Adaptive Transmission With Finite Code Rates

Lang Lin, *Member, IEEE*, Roy D. Yates, *Member, IEEE*, and Predrag Spasojević, *Member, IEEE*

Abstract—This work examines a transmission system which adapts a finite set of code rates and a continuously varying transmit power. We propose a technique for finding the average reliable throughput (ART)-maximizing policy satisfying an average power constraint for a slow fading additive white Gaussian noise (AWGN) channel. ART is a measure motivated by the information outage and can, for example, be argued to characterize the long-term average throughput of a data packet transmission system with a transmit queue and a feedback protocol which requests retransmission of erroneously received packets.

Given the size of the code rate set L, the ART-maximizing policy has the following properties.

1. For a given set of code rates, the optimum allocation policy suggests quantizing the fading state space into a set of $L + 1$ corresponding intervals. For each quantization interval the optimal policy specifies a minimum transmitted power assignment which guarantees zero information outage. The optimum average power assignments across quantization intervals have a waterfilling relationship with respect to the interval channel quality measure.

2. The joint optimization of quantization intervals and the corresponding rate assignments are shown to have multiple local maxima. Nevertheless, this optimization problem can be reduced to a simple one-dimensional search over a parameter which determines the outage interval.

Numerical results show that, in a Rayleigh-fading channel, there is only a 1-dB gap between the ergodic capacity and the throughput of a two-rate adaptive transmission system when the throughput is less than 6 bits/s/Hz. A special case of our optimal policy assignment is the optimal power and rate policy for an adaptive M-QAM system.

Index Terms—Adaptive coded modulation, adaptive transmission, average reliable throughput (ART), bit loading, block-fading channels, delayed constrained communications, information outage, M-QAM.

I. INTRODUCTION

IN [1], two adaptive transmission policies are compared: the truncated inversion policy with continuously variable transmitted power but a fixed code rate and the fixed transmitted power with continuously variable code rates. It is shown that the latter is superior in terms of the maximum average throughput. Both policies represent limiting cases of practical policies that support a finite set of code rates and power levels.

In [2], the optimization of a discrete adaptive transmission design based on information outage has been studied. In this case, the problem formulation was limited to the case where the number of code rates and transmitted power levels are the same. However, this restriction may not necessarily reflect practical design constraints; it is common to have significantly fewer code rates than power levels. For instance, in an IS-95 system [3], which has been available commercially for more than 10 years, the transmission power adapts on a grid of 1-dB steps over a dynamic range of 60 dB or more. On the other hand, even for the most recent adaptive system designs [4], [5], the number of code rates is only around 10.

In this paper, we examine an extreme case of a system design with a finite number of code rates and a continuously variable transmitted power. Though similar problems for adaptive M-QAM systems have been studied in [6], [7], this paper emphasizes the optimum system design with the physical constraints of average power and a number of codebooks. The primary design problem will be the selection of code rates and the corresponding assignments of rate and power for any given channel state.

We assume additive white Gaussian noise (AWGN) and a slow multiplicative fading environment with a channel state which is constant during the transmission of a codeword. It is assumed that the exact current channel state information is known at both the transmitter and the receiver. The channel state space is partitioned into a countable number of intervals. Upon each transmission, a message is encoded at a rate corresponding to the current channel state interval and a power level corresponding to the current channel state. Since each codeword experiences an AWGN channel, random Gaussian codes organized in multiple codebooks are employed. Similar scenarios with channel state uncertainty are examined in [8].

For the proposed adaptive system, it is possible that the instantaneous mutual information corresponding to a channel state is less than the assigned code rate. In this case, an information outage event occurs. The information outage is an intrinsic characteristic of communications over fading channels with a decoding delay constraint [9], [10] or, alternatively, with codewords not long enough to experience ergodic fading. One extreme case of such a scenario is the above-mentioned slow-fading channel assumption where the channel is constant during the transmission of a codeword. For delay-limited cases, the strict sense Shannon capacity is zero [9]. During an outage, a transmission is not considered reliable and, thus, it is frequently convenient to assume that the transmitted data can be ignored [11]. This assumption leads to the capacity versus outage problem which focuses on the tradeoff between the outage probability and the supportable rate; see, for example, [12], [13]. The practice of ignoring data received during an

outage is supported by the fact that the outage probability matches well the error probability of actual codes [14], [15]. Consequently, we characterize the performance of a system design based on the concept of average reliable throughput (ART), defined as the average data rate assuming zero rate when the channel is in outage [2].

As a motivation for the ART metric, consider a system where the transmitter has a transmit queue, and in each block, a packet of $r(s)n$ bits, where n is the length in dimensions of the slot, and s is the fading state, is transmitted over the channel at rate $r(s)$. In addition, suppose that any decoding error is revealed with probability 1. When an information outage occurs and results in a decoding error, the corresponding packet is left in the queue; otherwise, the packet is removed from the queue. In this case, the ART is the average service rate. Under mild conditions, if the queue is fed by an ergodic stationary arrival process of rate λ, also measured in bits per dimension, then the queue is stable if and only if λ is less than ART. Since most existing data packet systems have some feedback protocols that ask for retransmission of erroneously received packets, ART characterizes the long-term average throughput of such systems and determines the arrival rates for which there exists a policy that makes the queue stable.

Following the formulation of the finite code rate set problem, we explore optimum (ART-maximizing) policies where a *policy* is defined by a channel state space partition together with the corresponding transmitted power and rate allocation. In this work, we will show that, for an optimum policy, there are no outages in the sense that the transmitted power is nonzero only if the sender employs a code rate that allows for reliable decoding at the receiver. Furthermore, the transmitter employs a zero transmission power policy only for a subset of worst channel states. We show that an optimum policy employing L codebooks can uniquely be characterized by a partition of the channel state space with $L + 1$ intervals: the zero rate/power interval in addition to L intervals corresponding to L nonzero code rates. In particular, the optimum power allocation has a water-filling character which is uniquely determined by the channel state space partition.

Unlike in [2], the joint optimization over the channel state space partitions and the code rate/power assignments considered here can be reduced to a one-parameter search. This outcome is quite surprising since the throughput maximization problem is generally not a convex optimization problem and can have multiple local maxima for an arbitrary distribution of the channel state. In this paper, the channel state distribution is only constrained to be continuous and differentiable.

In addition, we have obtained the optimum partition for a given set of rates to be assigned and a given average power constraint. Such a solution is of particular interest to practical heuristic designs where codes can be selected only from a limited set of good channel codes before one chooses an optimum channel state partition. Though not surprisingly, our derivation also shows that a partition of $L + 1$ channel state intervals is indeed the optimum choice. As a byproduct, we also provide an optimum solution for the M-QAM spectral efficiency maximization problem introduced in [6] and addressed in [7], [16].

Finally, numerical results show that, for a Rayleigh-fading channel, there is a gap of only 1 dB between the ergodic capacity and the throughput of a two-rate adaptive transmission system when the throughput is less than 6 bits/s/Hz. Thus, in comparison with the results in [2], we find that power adaptation can indeed be very helpful for adaptive transmission with discrete code rates. This agrees with the conclusion for power adaptation in M-QAM systems in [7].

II. SYSTEM MODEL AND PROBLEM FORMULATION

We consider a multiplicative flat-fading channel model similar to that in [1]. The complex received signal

$$Y = \sqrt{S}X + W \qquad (1)$$

where S is the channel (fading) state, X is the complex transmitted signal, and W is a circularly symmetric AWGN with variance N_0. The channel state S is a real random variable of unit mean with a probability density function (pdf) $f(s)$, a continuous cumulative distribution function (cdf) $F(s)$, and a domain $\mathcal{S} = \{s | s \geq 0\}$. Only in Section IV, is $F(s)$ assumed to be continuous, differentiable, and strictly increasing in s. In this paper, fading is assumed to be sufficiently *slow* so that the channel state is constant during the transmission of a codeword.

A generalized adaptive transmission system can be modeled as follows. At any channel state s, the transmitter transmits codewords coded at a rate $r(s)$ with a power level $p(s) = E\{|X|^2 |s\}$, where $E\{\cdot\}$ denotes expectation. The code rate $r(s)$ is chosen from a set $\mathcal{R}_0 = \{r_0 = 0\} \cup \mathcal{R}$ where $\mathcal{R} = \{r_1, \ldots, r_L\}$. Without loss of generality, we assume that $r_{l-1} < r_l$ for $l = 1, \ldots, L$. An adaptive *transmission policy* can uniquely be specified by the assigned code rate $r(s)$ and the corresponding allocated power level $p(s)$. Such a policy is denoted by the tuple $(\mathcal{R}, r(s), p(s))$. Let

$$\mathcal{V}_l = \{s | r(s) = r_l\}, \qquad l = 0, \ldots, L \qquad (2)$$

denote the set of channel fading states in which rate r_l is employed. Each policy specifies a partition of the set of channel states $\mathcal{S} = \cup_{l=0}^{L} \mathcal{V}_l$. In general, $\mathcal{V}_l$ can be a countable union of intervals in $\mathcal{S}$ or simply a measurable set of channel states. Since there are L nonzero code rates, we call $(\mathcal{R}, r(s), p(s))$ an L-level policy.

Since (1) is an AWGN channel for any given $s \in \mathcal{S}$, the corresponding maximum mutual information is given by $\log(1 + p(s)s/N_0)$. Adopting the notation

$$R(\phi) = \log(1 + \phi/N_0), \qquad \phi \geq 0 \qquad (3)$$

the maximum mutual information associated with any state s is $R(p(s)s)$. For any $s \in \mathcal{V}_l$, given a code rate r_l of a capacity-achieving Gaussian codebook, and a power allocation $p(s)$, the information is guaranteed to be successfully received iff $R(p(s)s) \geq r_l$. We define the binary outage indicator function

$$I_{\text{out}}(r, s, p(s)) = \begin{cases} 1, & R(p(s)s) < r \\ 0, & \text{otherwise.} \end{cases} \qquad (4)$$

The expected value of $I_{\text{out}}(r, s, p(s))$ over the channel states S is the information outage introduced in [13].

