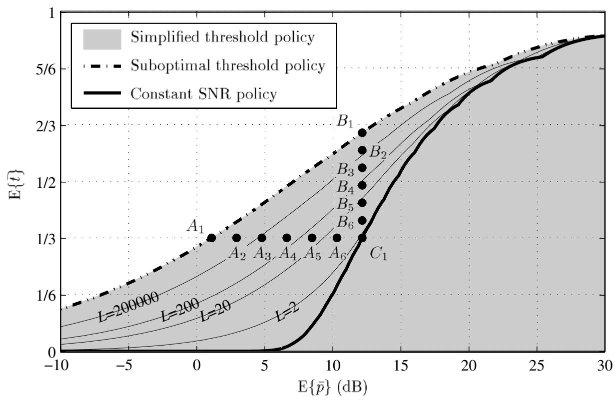

Fig. 2. Throughput vs. power characteristics of the threshold and constant SNR policies

	r_d (dB)	τ_x
A_1	16.75	-0.076
A_2	13.92	-0.593
A_3	13.05	-0.959
A_4	12.60	-1.329
A_5	12.31	-1.770
A_6	12.18	-2.380

	r_d (dB)	τ_x
B_1	20.09	-1.006
B_2	16.63	-1.629
B_3	15.36	-1.941
B_4	14.39	-2.240
B_5	13.60	-2.565
B_6	12.86	-3.023

TABLE I
REALIZATIONS OF THE SIMPLIFIED THRESHOLD POLICY

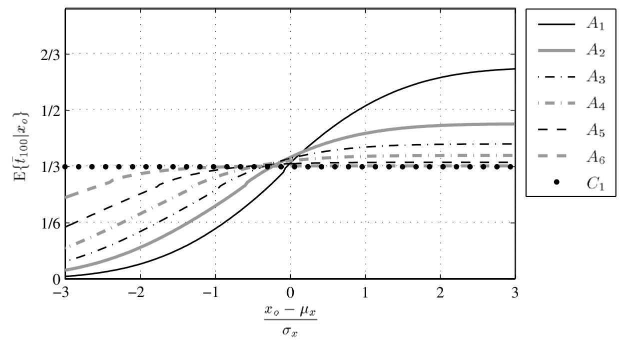

Fig. 3. Location-fairness of realizations for points A_1–A_6 and C_1

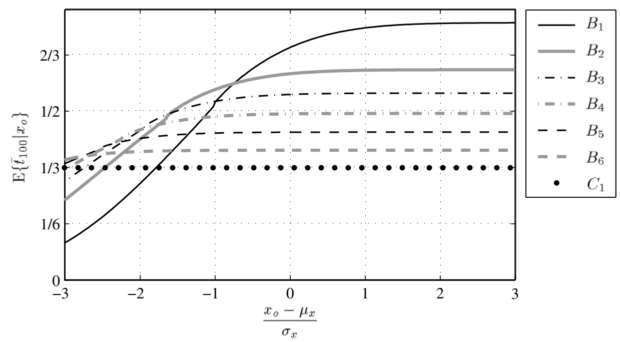

Fig. 4. Location-fairness of realizations for points B_1–B_6 and C_1

example, points A_1–A_6 and B_1–B_6 correspond to realizations listed in Table I. Each point on the dashdot curve corresponds to a realization of the suboptimal threshold policy, where r_d and τ_x are selected so that $\mathrm{E}\{\bar{p}\}$ is minimized subject to some desired $\mathrm{E}\{\bar{t}\}$; e.g., A_1 and B_1. Each point on the solid curve corresponds to a realization of the constant SNR policy; e.g., C_1 corresponds to $r_d = 12.14\,\mathrm{dB}$. (The meaning of the thin solid curves in Figure 2 will be explained in Section VI-C.)