Given a policy $(\mathcal{R}, r(s), p(s))$, the ART is

$$R_L(\mathcal{R}, r(s), p(s)) = \sum_{l=0}^{L} r_l \int_{\mathcal{V}_l} \left[1 - I_{\text{out}}(r_l, s, p(s))\right] f(s) \, ds. \tag{5}$$

Let

$$f_l(s) = \begin{cases} f(s)/\Pr\left[s \in \mathcal{V}_l\right], & s \in \mathcal{V}_l \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

denote the conditional pdf of the channel state given that it belongs to $\mathcal{V}_l$. The conditional average power given $s \in \mathcal{V}_l$ is

$$P_l = \int_{\mathcal{V}_l} p(s) f_l(s) \, ds. \tag{7}$$

Also, the average transmitted power for the policy $(\mathcal{R}, r(s), p(s))$ is

$$\rho_L(\mathcal{R}, r(s), p(s)) = \sum_{l=0}^{L} P_l \Pr\left[s \in \mathcal{V}_l\right]. \tag{8}$$

Throughout this paper, we only consider the policies with $r(s)$ and $p(s)$ such that both (5) and (7) are meaningful, i.e., the integrals are either Riemann or Lebesgue integrable. The objective will be to maximize ART $R_L(\mathcal{R}, r(s), p(s))$ subject to an average power constraint and a constraint that there are L codebooks of L distinct nonzero rates

$$\max_{\mathcal{R}, r(s), p(s)} \quad R_L(\mathcal{R}, r(s), p(s)) \tag{9}$$

$$\text{subject to} \quad \rho_L(\mathcal{R}, r(s), p(s)) \leq \overline{p} \tag{9a}$$

$$|\mathcal{R}| = L \tag{9b}$$

$$r(s) \in \mathcal{R}_0 = \{r_0 = 0\} \cup \mathcal{R} \tag{9c}$$

$$p(s) \geq 0 \tag{9d}$$

$$r(s) \geq 0. \tag{9e}$$

In (5), for any $\mathcal{V}_l$ with either $r_l = 0$ or $P_l = 0$, there is zero contribution toward ART and, consequently, it is optimum to assign $p(s) = 0$ over all such $\mathcal{V}_l$. For a $\mathcal{V}_l$ with nonzero r_l and P_l, we observe that an optimum policy must be locally optimum over $\mathcal{V}_l$. Local optimality requires that given $\mathcal{V}_l$, P_l, and r_l, $p(s)$ must be the solution of

$$\max_{p(s)} \quad r_l \int_{\mathcal{V}_l} \left[1 - I_{\text{out}}(r_l, s, p(s))\right] f_l(s) \, ds \tag{10}$$

$$\text{subject to} \quad \int_{\mathcal{V}_l} p(s) f_l(s) \, ds \leq P_l \tag{10a}$$

$$p(s) \geq 0. \tag{10b}$$

Given any nonnegative r_l, (10) is equivalent to the following local outage minimization problem:

$$\min_{p(s)} \quad \int_{\mathcal{V}_l} I_{\text{out}}(r_l, s, p(s)) f_l(s) \, ds \tag{11}$$

$$\text{subject to} \quad \int_{\mathcal{V}_l} p(s) f_l(s) \, ds \leq P_l \tag{11a}$$

$$p(s) \geq 0. \tag{11b}$$

The solution of (11) is presented in [12, Proposition 4] and will be summarized here. We define a channel inversion power allocation

$$\psi(s, r) = \begin{cases} \frac{N_0}{s} \left(e^r - 1\right), & s > 0, r > 0 \\ 0, & r = 0 \end{cases} \tag{12}$$

which represents the minimum power required to communicate reliably at rate r for channel state s. Let

$$r_l^\dagger = \max \left\{ r \,\middle|\, \int_{\mathcal{V}_l} \psi(s, r) f_l(s) \, ds \leq P_l \right\} \tag{13}$$

denote the largest possible assigned rate over $\mathcal{V}_l$ without outage given the average power allocation P_l over $\mathcal{V}_l$. If $r_l \leq r_l^\dagger$, the solution of (11) is trivial; we allocate power $\psi(s, r_l)$ and achieve zero outage over $\mathcal{V}_l$. On the other hand, if $r_l > r_l^\dagger$, outage within $\mathcal{V}_l$ is inevitable and the corresponding optimum power allocation is the truncated channel inversion

$$p(s) = \begin{cases} 0, & s \in \mathcal{V}_l \backslash \bar{\mathcal{V}}_l \\ \psi(s, r_l), & s \in \bar{\mathcal{V}}_l \end{cases} \tag{14}$$

where $\bar{\mathcal{V}}_l \subset \mathcal{V}_l$ is a subset of better channel states in $\mathcal{V}_l$, i.e., $s_1 \in \mathcal{V}_l \backslash \bar{\mathcal{V}}_l$ and $s_2 \in \bar{\mathcal{V}}_l$ implies $s_1 < s_2$. Moreover, the set $\bar{\mathcal{V}}_l$ is chosen to satisfy the average power constraint

$$P_l = \int_{\bar{\mathcal{V}}_l} \psi(s, r_l) f_l(s) \, ds. \tag{15}$$

Note that $\psi(s, r)$ is not defined for $s = 0$ and $r > 0$. This is due to the fact that, in the vicinity of $s = 0$, the channel is too poor to support any positive rate. Thus, the following must hold: $r(0) = 0$ and $p(0) = 0$.

In the optimization problem (10), we assume that r_l is known. However, for any $\mathcal{V}_l$, given the conditional average power P_l, we can also choose r_l to maximize the conditional ART. Specifically, local optimality implies that the optimum rate/power allocation must solve

$$\max_{r_l} \max_{p(s)} \quad r_l \int_{\mathcal{V}_l} \left[1 - I_{\text{out}}(r_l, s, p(s))\right] f_l(s) \, ds \tag{16}$$

$$\text{subject to} \quad \int_{\mathcal{V}_l} p(s) f_l(s) \, ds \leq P_l \tag{16a}$$

$$p(s) \geq 0. \tag{16b}$$

Since rate $r_l^\dagger$ defined by (13) is achievable under (16), any rate $r_l < r_l^\dagger$ is suboptimal for (16). For any $r_l' > r_l^\dagger$, outage probability within $\mathcal{V}_l$ is nonzero. Thus, for any optimum rate r_l, the optimum power allocation in (16) has the form (14). Since there is no transmission for $s \in \mathcal{V}_l \backslash \bar{\mathcal{V}}_l$, we can incorporate this set of channel states into the set of zero-power zero-rate channel states $\mathcal{V}_0 = \{s | r(s) = 0\}$ and redefine $\mathcal{V}_l = \bar{\mathcal{V}}_l$. For the new policy, there is no outage for states $s \in \mathcal{V}_0$; we reliably achieve zero rate by using zero power. Moreover, the assigned rate $r(s)$ and the corresponding power $p(s)$ satisfy

$$p(s) = \psi(s, r_l), \qquad s \in \mathcal{V}_l \tag{17}$$

$$r(s) = \log(1 + p(s)s). \tag{18}$$

Note that (17) and (18) are consistent with the definition of $\mathcal{V}_l$ which implies that $s \in \mathcal{V}_l$ whenever $r(s) = r_l$. Consequently, it

is clearly sufficient to denote any optimum policy by $(\mathcal{R}, r(s))$. Moreover, (18) and (9d) imply (9e).

From (5), (7), and (8), the corresponding overall average power and ART are

$$\rho_L(\mathcal{R}, r(s)) = \sum_{l=1}^{L} \int_{\mathcal{V}_l} \psi(s, r_l) f(s)\, ds \tag{19}$$

$$R_L(\mathcal{R}, r(s)) = \sum_{l=1}^{L} \int_{\mathcal{V}_l} r_l f(s)\, ds. \tag{20}$$

Our objective is to maximize $R_L(\mathcal{R}, r(s))$ subject to both the average power constraint $\rho_L(\mathcal{R}, r(s)) \leq \overline{p}$ and the constraint that the rate set $\{r(s)\}$ has a cardinality $L + 1$ and includes the zero rate $r_0 = 0$.

Even though (20) is much simpler than (5), a simple solution is not available. In the next two sections, we will obtain necessary and sufficient conditions for optimum policies by applying the Lagrange multiplier method and the Karush–Kuhn–Tucker (KKT) conditions [17]. Given these conditions, the optimum policies can be found by a relatively simple search method.

III. PARTITION OPTIMIZATION

In this section, we address the subproblem of finding the optimum $r(s)$ given a specific rate set $\mathcal{R}_0$. This problem is of interest since a valid strategy for designing adaptive transmission systems is to choose a subset of good error control codes before deciding $r(s)$, i.e., the channel state partition $\{\mathcal{V}_l\}$. Furthermore, since this subproblem and the M-QAM spectral efficiency maximization problem in [6] (solved only in a suboptimal manner in [6]) are closely related, the optimum solution presented here is also the optimum solution of the problem in [6].

For a given rate set $\mathcal{R}$ including L distinct positive rates, the throughput maximization problem (9) becomes

$$\max_{r(s)} \quad R_L(\mathcal{R}, r(s)) \tag{21}$$

$$\text{subject to} \quad \rho_L(\mathcal{R}, r(s)) \leq \overline{p} \tag{21a}$$

$$r(s) \in \mathcal{R}_0. \tag{21b}$$

In comparison with the problem (9), (21) has fewer constraints. Specifically, given $\mathcal{R}$, with (21b), (9b) and (9e) are redundant. In addition, with (21a), we implicitly take $p(s)$ in the form specified in (17), which further implies that (9d) is automatically satisfied.

The maximization problem (21) is a variation of the bit-loading problem in [18]. Both problems belong to the general class of Knapsack problems (KPs) [19]. The traditional bit-loading problem is to optimize the rate/power allocation over a finite number of parallel channels where any rate assigned to a channel can only be an integer. In this work, the rate/power allocation is over $\mathcal{S}$, which is an uncountable set. Furthermore, elements of $\mathcal{R}$ are not necessarily integers.

A. Related KPs

A good example of a KP is the Fractional Knapsack problem (FKP). For an FKP, there is a knapsack of size $c > 0$ and an integer number n objects each with a reward value $r_i > 0$ and

a weight $w_i > 0$, $i = 1, \ldots, n$. An FKP studies how to fill the knapsack in order to maximize the total value of the objects inside the knapsack

$$\max_{(x_1, \ldots, x_n)} \quad \sum_{i=1}^{n} x_i r_i \tag{22}$$

$$\text{subject to} \quad \sum_{i=1}^{n} x_i w_i \leq c \tag{22a}$$

$$x_i \in [0, 1], \qquad i = 1, \ldots, n. \tag{22b}$$

Because the FKP allows a fraction of an object to be placed in the knapsack, (22) can be solved by a greedy algorithm that keeps putting the objects into the knapsack in descending value-size ratio r_i / w_i order until the capacity c is reached.

Without loss of generality, we assume that the value-size ratios corresponding to the n objects are *distinct*. Therefore, an optimum solution can be described by two parameters, i.e., a boundary value-size ratio α and a fraction value $x \in [0, 1]$. In order to achieve the optimum, all objects with a value-size ratio larger than α must be put into the knapsack. In addition, there is an item with a value-size ratio equal to α and only a fraction x of that item is put in the knapsack. The rest of the objects are left untouched.