Analyzing Figure 2 we observe that:

(a) To get $\mathrm{E}\{\bar{t}\} = \frac{1}{3}$, the suboptimal threshold policy needs $\mathrm{E}\{\bar{p}\} = 1.09\,\mathrm{dB}$ (point A_1), whereas the constant SNR policy needs $\mathrm{E}\{\bar{p}\} = 12.14\,\mathrm{dB}$ (point C_1). Thus, the former is $11.05\,\mathrm{dB}$ more power-efficient than the latter.

(b) With $\mathrm{E}\{\bar{p}\} = 12.14\,\mathrm{dB}$, the suboptimal threshold policy gives $\mathrm{E}\{\bar{t}\} = 0.64$ (point B_1), whereas the constant SNR policy gives $\mathrm{E}\{\bar{t}\} = \frac{1}{3}$ (point C_1). Hence, the former yields 92% throughput improvement over the latter.

(c) Depending on the choice of r_d and τ_x, the simplified threshold policy may slightly (e.g., points A_6 and B_6) or significantly (e.g., A_2 and B_2) outperform the constant SNR policy (e.g., C_1). It may also underperform, if r_d and τ_x are not chosen properly (e.g., any point below and to the right of C_1).

(d) Observations (a)–(c) are independent of ζ_0, ζ_1, ζ_2, and μ_x because ζ_0, ζ_1, and ζ_2 do not affect $\mathrm{E}\{\bar{p}\}$ and $\mathrm{E}\{\bar{t}\}$, and a change in μ_x shifts the dashdot and solid curves and the gray region in Figure 2 horizontally at the same rate.

From these observations, we conclude that the simplified and suboptimal threshold policies provide remarkable improvements in power-efficiency and throughput over the constant SNR policy. These improvements take place because the user is not forced to transmit when the channel conditions are bad, thereby saving a substantial amount of transmit power that can be allocated for more aggressive transmissions when the channel conditions are good. The simplified and suboptimal threshold policies, however, are inferior to the constant SNR policy in terms of location-fairness and downtime, as shown below.

D. Lack of Location-Fairness

Recall that location-fairness is characterized by $\mathrm{E}\{\bar{t}(k_1,k_2)|x(k_1) = x_o\}$ or $\mathrm{E}\{\bar{t}_K|x_o\}$, i.e., Performance Measure P3, and that a large (small) x_o corresponds to a good (bad) location. Clearly, if $\mathrm{E}\{\bar{t}_K|x_o\}$ is roughly the same for all $x_o \in \mathbb{R}$, location-fairness is ensured. Figures 3 and 4 represent $\mathrm{E}\{\bar{t}_{100}|x_o\}$ as a function of $\frac{x_o-\mu_x}{\sigma_x}$ for realizations of the simplified threshold policy for points A_1–A_6 and B_1–B_6 (suboptimal threshold policy for points A_1 and B_1) and realization of the constant SNR policy for point C_1. From these figures we observe that:

(a) The suboptimal threshold policy has the worst location-fairness, while the constant SNR policy has the best. The simplified threshold policy lies somewhere in between, depending on the realizations.

(b) Realizations for points $A_1, A_2, \ldots, A_6$ yield ever improving location-fairness (see Figure 3), ever increasing $\mathrm{E}\{\bar{p}\}$ (see Figure 2), and the same $\mathrm{E}\{\bar{t}\} = \frac{1}{3}$. Hence, for a fixed $\mathrm{E}\{\bar{t}\}$, the more "uniform" $\mathrm{E}\{\bar{t}_K|x_o\}$ is, the larger $\mathrm{E}\{\bar{p}\}$ would be, i.e., there is a trade-off between location-fairness and power-efficiency improvement.