When x_i is constrained to be 0 or 1 in (22), the problem becomes a 0–1 KP. The optimum solution of a 0–1 KP can be obtained by dynamic programming using Bellman recursion [20]. A more efficient algorithm can be obtained based on the results in [21]. Different from FKP, the optimum solution of a 0–1 KP cannot typically be described analytically by a few parameters. However, it should be clear that the solution of an FKP provides an upper bound of the solution of a 0–1 KP.

A more complex knapsack problem is the generalized multiple-choice knapsack problem (GMCKP) introduced by Pisinger [22]. For a GMCKP, there are L knapsacks. For knapsack l, there is a set of n_l objects where the ith object has reward r_{li} and weight w_{li}, respectively. A simplified special case of the GMCKP is then

$$\max_{\{x_{li}\}} \quad \sum_{l=1}^{L} \sum_{i=1}^{n_l} x_{li} r_{li} \tag{23}$$

$$\text{subject to} \quad \sum_{l=1}^{L} \sum_{i=1}^{n_l} x_{li} w_{li} \leq c \tag{23a}$$

$$x_{li} \in \{0, 1\}, \quad l = 1, \ldots, L, i = 1, \ldots, n_l. \tag{23b}$$

Again, a GMCKP does not typically have a simple analytical solution due to (23b). With a proper procedure, it is demonstrated in the next subsection that (23) is similar to the fixed rate set problem (21). For (21), the number of knapsacks is infinite and all objects can only contribute an infinitely small amount in either value or size. Hence, similar to FKP, we can find a simple analytical solution for (21). However, a practical system may not have the advantage from the fractional scenario and, thus, the references provided here become useful.

B. Optimum Partition

In [18] (see also [23]), it is proved that a greedy rate/power allocation is the optimum solution for the bit-loading problem
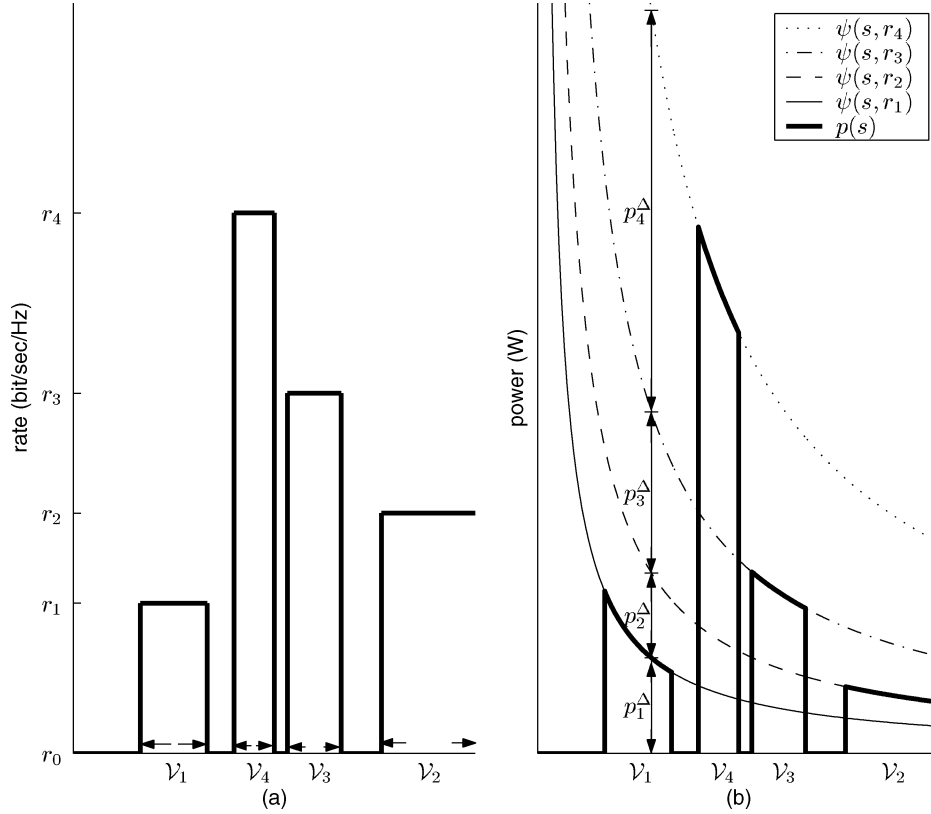
Fig. 1.   Illustration of an arbitrary policy satisfying (17) and (18). Note that $\mathcal{V}_0 = \mathcal{S}\backslash\cup_{l=1}^{4}\mathcal{V}_l$ is not shown. (a) Rate assignment. (b) Power allocation.

in parallel channels. We will show that the same allocation is indeed optimum for (21). Moreover, such a solution implies that quantization is the optimum partition. The optimum quantization partition boundaries are also derived.

Fig. 1 depicts a policy with an arbitrary, and not particularly intelligent, rate and partition assignment. The wave-like power assignment in Fig. 1(b) is the result of the local optimization in (16). The average rate $R_L(\mathcal{R}, r(s))$ in (20) is the integral of the rate pulses in Fig. 1(a) weighted by the channel state pdf $f(s)$. Similarly, the average power $\rho_L(\mathcal{R}, r(s))$ in (19) is the integral of the wave crests of Fig. 1(b) weighted by $f(s)$. To express these integrals in a more useful form, we introduce the incremental rate and power functions

$$r_l^\Delta = r_l - r_{l-1} \tag{24}$$
$$p_l^\Delta(s) = \psi(s, r_l) - \psi(s, r_{l-1}) \tag{25}$$

for $l = 1, \ldots, L$. Since elements in $\mathcal{R}$ are a strictly increasing sequence, the incremental rates r_l^Δ and power $p_l^\Delta(s)$ are both positive.

At any channel state s, the transmitted code rate and the corresponding required transmitted power assignments can be expressed as sums of r_l^Δ and $p_l^\Delta(s)$

$$r(s) = \sum_{l=1}^{L} I_l(s) r_l^\Delta \tag{26}$$

$$p(s) = \sum_{l=1}^{L} I_l(s) p_l^\Delta(s) \tag{27}$$

where the coefficients, $\{I_l(s)\,|\, l = 1, \ldots, L\}$, in (26) and (27) is a set of binary 0/1-value functions. Since $r_l \geq r_{l'}$ for $l' = 1, \ldots, l$, (26) implies that

$$\mathcal{V}_l = \left\{ s \,\Big|\, I_{l'}(s) = 1, l' \leq l, \text{ and } I_{l'}(s) = 0, l' > l \right\}. \tag{28}$$

A given rate set $\mathcal{R}$ specifies the incremental rates $\{r_l^\Delta\}$ and the incremental powers $\{p_l^\Delta(s)\}$. Thus, given $\mathcal{R}$, $\mathcal{I} = \{I_l(s)\,|\, l = 1, \ldots, L\}$ describes a policy of interest. The corresponding ART and average power are

$$R_L(\mathcal{I}) = \sum_{l=1}^{L} \int_0^\infty I_l(s) r_l^\Delta f(s)\, ds \tag{29}$$

$$\rho_L(\mathcal{I}) = \sum_{l=1}^{L} \int_0^\infty I_l(s) p_l^\Delta(s) f(s)\, ds. \tag{30}$$

The maximization problem (21) now becomes a problem of searching for the optimum $\mathcal{I}$. Note that (28) requires a valid policy to satisfy the precedence constraint

$$I_{l'}(s) \geq I_l(s), \qquad l' \leq l. \tag{31}$$

The precedence constraint simply says that if $I_l(s) = 1$, then $I_{l'}(s) = 1$ for all $l' \leq l$. In addition, if $I_l(s) = 0$, then $I_{l'}(s) = 0$ for all $l' > l$. Later, it is shown that the precedence constraint can be lifted because it is always satisfied by an optimal policy. In the absence of the precedence constraint, the maximization problem (21) can be viewed as a version of the GMCKP (23) in which each knapsack can have an infinite number of objects

and the objects have infinitesimal rewards $r_{li} = r_l^\Delta f(s)\,ds$ and weights $w_{li} = p_l^\Delta(s)f(s)\,ds$.

In order to identify a policy that provides an optimum solution to the throughput maximization problem (21), we introduce the *incremental efficiency* (or, simply, *efficiency*)

$$\eta_l(s) = \begin{cases} \dfrac{r_l^\Delta f(s)}{p_l^\Delta(s)f(s)} = \dfrac{s}{N_0}\dfrac{r_l - r_{l-1}}{e^{r_l} - e^{r_{l-1}}}, & f(s) \neq 0 \\ 0, & f(s) = 0 \end{cases} \quad (32)$$

which is a ratio between an increment in the throughput from $r_{l-1}f(s)$ to $r_l f(s)$ at state s and the corresponding power expenditure $p_l^\Delta(s)f(s)$. Note that in (32), it requires nonzero $p_l^\Delta(s)$ for $s \in \mathcal{S}$ that is satisfied by the definitions of $\psi(s, r)$ and $p_l^\Delta(s)$ in (12) and (25), respectively.

For optimization with integer rates, the efficiency concept may not be necessary [23]. Nevertheless, it is a key for solving the problem with noninteger code rates.

Lemma 1: The incremental efficiency $\eta_l(s)$ has the following properties:

a) $\eta_l(0) = 0$ for all l;
b) for fixed l, $\eta_l(s)$ increases in s;
c) for fixed s, $\eta_l(s)$ decreases in l.

From (29), ART can now be expressed in terms of the incremental efficiency as

$$R_L(\mathcal{I}) = \sum_{l=1}^{L} \int_0^\infty I_l(s)\eta_l(s)p_l^\Delta(s)f(s)\,ds. \quad (33)$$

Together with the constraint (30), the maximization of $R_L(\mathcal{I})$ forms a KP [19], which is solved by the following policy.

Definition 1: The most power-efficient quantization (MPEQ) is a policy $\mathcal{I}^* = \{I_l^*(s)\}$ where

$$I_l^*(s) = \begin{cases} 1, & \eta_l(s) \geq \lambda_M(\boldsymbol{r}) \\ 0, & \text{otherwise} \end{cases} \quad l = 1, \ldots, L \quad (34)$$

and the positive constant $\lambda_M(\boldsymbol{r})$ is determined by the average power constraint $\rho_L(\mathcal{I}^*) = \overline{p}$.

The existence of an MPEQ is guaranteed. The procedure for finding the MPEQ solution is the same as that for an FKP. Starting from all $I_l(s) = 0$, we repeatedly assign $I_l(s) = 1$ to the highest remaining $\eta_l(s)$ until $\overline{p}$ is reached. Such a procedure is always successful since in (30), $p_l^\Delta(s)$ is continuous and $f(s)$ is bounded due to the continuity of $F(s)$. Note that the uniqueness of the efficiency lower bound $\lambda_M(\boldsymbol{r})$ and MPEQ is an issue discussed later in the subsection.

For any MPEQ, $s \in \mathcal{V}_{l-1}$ and $s' \in \mathcal{V}_l$ imply that for some $\lambda_M(\boldsymbol{r}) > 0$, $I_l^*(s) = 0$ and $I_l^*(s') = 1$, i.e., $\eta_l(s) < \eta_l(s')$ and, consequently, $s < s'$. Thus, MPEQ leads to a quantization in the sense that $s \in \mathcal{V}_{l-1}$ and $s' \in \mathcal{V}_l$ imply that $s < s'$. In addition, Lemma 1 part c) guarantees that MPEQ satisfies the precedence constraint (31).

Theorem 1: For any policy $\mathcal{I} = \{I_l(s)\}$ satisfying the average power constraint $\rho_L(\mathcal{I}) \leq \overline{p}$

$$R_L(\mathcal{I}) \leq R_L(\mathcal{I}^*). \quad (35)$$

We emphasize that Theorem 1 holds regardless of whether the policy $I_l(s)$ satisfies the precedence constraint (31) or not.

The MPEQ solution is easy to obtain analytically. An MPEQ policy is fully determined by the efficiency threshold $\lambda_M(\boldsymbol{r})$ where $r(s) \geq r_l$ iff $\eta_l(s) \geq \lambda_M(\boldsymbol{r})$. From Lemma 1 and the continuity of the domain $\mathcal{S}$, there exists q_l such that $\eta_l(q_l) = \lambda_M(\boldsymbol{r})$ and $\eta_l(s) \geq \lambda_M(\boldsymbol{r})$ iff $s \geq q_l$. Thus, q_l is the boundary separating $\mathcal{V}_{l-1}$ and $\mathcal{V}_l$. It follows from (32) and the equality $\eta_l(q_l) = \lambda_M(\boldsymbol{r})$ that

$$q_l = \frac{\lambda_M(\boldsymbol{r})N_0(e^{r_l} - e^{r_{l-1}})}{r_l - r_{l-1}}, \qquad l = 1, \ldots, L. \quad (36)$$

Since $\eta_l(s)$ decreases in l for fixed s, (36) is an increasing sequence of L boundaries $q_1, \ldots, q_L$. Moreover, these boundaries indicate that for a given set of rates $\mathcal{R}$, an optimum policy is given by a quantization of the channel state set $\mathcal{S}$ into exactly $L + 1$ intervals corresponding to the set of rates $\mathcal{R}$. Thus, an optimum policy for a given $\mathcal{R}$ can be represented by the vector

$$\boldsymbol{q} = [q_1, \ldots, q_L]^\top. \quad (37)$$

An MPEQ policy is illustrated in Fig. 2, where $q_0 = 0$.

To study the uniqueness of MPEQ, we assume $\mathcal{I}^* = \{I_l^*(s)\}$ and $\mathcal{I}^{*\prime} = \{\mathcal{I}_l^{*\prime}(s)\}$ be two MPEQs corresponding to $\lambda_M(\boldsymbol{r})$ and $\lambda_M'(\boldsymbol{r})$, respectively. The corresponding boundaries of the partition intervals are $\boldsymbol{q}$ and $\boldsymbol{q}' = [q_1', \ldots, q_L']^\top$.

Proposition 1: If $\lambda_M(\boldsymbol{r}) > \lambda_M'(\boldsymbol{r})$, we have

a) $I_l^{*\prime}(s) = 1$ wherever $I_l^*(s) = 1$;
b) $\rho_L(\mathcal{I}^*) \leq \rho_L(\mathcal{I}^{*\prime})$;
c) $q_l' < q_l$, for all $l = 1, \ldots, L$;
d) $\rho_L(\mathcal{I}^*) = \rho_L'(\mathcal{I}^{*\prime})$ if $f(s) = 0$, for all $s \in [q_l', q_l)$, $l = 1, \ldots, L$.

Proposition 1 part d) implies the following theorem.

Theorem 2: MPEQ is unique if $f(s) > 0$ for $s \in \mathcal{S} \setminus \{0\}$.

C. MQAM Spectral Efficiency Maximization

Problem (21) is closely related to a special case of the spectral efficiency maximization problem for an adaptive M-QAM system [6], which suggests employing a set of $L + 1$ predetermined M-QAM constellations of sizes in $\mathcal{M} = \{M_0 = 1, M_1, \ldots, M_L\}$. The $M_0 = 1$ constellation corresponds to turning off the transmitter. The transmission scheme requires that given a channel state s in $\mathcal{V}_l$, a transmitter transmits a quadrature amplitude modulation (QAM) symbol constellation of size $M(s) = M_l$, $M_l \in \mathcal{M}$, with power $p(s)$.

Without loss of generality, we assume $M_{l-1} < M_l$ for $l = 1, \ldots, L$. To guarantee a specified bit-error rate (BER) P_b for all channel states, the transmitted power is

$$p(s) = \frac{M(s) - 1}{Ks} \quad (38)$$

where $K = -1.5/\log(5P_b)$ [6].

Let $r_l = \log(M_l)$, $l = 0, \ldots, L$. Consequently, $r(s) = \log(1 + p(s)sK)$. The spectral efficiency maximization prob-
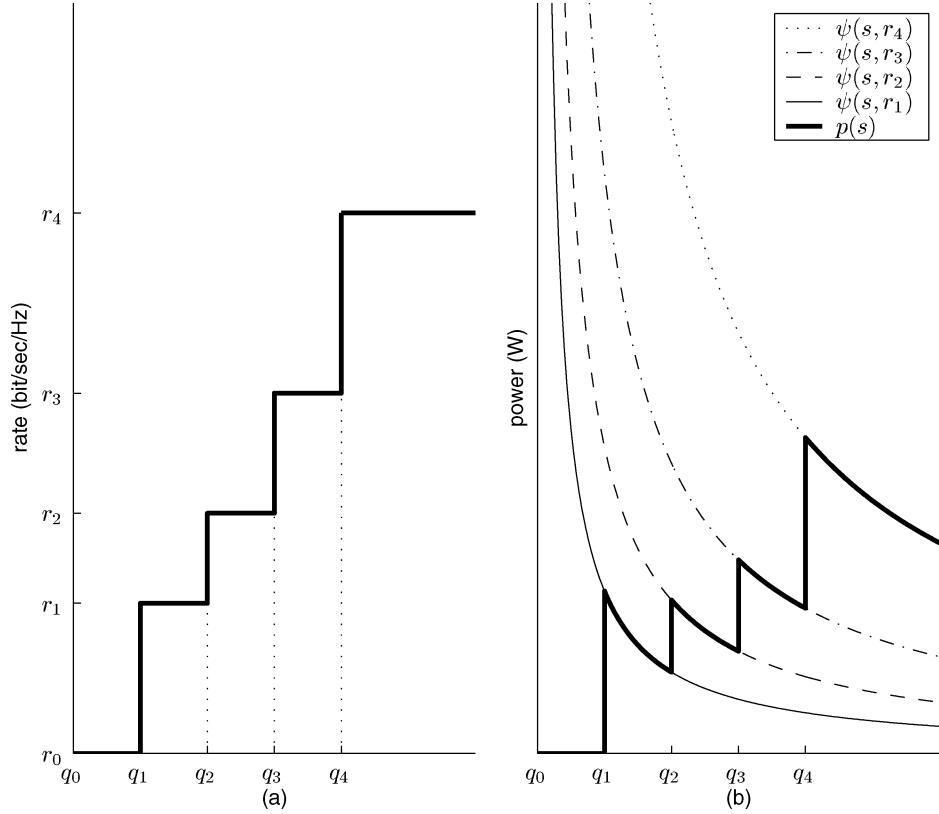
Fig. 2. Illustration of an MPEQ policy. (a) Rate staircase. (b) Power staircase.

lem for $L+1$ QAM signaling and a continuously varying power is as follows:

$$\max_{r(s)} \quad R_L(\mathcal{R}, r(s)) \tag{39}$$

$$\text{subject to} \quad \int_{\mathcal{S}} p(s)f(s)\,ds \leq \overline{p} \tag{39a}$$

$$r(s) \in \mathcal{R}_0 \tag{39b}$$

$$p(s) = \frac{e^{r(s)} - 1}{Ks} = \frac{\psi(s, r(s))}{KN_0} \tag{39c}$$

where $\mathcal{R}_0 = \{r_0\} \cup \mathcal{R}$ and $\mathcal{R} = \{r_1, \ldots, r_L\}$. Following the procedure derived in this section, we obtain the optimum partition as a quantization with boundaries

$$q_l = \frac{\lambda \left(e^{r_l} - e^{r_{l-1}}\right)}{K(r_l - r_{l-1})}, \qquad l = 1, \ldots, L \tag{40}$$

where λ is determined by the average power constraint. An evaluation of this policy in Rayleigh fading is given in Section V-C.

IV. OPTIMUM POLICIES

In the previous sections, we have demonstrated that an ART-maximizing policy, defined in Section II, with a rate set $\mathcal{R}_0$ must be an MPEQ corresponding to $\mathcal{R}_0$. More specifically, Theorem 1 shows that any candidate ART-maximizing policy with a rate set $\mathcal{R}_0$ must have channel states partitioned into $|\mathcal{R}_0|$ intervals. Furthermore, we found that the optimum solution employs rate $r(s) = r_l \in \mathcal{R}_0$ in the interval $\mathcal{Q}_l = [q_l, q_{l+1})$, where q_l is defined in (36) with $q_0 = 0$ and $q_{|\mathcal{R}_0|} = \infty$. In addition

$$p(s) = \psi(s, r_l), \qquad s \in \mathcal{Q}_l, \; l = 0, 1, \ldots, |\mathcal{R}|. \tag{41}$$

Consequently, any policy of interest can be specified by $(\boldsymbol{r}, \boldsymbol{q})$ where the vectors

$$\boldsymbol{r} = [r_1, \ldots, r_{|\mathcal{R}|}]^\top \tag{42}$$

$$\boldsymbol{q} = [q_1, \ldots, q_{|\mathcal{R}|}]^\top \tag{43}$$

with the corresponding $p(s)$ given by (41). An optimum policy is denoted by $(\boldsymbol{r}^*, \boldsymbol{q}^*)$.