(c) Realizations for points $B_1, B_2, \ldots, B_6$ yield ever improving location-fairness (see Figure 4), ever decreasing

	$\mathrm{E}\{d \mid d>0\}$
A_1	23.38
A_2	16.37
A_3	13.27
A_4	11.04
A_5	9.14
A_6	7.32

	$\mathrm{E}\{d \mid d>0\}$
B_1	12.94
B_2	9.68
B_3	8.55
B_4	7.68
B_5	6.90
B_6	6.03

TABLE II

MEAN DOWNTIME OF THE SIMPLIFIED THRESHOLD POLICY

$\mathrm{E}\{\bar{t}\}$ (see Figure 2), and the same $\mathrm{E}\{\bar{p}\} = 12.14\,\mathrm{dB}$. Thus, for a fixed $\mathrm{E}\{\bar{p}\}$, the more "uniform" $\mathrm{E}\{\bar{t}_K|x_o\}$ is, the smaller $\mathrm{E}\{\bar{t}\}$ would be, i.e., there is a trade-off between location-fairness and throughput improvement.

E. Long Downtime

Recall that mean downtime is represented by $\mathrm{E}\{d(k)|d(k) > 0\}$ or $\mathrm{E}\{d|d > 0\}$, i.e., Performance Measure P4. Clearly, small $\mathrm{E}\{d|d > 0\}$ is desirable, particularly in delay-sensitive applications. Table II lists $\mathrm{E}\{d|d > 0\}$ for realizations of the simplified threshold policy for points A_1–A_6 and B_1–B_6 (suboptimal threshold policy for points A_1 and B_1). From Table II we observe that:

(a) The suboptimal threshold policy has the longest downtime, while the constant SNR policy has zero downtime. The simplified threshold policy again lies somewhere in between, depending on the realizations.

(b) Realizations for points $A_1, A_2, \ldots, A_6$ result in ever decreasing mean downtime. Along with Figure 2, they imply that, for a fixed $\mathrm{E}\{\bar{t}\}$, decreasing $\mathrm{E}\{\bar{p}\}$ increases $\mathrm{E}\{d|d > 0\}$, i.e., trade-off exists between power-efficiency improvement and downtime.

(c) Realizations for points $B_1, B_2, \ldots, B_6$ exhibit the same trends as $A_1, A_2, \ldots, A_6$. Along with Figure 2, they imply that, for a fixed $\mathrm{E}\{\bar{p}\}$, increasing $\mathrm{E}\{\bar{t}\}$ increases $\mathrm{E}\{d|d > 0\}$, i.e., trade-off exists between throughput improvement and downtime.

As it follows from the above observations, a network operating under the suboptimal threshold policy is very power-efficient but suffers from lack of location-fairness and long downtime, relative to the constant SNR policy. The same can be said, although to a lesser extent, about the simplified threshold policy. Therefore, the simplified and suboptimal threshold policies may be suitable only for delay-insensitive applications and are inadequate otherwise. This necessitates the development of an adaptive threshold policy discussed next.

VI. ADAPTIVE THRESHOLD POLICY

A. Policy Formulation

The lack of location-fairness of the simplified threshold policy (6) is due to the fact that the threshold τ is independent of the user's location relative to the base station. Obviously, while the user is in a good location, the log-channel gain $x(k)$ is on the average large and the condition $x(k) > \tau$ is more likely to be satisfied than when the user is in a bad location. Thus, to compensate for this "built-in" unfairness, the threshold level should be adjusted to location, for instance, according to

$$x(k) > \tau + \bar{x}(k-L, k-1), \tag{8}$$

where $\bar{x}(k-L, k-1)$ is the moving average of the log-channel gain over the past $L \in \mathbb{Z}_+$ time slots, i.e.,

$$\bar{x}(k-L, k-1) = \frac{1}{L}\sum_{\ell=1}^{L} x(k-\ell). \tag{9}$$

When $L \to \infty$, the averaging in (9) eliminates the location-dependence of $\bar{x}(k-L, k-1)$; when $L = 1$, there is no averaging. Therefore, there must be an L^* such that fading dips are averaged out but the location-dependence of $\bar{x}(k-L, k-1)$ is preserved. With this L^*, as it follows from (8) and (9), the new threshold, $\tau + \bar{x}(k-L^*, k-1)$, is large when the user is in a good location and small otherwise. This implies that the user has approximately equal probability to transmit, regardless of its location *vis-à-vis* the base station, and (8) would prevent transmission mostly under conditions of occasional fading dips.