Clearly, for a policy specified by $(\boldsymbol{r}, \boldsymbol{q})$, if the q_l are not distinct, the policy degrades to a policy with fewer than $|\mathcal{R}_0|$ distinct rates. Consequently, without loss of generality, we concentrate on $\boldsymbol{q}$ with $|\mathcal{R}|$ distinct elements. For an $L = |\mathcal{R}|$ level policy $(\boldsymbol{r}, \boldsymbol{q})$, the corresponding ART is

$$R_L(\boldsymbol{r}, \boldsymbol{q}) = \sum_{l=1}^{L} r_l F_l \tag{44}$$

where

$$F_l = \Pr\left[s \in \mathcal{Q}_l\right]. \tag{45}$$

The conditional average power in the interval $\mathcal{Q}_l$ is

$$P_l(\boldsymbol{r}, \boldsymbol{q}) = \int_{\mathcal{Q}_l} \psi(s, r_l) f_l(s)\,ds \tag{46}$$

where the conditional pdf of channel state s is

$$f_l(s) = \begin{cases} f(s)/F_l, & s \in \mathcal{Q}_l \\ 0, & \text{otherwise.} \end{cases} \tag{47}$$
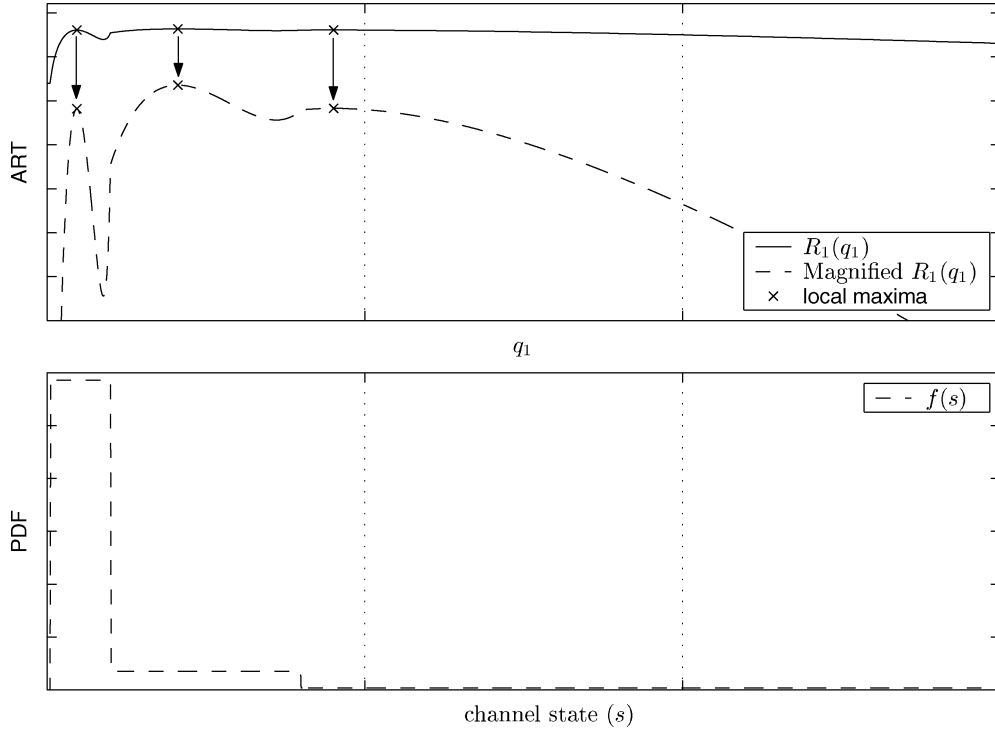
Fig. 3. A $R_1(q_1)$ with a simple piecewise linear $f(s)$ has three local maxima. Arrows show the corresponding local maxima before and after magnification.

For an interval $\mathcal{Q}_l$, the conditional average channel quality ω_l is defined by

$$\frac{1}{\omega_l} = \int_{\mathcal{Q}_l} \frac{N_0}{s} f_l(s)\, ds, \qquad l = 1, \ldots, L. \tag{48}$$

Here, ω_l is normalized by N_0 in order to simplify the latter derivations and $0 < w_1 < \cdots < w_L$. Equations (12), (46), and (48) imply

$$P_l(\boldsymbol{r}, \boldsymbol{q}) = \frac{(e^{r_l} - 1)}{\omega_l}. \tag{49}$$

The original ART maximization problem (9) now becomes

$$\max_{\boldsymbol{r},\boldsymbol{q}} \quad R_L(\boldsymbol{r}, \boldsymbol{q}) \tag{50}$$

$$\text{subject to} \quad \rho_L(\boldsymbol{r}, \boldsymbol{q}) \leq \overline{p} \tag{50a}$$

where

$$R_L(\boldsymbol{r}, \boldsymbol{q}) = \sum_{l=1}^{L} F_l r_l = \sum_{l=1}^{L} r_l \int_{\mathcal{Q}_l} f(s)\, ds \tag{51}$$

$$\rho_L(\boldsymbol{r}, \boldsymbol{q}) = \sum_{l=1}^{L} F_l P_l(\boldsymbol{r}, \boldsymbol{q}) = \sum_{l=1}^{L} (e^{r_l} - 1) \int_{\mathcal{Q}_l} \frac{N_0}{s} f(s)\, ds. \tag{52}$$

We note that problem (50) can have multiple local maxima. For example, when $L = 1$, $\rho_L(\boldsymbol{r}, \boldsymbol{q})$ becomes

$$\rho_1(r_1, q_1) = (e^{r_1} - 1) \int_{q_1}^{\infty} \frac{N_0}{s} f(s)\, ds. \tag{53}$$

Now, since $\rho_1(r_1, q_1) = \overline{p}$ is necessary to achieve the maximum ART, r_1 must be a function of q_1 defined by

$$r_1 = \log\left(1 + \frac{\overline{p}}{\int_{q_1}^{\infty} \frac{N_0}{s} f(s)\, ds}\right). \tag{54}$$

Consequently, $R_L(\boldsymbol{r}, \boldsymbol{q})$ becomes $R_1(r_1, q_1)$ which is simply a function of q_1 given by

$$R_1(q_1) = \log\left(1 + \frac{\overline{p}}{\int_{q_1}^{\infty} \frac{N_0}{s} f(s)\, ds}\right) \int_{q_1}^{\infty} f(s)\, ds. \tag{55}$$

An illustration of $R_1(q_1)$ with multiple local maxima is shown in Fig. 3. Note that in Fig. 3, the $f(s)$ violates the continuity assumption in Section II. However, it is not very hard to imagine that a continuous $f(s)$ similar to that in Fig. 3 will lead to a similar scenario with multiple local maxima.

A. Water-Filling Power Allocation

According to (49), given a channel state partition $\boldsymbol{q}$ that specifies the conditional average channel quality ω_l, the rate r_l is uniquely expressed in terms of the conditional average power P_l in the interval $\mathcal{Q}_l$ as

$$r_l = \log\left(1 + P_l \omega_l\right). \tag{56}$$

Given a fixed $\boldsymbol{q}$, the ART maximization problem (50) can be written as

$$\max_{P_1, \ldots, P_L} \quad \sum_{l=1}^{L} \log\left(1 + \frac{P_l \omega_l}{N_0}\right) F_l \tag{57}$$

$$\text{subject to} \quad \sum_{l=1}^{L} P_l F_l \leq \overline{p} \tag{57a}$$

$$P_l \geq 0, \qquad l = 1, \ldots, L. \tag{57b}$$

It is straightforward to show using the KKT conditions that the optimum solution of (57) for given $\boldsymbol{q}$ is

$$P_l'(\boldsymbol{q}) = \left(\frac{1}{\lambda_{\mathrm{W}}(\boldsymbol{q})} - \frac{1}{\omega_l} \right)^+ \qquad (58)$$

where $1/\lambda_{\mathrm{W}}(\boldsymbol{q})$ is the water-filling level which satisfies

$$\overline{p} = \sum_{l=1}^{L} F_l P_l'(\boldsymbol{q}). \qquad (59)$$

We have shown that an optimum policy $(\boldsymbol{r}^*, \boldsymbol{q}^*)$ must have $|\mathcal{R}| = L$. Hence, (56) implies that $P_l^* = P_l'(\boldsymbol{q}^*) > 0$ for $l = 1, \ldots, L$. That is, for optimum policies, the operator $(\cdot)^+$ in (58) should not have any impact. This in turn enforces the implicit requirement on $\boldsymbol{q}^*$ in the form of $\lambda_{\mathrm{W}}(\boldsymbol{q}^*) < \omega_l^*$ for $l = 1, \ldots, L$, for policies of interest. Consequently, (59) implies

$$\frac{1}{\lambda_{\mathrm{W}}(\boldsymbol{q})} = \frac{\overline{p} + \int_{q_1}^{\infty} \frac{N_0}{s} f(s) \, ds}{1 - F(q_1)}. \qquad (60)$$

and (56) implies that the optimum rates corresponding to $\boldsymbol{q}$ are given by

$$r_l' = \log \left(\frac{\omega_l}{\lambda_{\mathrm{W}}(\boldsymbol{q})} \right). \qquad (61)$$

Since ω_l characterizes the channel states of interval $\mathcal{Q}_l$, the water-filling result is analogous to those in both the original continuous adaptive transmission problem [1] and the parallel Gaussian channel problem [24].

Note that the water-filling power allocation (58) is optimum within the set of policies $(\boldsymbol{r}, \boldsymbol{q})$ that explicitly require $p(s)$ in the form of (41). For a poorly chosen partition $\boldsymbol{q}$, $(\boldsymbol{r}, \boldsymbol{q})$ policies with $p(s)$ given by (41) may not be even locally optimum within the interval $\mathcal{Q}_l$. Therefore, the water-filling power allocation (58) must be used with some caution.

Moreover, because $R_L(\boldsymbol{r}, \boldsymbol{q})$ can have multiple local maxima, hill-climbing techniques based on alternating optimization of the partition $\boldsymbol{q}$ and rates $\boldsymbol{r}$ can at best only reach a local maxima. In the following, we show how the necessary conditions for optimality and q_1^*, the first element of $\boldsymbol{q}^*$ uniquely specify an optimum policy. Thus, the search for optimum policies is reduced to a line search over q_1.

B. Necessary Conditions for Optimum Policies

The corresponding Lagrangian function for (50) is

$$L(\boldsymbol{r}, \boldsymbol{q}, \lambda) = R_L(\boldsymbol{r}, \boldsymbol{q}) + \lambda[\overline{p} - \rho_L(\boldsymbol{r}, \boldsymbol{q})]. \qquad (62)$$

The Lagrangian function has a positive multiplier λ. The unit of the Lagrangian function is the same as ART.

Given fixed $\boldsymbol{r}$ and variable $\boldsymbol{q}$, (50) becomes the partition optimization problem studied in Section III. Formally, the partial derivative of the Lagrangian function with respect to q_l is

$$\frac{\partial L(\boldsymbol{r}, \boldsymbol{q}, \lambda)}{\partial q_l} = f(q_l) \left\{ -(r_l - r_{l-1}) + \lambda (e^{r_l} - e^{r_{l-1}}) \frac{N_0}{q_l} \right\}. \qquad (63)$$

Note that it is here where $F(s)$ is required to be continuous and differentiable. However, this is not required for the MPEQ in Section III. The necessary conditions for optimality

$$\left. \frac{\partial L(\boldsymbol{r}, \boldsymbol{q}, \lambda)}{\partial q_l} \right|_{\boldsymbol{r} = \boldsymbol{r}^*, \boldsymbol{q} = \boldsymbol{q}^*, \lambda = \lambda^*} = 0, \qquad l = 1, \ldots, L \qquad (64)$$

imply

$$\lambda^* = \frac{r_l^* - r_{l-1}^*}{\frac{N_0}{q_l^*} \left(e^{r_l^*} - e^{r_{l-1}^*} \right)} = \lambda_{\mathrm{M}}(\boldsymbol{r}^*) \qquad (65)$$

where the last equality is from (36). Therefore, the optimum Lagrange multiplier λ^* is equal to the incremental efficiency at q_l^* defined in (32).

On the other hand, a similar optimization problem can be formulated from (50) for given $\boldsymbol{q}$ and variable $\boldsymbol{r}$. This problem is the power allocation problem studied in Section IV-A. The necessary conditions for the optimum policies are

$$\left. \frac{\partial L(\boldsymbol{r}, \boldsymbol{q}, \lambda)}{\partial r_l} \right|_{\boldsymbol{r} = \boldsymbol{r}^*, \boldsymbol{q} = \boldsymbol{q}^*, \lambda = \lambda^*}$$
$$= F_l \left\{ 1 - \lambda e^{r_l} \frac{1}{\omega_l} \right\} \bigg|_{\boldsymbol{r} = \boldsymbol{r}^*, \boldsymbol{q} = \boldsymbol{q}^*, \lambda = \lambda^*} = 0, \quad l = 1, \ldots, L \qquad (66)$$

implying

$$\lambda^* = e^{-r_l^*} \omega_l^* = \lambda_{\mathrm{w}}(\boldsymbol{q}^*) \qquad (67)$$

where the last equality is due to (61). Therefore, the optimum Lagrange multiplier λ^* is the reciprocal of the water level of the water-filling power allocation obtained in Section IV-A.

Both (65) and (67) imply that for an optimum policy $(\boldsymbol{r}^*, \boldsymbol{q}^*)$

$$\lambda_{\mathrm{M}}(\boldsymbol{r}^*) = \lambda_{\mathrm{W}}(\boldsymbol{q}^*). \qquad (68)$$

For convenience, we define $\omega_0^* = \lambda_{\mathrm{W}}(\boldsymbol{q}^*)$ obtained from (58). Thus, from (65), (67), and (68), we obtain a set of necessary conditions for optimum policies as

$$q_l^* = \lambda_{\mathrm{W}}(\boldsymbol{q}^*) \left[\frac{\left(e^{r_l^*} - e^{r_{l-1}^*} \right)}{r_l^* - r_{l-1}^*} \right] \qquad (69)$$

$$= \frac{\omega_l^* - \omega_{l-1}^*}{\log \left(\omega_l^* \right) - \log \left(\omega_{l-1}^* \right)}, \qquad l = 1, 2, \ldots, L \qquad (70)$$

by substituting (61).

C. Search for Optimum Policies

Since ω_l^* and $\lambda_{\mathrm{W}}(\boldsymbol{q}^*)$ depend on $\boldsymbol{q}^*$, (70) provides L equations for the L unknowns q_l^*. The only inconvenience is that (70) are nonlinear and q_l^* are limits of integrals defining ω_l^* and $\lambda_{\mathrm{W}}(\boldsymbol{q}^*)$. Therefore, it is hard to solve q_l^* by using any direct substitution. However, in the following, based on the monotonic relationship between q_l^* and ω_l^*, a simple algorithm can be found to search for the optimum solution.

Naturally, the average channel quality ω_l becomes better when $\mathcal{Q}_l$ is enlarged by adding more good channel states. Such an intuition brings two monotonic relations below.

Lemma 2: q_l^* is strictly increasing in ω_l^*.

Lemma 3: ω_{l-1} is strictly increasing in q_l.

1) $l = 1$:
   a) Given q_1^*, use (60) to uniquely solve for $\lambda_W(\mathbf{q}^*)$ and, thus, obtain $w_0^* = \lambda_W(\mathbf{q}^*)$;
   b) Given q_1^* and w_0^*, use (70) to uniquely solve for w_1^* [Lemma 2].

2) $l > 1$:
   a) Given q_{l-1}^* and ω_{l-1}^*, use (71) to uniquely solve for q_l^* [Lemma 3];
   b) Given q_l^* and ω_{l-1}^*, use (70) to uniquely solve for ω_l^* [Lemma 2].

3) Repeat step 2) from $l = 2$ to $l = L$, we will have $\mathbf{q}^*$.

Fig. 4.   An algorithm searching for optimum policies started from q_1^*.

In particular, Lemma 3 follows from the following restatement of (48):

$$\frac{1}{\omega_{l-1}} = \int_{q_{l-1}}^{q_l} \frac{N_0}{s} f_l(s)\, ds, \qquad l = 2, \ldots, L. \tag{71}$$

These two lemmas are the building blocks of a constructive procedure for finding the optimum policy given q_1^* described in Fig. 4. Particularly, in steps 1b and 2b, Lemma 2 implies the existence of a unique solution ω_l^* and in the step 2a, Lemma 3 implies the existence of the unique solutions q_l^*.

Theorem 3: For each value of q_1^*, there exists a unique policy $(\mathbf{r}^*, \mathbf{q}^*)$ that satisfies the necessary conditions (70).

Overall, (48) and (70) offer $2L$ equalities. With q_1^*, we repeatedly use (48) and (70) to uniquely determine other q_l^* and all ω_l^*. Indeed, at the step $l = L$, given q_L^*, ω_L^* can be determined by either (70) or by (48) with $q_{L+1} = \infty$. These two approaches are distinct and must agree for the optimum policies. In this sense, the $2L$ equalities (48) and (70) implicitly restrict q_1^*. However, since (70) are only necessary conditions obtained from (68), it implies that all local optimum satisfy all $2L$ equalities, i.e., (48) and (70). Therefore, the optimum q_1^* will not be unique when multiple local maxima achieve the same objective value. In this case, there will be multiple optimum policies $(\mathbf{r}^*, \mathbf{q}^*)$.

Since there is only one undetermined parameter q_1^*, a line search over q_1 solves the problem of maximizing ART subject to a power constraint and a finite code rate set. However, it is not yet clear whether the search is well behaved for general fading distributions.

Finally, let the maximum ART for any L-level policy be

$$\bar{C}_L = R_L(\mathbf{r}^*, \mathbf{q}^*) \tag{72}$$

where $(\mathbf{r}^*, \mathbf{q}^*)$ is any L-level optimum policy.

Theorem 4: For all $L > 1$

$$\bar{C}_L > \bar{C}_{L-1}. \tag{73}$$

V. NUMERICAL RESULTS

In this section, following the approach in [12], we let $N_0 = 1$ and evaluate the system performance numerically.

A. Optimum Code Rates and Partition

Theorem 3 suggests that it is possible to perform a line-search on the single parameter q_1 to find $(\mathbf{r}^*, \mathbf{q}^*)$ and obtain $\bar{C}_L$. In Fig. 5, we present a comparison between $\bar{C}_L$ and several known

throughputs for a Rayleigh-fading channel. C is the ergodic capacity given in [1]. C_L, proposed in [2], is the maximum ART corresponding to a discrete adaptive transmission policy with L code rates and L power levels. C_2 requires about 3 dB less average transmitted power than C_1 to achieve the same throughput.

Since C_L is obtained through exhaustive search, the required computation complexity increases exponentially with respect to L and it is not desirable to evaluate C_L for large L. In [2], by employing a greedy iterative algorithm initialized with an asymptotically optimum solution of C_L, a good lower bound $R_L^\ddagger$ of C_L is found. In Fig. 5, $R_{10}^\ddagger$ is about 1 dB away from C.

We observed that $\bar{C}_2 > R_{10}^\ddagger$, which suggests employing adaptive systems with a small number of code rates and a large number of power levels. Finally, it can be shown that $\bar{C}_4$ is only a fraction of a decibel away from C for throughput values less than 8 bits/s/Hz.

Therefore, an important message embedded in these results is that for a practical power-limited adaptive system design, a combination of a few optimized coding rates and variable transmitted power buys almost all the available ergodic capacity. Moreover, it is shown in [8] that the claim still holds for scenarios when the precise channel state information is not available.

B. Optimum Partition for Preset Code Rates

Here, we have a comparison between the performance of the optimum partition and that of a partition motivated by the design in [6]. Given a code rate set $\mathbf{r}$ with $r_0 = 0$, a suboptimal partition is

$$q_l = \begin{cases} 0, & l = 0 \\ q e^{r_l}, & l = 1, 2, \ldots, L - 1 \end{cases} \tag{74}$$

where q is a parameter tuned to satisfy the power constraint (21a) since $p(s) = \psi(s, r_l)$, $s \in \mathcal{Q}_l$.

Fig. 6 shows the results of two partition methods for

$$r_l = \begin{cases} 0, & l = 0 \\ 1, & l = 1 \\ 2(l-1), & l = 2, 3, \ldots, L - 1. \end{cases} \tag{75}$$