Introducing the quantity

$$\tilde{x}(k) = x(k) - \bar{x}(k-L, k-1), \tag{10}$$

(8) can be written as $\tilde{x}(k) > \tau$. Using this expression, we define the *adaptive threshold policy* as

$$p(k) = \begin{cases} r_d e^{-x(k)}, & \text{if } \tilde{x}(k) > \tau, \\ 0, & \text{otherwise}, \end{cases} \quad k \in \mathbb{Z}, \tag{11}$$

where $r_d > 0$ is the desired SNR and $\tau \in \mathbb{R}$ is the threshold.

B. Performance Analysis

Theorem 5. *Consider a network described by* (1) *and* (2)*, with $x(k)$ specified by Assumption A2 and Φ by Assumption B2. Suppose it operates under the adaptive threshold policy* (9)–(11)*. Then, for any $k_1, k_2 \in \mathbb{Z}$, $k_1 \le k_2$, and any $k \in \mathbb{Z}$, Performance Measures P1–P4 are given by*

$$\mathrm{E}\{\bar{p}(k_1,k_2)\} = r_d e^{\frac{\sigma_x^2}{2}-\mu_x} Q_1(\tau_{\tilde{x}} + \sigma_x \rho_{x\tilde{x}}(0)) \triangleq \mathrm{E}\{\bar{p}\},$$

$$\mathrm{E}\{\bar{t}(k_1,k_2)\} = \Phi(r_d) Q_1(\tau_{\tilde{x}}) \triangleq \mathrm{E}\{\bar{t}\},$$

$$\mathrm{E}\{\bar{t}(k_1,k_2)|x(k_1)=x_o\} = \frac{\Phi(r_d)}{K}\sum_{\ell=0}^{K-1} Q_1\Bigg(\frac{\tau_{\tilde{x}} - \rho_{x\tilde{x}}(-\ell)\frac{x_o-\mu_x}{\sigma_x}}{\sqrt{1-\rho_{x\tilde{x}}^2(-\ell)}}\Bigg) \triangleq \mathrm{E}\{\bar{t}_K|x_o\},$$

$$\mathrm{E}\{d(k)|d(k)>0\} = \frac{1-Q_1(\tau_{\tilde{x}})}{Q_1(\tau_{\tilde{x}}) - Q_2(\tau_{\tilde{x}}; \rho_{\tilde{x}}(1))} \triangleq \mathrm{E}\{d|d>0\},$$

where $\tau_{\tilde{x}} = \frac{\tau}{\sigma_{\tilde{x}}}$, $K = k_2 - k_1 + 1$,

$$\sigma_{\tilde{x}}^2 = \sigma_x^2\Big[1 + \frac{1}{L} - \frac{2}{L^2}\sum_{\ell_1=1}^{L} \ell_1 \rho_x(\ell_1)\Big],$$

$$\rho_{\tilde{x}}(\ell) = \frac{(1+\frac{1}{L})\rho_x(\ell) - \frac{1}{L^2}\sum_{\ell_1=-L}^{L} |\ell_1| \rho_x(\ell+\ell_1)}{1 + \frac{1}{L} - \frac{2}{L^2}\sum_{\ell_1=1}^{L} \ell_1 \rho_x(\ell_1)},$$

$$\rho_{x\tilde{x}}(\ell) = \frac{\rho_x(\ell) - \frac{1}{L}\sum_{\ell_1=1}^{L} \rho_x(\ell+\ell_1)}{\sqrt{1 + \frac{1}{L} - \frac{2}{L^2}\sum_{\ell_1=1}^{L} \ell_1 \rho_x(\ell_1)}}.$$