Using the optimum partition can reduce the transmitted power requirement by as much as 0.5 dB in comparison with using the other suboptimal partition for $L = 5$. However, for a small L, there is only a negligible difference between the results corresponding to the optimum partition and the suboptimal partition.
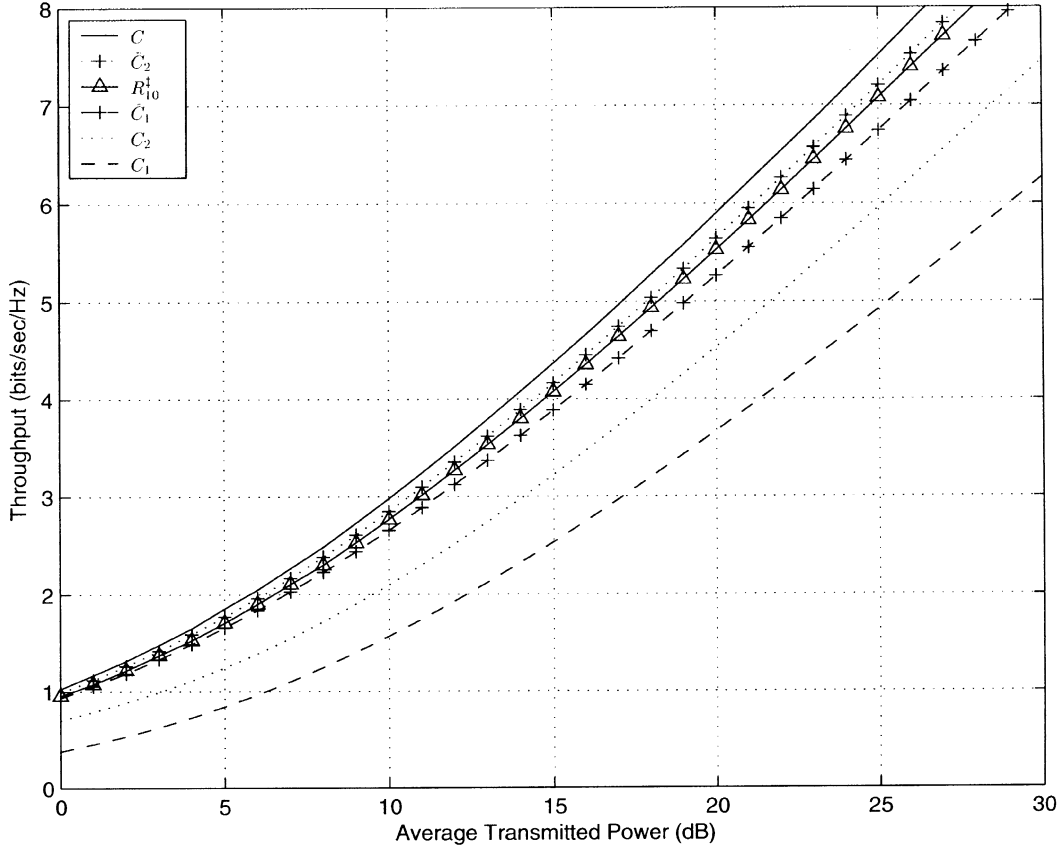
Fig. 5. A comparison of C, C_L, $R_L^\ddagger$, and $\bar{C}_L$ for Rayleigh fading.

C. Spectral Efficiency

In Fig. 7, we show the results of spectral efficiency of adaptive M-QAM systems with a $P_b = 10^{-6}$. The results show that for three regions ($L = 3$), there is a negligible difference between the spectral efficiency corresponding to the optimum partition and the suboptimal partition [6]. The difference becomes distinguishable when there are more regions (larger L). Overall, the suboptimal solution of Goldsmith and Chua [6] for variable M-QAM is indeed very close to the optimum.

VI. CONCLUSION

In this paper, the average reliable throughput maximization problem for adaptive transmission systems with a finite number of code rates and continuously varying power level is formulated. By exploring the properties of the optimum policies, we can obtain $\bar{C}_L$ through a simple line-search algorithm.

Moreover, while studying the properties of the optimum policies, we discovered the optimum partition given an increasing rate assignment. This is particularly useful for designing adaptive systems with the channel codes selected from a limited set of good codes. The approach used in solving our maximization problem can be applied to solve a spectrum efficiency maximization problem formulated in [6].

APPENDIX
PROOFS

A. Proof: Lemma 1

Claim a) is straightforward since, regardless of how much the transmitted power is, we cannot have nonzero rate for reliable communication when $s = 0$. Combining (12), (25), (24), and (32), we have

$$\eta_l(s) = \frac{s}{N_0} \frac{r_l - r_{l-1}}{e^{r_l} - e^{r_{l-1}}}. \tag{76}$$

Hence, claim b) follows. For claim c), we notice that e^{r_l} is strictly convex in r_l. For any strictly convex function $g(x)$, we have for $x_2 > x_1$

$$\left. \frac{\partial g(x)}{\partial x} \right|_{x=x_1} < \frac{g(x_2) - g(x_1)}{x_2 - x_1} < \left. \frac{\partial g(x)}{\partial x} \right|_{x=x_2}. \tag{77}$$

Consequently, for any $x_1 < x_2 < x_3$

$$\frac{g(x_2) - g(x_1)}{x_2 - x_1} < \left. \frac{\partial g(x)}{\partial x} \right|_{x=x_2} < \frac{g(x_3) - g(x_2)}{x_3 - x_2}. \tag{78}$$

Therefore, with $g(x) = e^x$, it follows from (76) that

$$\frac{s}{N_0 \eta_l(s)} = \frac{e^{r_l} - e^{r_{l-1}}}{r_l - r_{l-1}} < \frac{e^{r_{l+1}} - e^{r_l}}{r_{l+1} - r_l} = \frac{s}{N_0 \eta_{l+1}(s)}. \tag{79}$$

Since $\eta_l(s)$ is positive, we have (c). $\square$

B. Proof of Theorem 1

Given an arbitrary power allocation $I_l(s)$ and the MPEQ allocation $I_l^*(s)$, we observe that

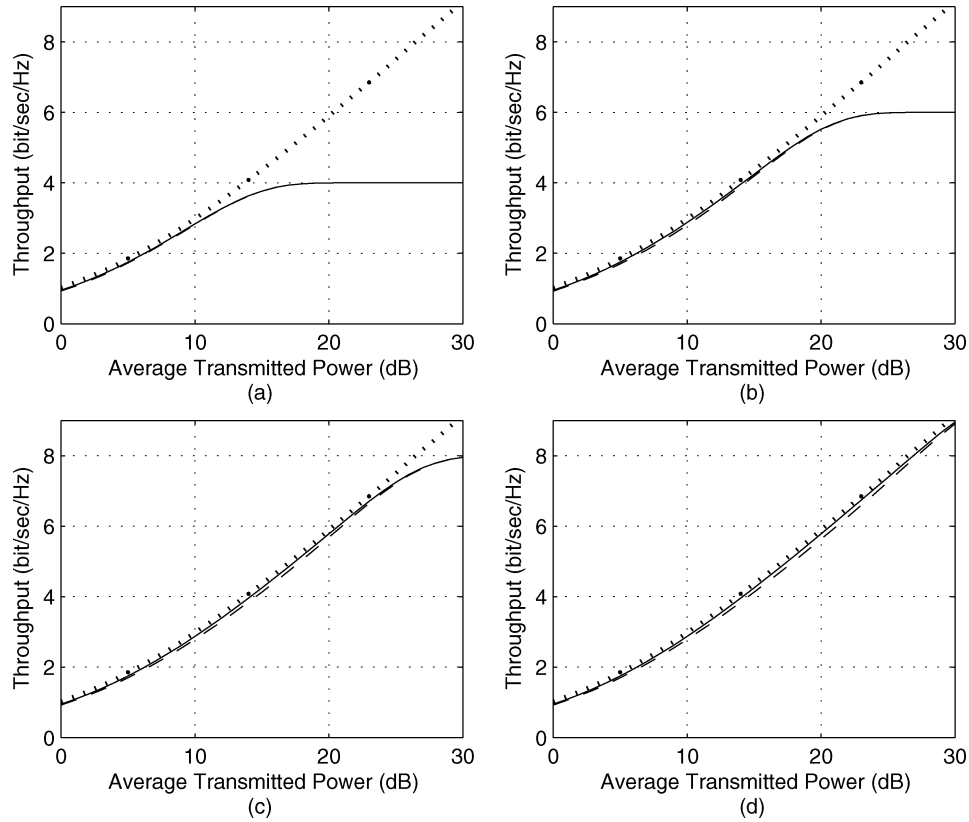$$I_l^*(s) = I_l^*(s) I_l(s) + I_l^*(s) [1 - I_l(s)]. \tag{80}$$

Fig. 6. A comparison of partitions with $r_0 = 0$ and $r_l = 2l$, $l = 1, 2, \ldots, L - 1$ for Rayleigh fading. (a) $L = 3$, (b) $L = 4$, (c) $L = 5$, and (d) $L = 6$. The dotted line indicates the ergodic capacity C, the solid line corresponds to the optimum partition, and the dash line corresponds to the suboptimal partition in [6].
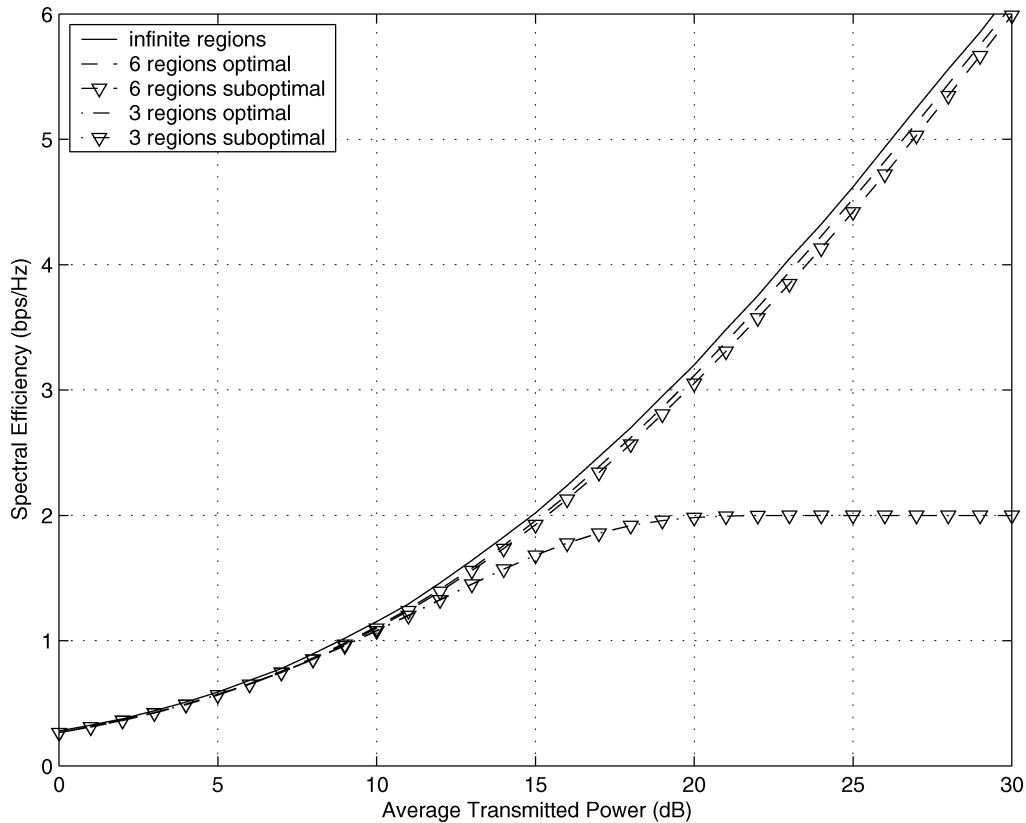

Fig. 7. Comparison of spectral efficiency obtained by the optimum partition and the suboptimal partition in [6] for $P_b = 10^{-6}$.