In Theorem 5, $\tau_{\tilde{x}}$ is the *normalized threshold* since the user would transmit if and only if $\frac{\tilde{x}(k)}{\sigma_{\tilde{x}}} > \tau_{\tilde{x}}$, and $\frac{\tilde{x}(k)}{\sigma_{\tilde{x}}}$ is a standard Gaussian random variable. As in Theorem 3, K is the number of time slots between k_1 and k_2, and the notations $\mathrm{E}\{\bar{p}\}$, $\mathrm{E}\{\bar{t}\}$, $\mathrm{E}\{\bar{t}_K|x_o\}$, and $\mathrm{E}\{d|d > 0\}$ are utilized to emphasize their independence with respect to k_1, k_2, k, and dependence on K.

Theorem 5 is used next to compare the performance of the adaptive threshold policy (9)–(11) with the simplified threshold policy (6) and the constant SNR policy (7).

	r_d (dB)	$\tau_{\tilde{x}}$	L
A_5	13.30	-0.821	1860
A_6	12.42	-1.550	820
B_5	13.86	-1.711	900
B_6	12.91	-2.264	920

TABLE III
REALIZATIONS OF THE ADAPTIVE THRESHOLD POLICY

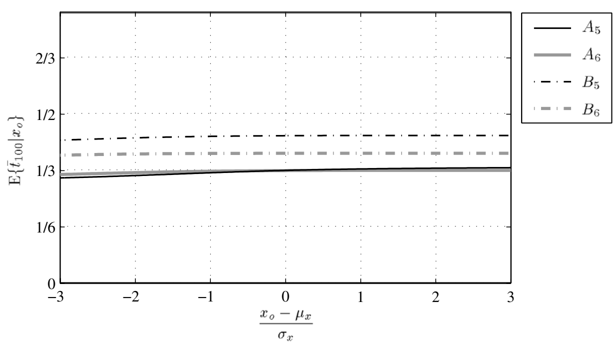

Fig. 5. Location-fairness of the adaptive threshold policy

C. Power-Efficiency and Throughput Comparisons

Recall from Section V-C that each point in the gray region of Figure 2 corresponds to a realization of the simplified threshold policy. It turns out that each of these points also corresponds to a realization of the adaptive threshold policy, for some r_d, $\tau_{\tilde{x}}$, and L. Hence, the simplified and adaptive threshold policies have *identical* performance, in terms of $\mathrm{E}\{\bar{p}\}$ and $\mathrm{E}\{\bar{t}\}$. This implies that the adaptive threshold policy also provides remarkable improvements in power-efficiency and throughput over the constant SNR policy, if r_d, $\tau_{\tilde{x}}$, and L are chosen appropriately.

The thin solid curves in Figure 2 are used to illustrate the effect of L on performance. The curve labeled "$L = 2$" is such that, if a realization of the adaptive threshold policy has $L = 2$, the resulting $\mathrm{E}\{\bar{p}\}$ and $\mathrm{E}\{\bar{t}\}$ must lie on or to the right of this curve, irrespective of r_d and $\tau_{\tilde{x}}$. The curves labeled "$L = 20$", "$L = 200$", and "$L = 200000$" are obtained similarly. Hence, a relatively large L is necessary for power-efficiency and throughput improvements to be substantial.

D. Improvements in Location-Fairness and Downtime

Recall from Sections V-D and V-E that the simplified threshold policy is location-unfair and has long downtime, relative to the constant SNR policy. Here, we show that these drawbacks can be alleviated to a certain extent using the adaptive threshold policy. Specifically, for points A_5, A_6, B_5, and B_6 of Figure 2, we construct realizations of the adaptive threshold policy, which ensure excellent location-fairness and shorter downtime.

	$\mathrm{E}\{d\|d > 0\}$
A_5	9.13
A_6	6.36

	$\mathrm{E}\{d\|d > 0\}$
B_5	5.96
B_6	4.86

TABLE IV
MEAN DOWNTIME OF THE ADAPTIVE THRESHOLD POLICY

Table III lists realizations of the adaptive threshold policy for points A_5, A_6, B_5, and B_6 of Figure 2. Figure 5 represents the resulting location-fairness, and Table IV lists the resulting mean downtime.

Comparing Figures 3 and 4 with Figure 5, we observe that, for points A_5, A_6, B_5, and B_6, realizations of the adaptive threshold policy ensure excellent location-fairness not achievable by realizations of the simplified threshold policy. Comparing Table II with Table IV, we observe that, for points A_5, A_6, B_5, and B_6, realizations of the adaptive threshold policy, on the average, reduce mean downtime by 12%.

Although realizations of the adaptive threshold policy for points A_1–A_4 and B_1–B_4 can also be constructed, the resulting improvements in location-fairness and downtime are insignificant. Therefore, we conclude that the adaptive threshold policy is superior to the simplified threshold policy when the power-efficiency and throughput improvements over the constant SNR policy are moderate, i.e., up to $3.6\,\mathrm{dB}$ and 30%, respectively.

VII. CONCLUSIONS

In this paper, power-efficient operation of wireless networks is analyzed. Under general assumptions, it is proved that the optimal transmission policy, which ensures a desired throughput, is necessarily of threshold nature. Unfortunately, detailed properties of this policy are fairly complicated and sensitive to channel and communication system models. Therefore, we formulated and analyzed a simplified threshold policy defined by two parameters—the desired SNR, r_d, and threshold, τ. We showed that this policy, optimized with respect to r_d and τ, offers up to $11\,\mathrm{dB}$ power-efficiency or 90% throughput improvement, compared to the constant SNR policy. Further analysis revealed that this policy lacks location-fairness and leads to long downtime. Although this behavior might be acceptable in data communications, voice and other delay-sensitive applications might not tolerate these deficiencies. Therefore, we proposed and analyzed an adaptive threshold policy, according to which τ is adapted using a moving average of the log-channel gain. We showed that this policy recovers location-fairness and reduces downtime. In this case, $3.6\,\mathrm{dB}$ power-efficiency or 30% throughput improvement can be achieved. Based on the above, this policy may be recommended as an alternative to the constant SNR policy for delay-sensitive applications.

REFERENCES

[1] J. M. Rulnick and N. Bambos, "Mobile power management for wireless communication networks," *Wireless Networks*, vol. 3, no. 1, pp. 3–14, 1997.

[2] J. M. Rulnick and N. Bambos, "Power-induced time division on asynchronous channels," *Wireless Networks*, vol. 5, no. 2, pp. 71–80, 1999.

[3] A. J. Goldsmith and S.-G. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Transactions on Communications*, vol. 45, no. 10, pp. 1218–1230, 1997.

[4] S. W. Kim and A. J. Goldsmith, "Truncated power control in code-division multiple-access communications," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 3, pp. 965–972, 2000.

[5] P. T. Kabamba, S. M. Meerkov, and C. Y. Tang, "Optimal, suboptimal and adaptive threshold policies for power efficiency of wireless networks," Control Group Report No. CGR-02-03, University of Michigan, Ann Arbor, MI, 2002, also submitted for publication in *IEEE Transactions on Information Theory*.

[6] P. T. Kabamba, S. M. Meerkov, W. E. Stark, and C. Y. Tang, "Feedforward control of data rate in wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 51, no. 5, pp. 1206–1222, 2002.

[7] G. J. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence," *IEEE Transactions on Vehicular Technology*, vol. 42, no. 4, pp. 641–646, 1993.

[8] S. A. Grandhi, J. Zander, and R. Yates, "Constrained power control," *International Journal of Wireless Personal Communications*, vol. 1, no. 4, pp. 257–270, 1994.