For $\rho_L(\mathcal{I})$ given by (30)

$$\rho_L(\mathcal{I}^*) = \rho_L(\{I_l^*(s)\})$$
$$= \rho_L(\{I_l^*(s)I_l(s)\}) + \rho_L(\{I_l^*(s)[1-I_l(s)]\}). \quad (81)$$

It follows from (33) and (80) that the MPEQ policy achieves ART

$$R_L(\mathcal{I}^*) = \sum_{l=1}^{L} \int_0^\infty I_l^*(s)\eta_l(s)p_l^\Delta(s)f(s)\,ds \quad (82)$$

$$= \kappa + \sum_{l=1}^{L} \int_0^\infty I_l^*(s)[1-I_l(s)]\,\eta_l(s)p_l^\Delta(s)f(s)\,ds \quad (83)$$

where

$$\kappa = \sum_{l=1}^{L} \int_0^\infty I_l^*(s)I_l(s)\eta_l(s)p_l^\Delta(s)f(s)\,ds. \quad (84)$$

Consequently

$$R_L(\mathcal{I}^*) \geq \kappa + \sum_{l=1}^{L} \int_0^\infty I_l^*(s)[1-I_l(s)]\lambda_M(\boldsymbol{r})p_l^\Delta(s)f(s)\,ds \quad (85)$$

$$= \kappa + \lambda_M(\boldsymbol{r})\left(\rho_L(\mathcal{I}^*) - \rho_L(\{I_l^*(s)I_l(s)\})\right) \quad (86)$$

$$\geq \kappa + \lambda_M(\boldsymbol{r})\left(\rho_L(\mathcal{I}) - \rho_L(\{I_l(s)I_l^*(s)\})\right) \quad (87)$$

where (85) holds since $\eta_l(s) \geq \lambda_M(\boldsymbol{r})$ for channel states s with $I_l^*(s) = 1$; (86) is due to (30) and (81); and (87) is true because $\bar{p} = \rho_L(\mathcal{I}^*) \geq \rho_L(\mathcal{I})$. We then observe that

$$R_L(\mathcal{I}^*) \geq \kappa + \sum_{l=1}^{L} \int_0^\infty I_l(s)[1-I_l^*(s)]\lambda_M(\boldsymbol{r})p_l^\Delta(s)f(s)\,ds \quad (88)$$

$$\geq \kappa + \sum_{l=1}^{L} \int_0^\infty I_l(s)[1-I_l^*(s)]\eta_l(s)p_l^\Delta(s)f(s)\,ds \quad (89)$$

$$= \sum_{l=1}^{L} \int_0^\infty I_l(s)\eta_l(s)p_l^\Delta(s)f(s)\,ds \quad (90)$$

$$= R_L(\mathcal{I}) \quad (91)$$

since applying (81) on $\mathcal{I}$ instead of $\mathcal{I}^*$ will confirm that the right-hand side of (88) equals the right-hand side of (87); (89) is valid because $\eta_l(s) < \lambda_M(\boldsymbol{r})$ for channel states s with $[1 - I_l^*(s)] = 1$; and (90) is found by substituting κ from (84) and applying the relation of (80) on $\mathcal{I}$ again. $\quad\square$

C. Proof of Proposition 1

The claim a) is a direct consequence of Definition 1. Specifically, Definition 1 implies that $I_l^{*\prime}(s) = 1$ wherever $I_l^*(s) = 1$ if $\lambda_M(\boldsymbol{r}) > \lambda_M'(\boldsymbol{r})$. Therefore,

$$I_l^*(s) = I_l^{*\prime}(s)I_l^*(s). \quad (92)$$

Consequently

$$\rho_L(\mathcal{I}^*) = \sum_{l=1}^{L} \int_0^\infty I_l^*(s)p_l^\Delta(s)f(s)\,ds \quad (93)$$

$$= \sum_{l=1}^{L} \int_0^\infty I_l^{*\prime}(s)I_l^*(s)p_l^\Delta(s)f(s)\,ds \quad (94)$$

$$\leq \sum_{l=1}^{L} \int_0^\infty I_l^{*\prime}(s)I_l^*(s)p_l^\Delta(s)f(s)\,ds$$
$$+ \sum_{l=1}^{L} \int_0^\infty I_l^{*\prime}(s)(1 - I_l^*(s))p_l^\Delta(s)f(s)\,ds \quad (95)$$

$$= \sum_{l=1}^{L} \int_0^\infty I_l^{*\prime}(s)p_l^\Delta(s)f(s)\,ds \quad (96)$$

$$= \rho_L(\mathcal{I}^{*\prime}) \quad (97)$$

where (95) is due to the fact that the second term on the right-hand side is nonnegative. Hence, the claim b) holds.

Note that (36) implies

$$q_l = \frac{\lambda_M(\boldsymbol{r})N_0\left(e^{r_l} - e^{r_{l-1}}\right)}{r_l - r_{l-1}} \quad (98)$$

$$> \frac{\lambda_M'(\boldsymbol{r})N_0\left(e^{r_l} - e^{r_{l-1}}\right)}{r_l - r_{l-1}} \quad (99)$$

$$= q_l'. \quad (100)$$

Thus, claim c) holds.

For $\lambda_M(\boldsymbol{r}) > \lambda_M'(\boldsymbol{r})$, if $\rho_L(\mathcal{I}^*) = \rho_L(\mathcal{I}^{*\prime}) = \bar{p}$, according to (95), we have

$$\sum_{l=1}^{L} \int_0^\infty I_l^{*\prime}(s)(1 - I_l^*(s))p_l^\Delta(s)f(s)\,ds = 0. \quad (101)$$

Since $p_l^\Delta(s) > 0$, we have

$$\sum_{l=1}^{L} \int_0^\infty I_l^{*\prime}(s)(1 - I_l^*(s))f(s)\,ds = 0. \quad (102)$$

For $s \in [q_l', q_l)$, $I_l^{*\prime}(s) = 1$ and $I_l^*(s) = 0$. Therefore, (102) implies

$$f(s) = 0, \qquad s \in [q_l', q_l), \ l = 1, \ldots, L. \quad (103)$$

Consequently, claim d) also holds. $\quad\square$

D. Proof of Lemma 2

Note that $\frac{x}{e^x - 1}$ decreases in x for $x \geq 0$.

Defining $x = [\log(\omega_l) - \log(\omega_{l-1})]$, (70) can be written as

$$q_l = \frac{e^x - 1}{x}\omega_{l-1}. \quad (104)$$

Since x strictly increases in ω_l, the lemma follows. $\quad\square$

E. Proof of Lemma 3

From (48), let $g(q_l) = \omega_{l-1} = \frac{1}{\int_{\mathcal{Q}_{l-1}} f_{l-1}(s)/s\,ds}$.

$$\frac{dg(q_l)}{dq_l} = \frac{f(q_l)}{\left[\int_{\mathcal{Q}_{l-1}} f(s)/s\,ds\right]^2}\left[\int_{\mathcal{Q}_{l-1}}\left(\frac{1}{s} - \frac{1}{q_l}\right)f(s)\,ds\right] > 0 \quad (105)$$

since $F(s)$ strictly increases in s. Therefore, ω_{l-1} strictly increases in q_l. $\quad\square$

F. Proof of Theorem 4

An optimum $(\boldsymbol{r}^*, \boldsymbol{q}^*)$ policy with $L - 1$ distinct nonzero q_l^* can be expanded to an $(\boldsymbol{r}, \boldsymbol{q})$ policy with L distinct nonzero q_l

by splitting any of its L intervals (including zero-power interval $[0, q_1^*)$) into two intervals.

If the first interval $[0, q_1^*)$ is split, without changing $p(s)$, the new $(\boldsymbol{r}, \boldsymbol{q})$ policy will have the same ART as the original.

On the other hand, we can split any intervals of the original $(\boldsymbol{r}^*, \boldsymbol{q}^*)$ policy with nonzero power and re-allocate power according to (58). The newly split intervals will increase their contribution to ART while the contribution from the other intervals stays the same. Therefore, the new $(\boldsymbol{r}, \boldsymbol{q})$ policy has a higher ART than that one corresponding to the original $(\boldsymbol{r}^*, \boldsymbol{q}^*)$ policy. We have

$$\bar{C}_L \geq R_L(\boldsymbol{r}, \boldsymbol{q}) > R_{L-1}(\boldsymbol{r}^*, \boldsymbol{q}^*) = \bar{C}_{L-1}. \tag{106}$$

□

REFERENCES

[1] A. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.

[2] L. Lin, R. Yates, and P. Spasojević, "Discrete adaptive transmission for fading channels," *IEEE Trans. Commun.*, vol. 51, no. 12, pp. 2115–2125, Dec. 2003.

[3] *EIA/TIA-95: Mobile Station-Base Station Compatibility Standard for Dual-Mode Wideband Spectrum Cellular System*, 1989.

[4] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushyana, and S. Viterbi, "CDMA/HDR: A bandwidth efficient high speed wireless data service for nomadic users," *IEEE Commun. Mag.*, vol. 38, no. 7, pp. 70–77, Jul. 2000.

[5] K. Balachandran, R. Ejzak, S. Nanda, S. Vitebskiy, and S. Seth, "GPRS-136: High-rate packet data service for North American TDMA digital cellular systems," *IEEE Pers. Commun.*, vol. 6, no. 3, pp. 34–47, Jun. 1999.

[6] A. Goldsmith and S. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, no. 10, pp. 1218–1229, Oct. 1997.

[7] C. Köse and D. Goeckel, "On power adaptation in adaptive signaling systems," *IEEE Trans. Commun.*, vol. 48, no. 11, pp. 1769–1773, Nov. 2000.

[8] L. Lin, R. Yates, and P. Spasojević, "Adaptive transmission with discrete code rates and channel state uncertainty," in *Proc. IEEE GLOBECOM*, vol. 3, San Francisco, CA, Dec. 2003, pp. 1771–1775.

[9] L. Ozarow, S. Shamai (Shitz), and A. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, no. 3, pp. 359–378, May 1994.

[10] S. Hanly and D. N. C. Tse, "Multi-access fading channels: Part II: Delay-limited capacities," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2816–2831, Nov. 1998.

[11] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[12] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1468–1489, Jul. 1999.

[13] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.

[14] R. Knopp and P. Humblet, "On coding for block fading channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 1, pp. 189–205, Jan. 2000.

[15] E. Malkamäki and H. Leib, "Coded diversity on block-fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 771–781, Mar. 1999.

[16] D. Goeckel, "Adaptive coding for time-varying channels using outdated fading estimates," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 844–855, Jun. 1999.

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

[18] J. Campello, "Optimal discrete bit loading for multicarrier modulation systems," in *Proc. IEEE Int. Symp. Information Theory*, Cambridge, MA, Aug. 1998, p. 193.

[19] B. Korte and J. Vygen, *Combinatorial Optimization: Theory and Algorithms*, 2nd ed. New York: Springer-Verlag , 2002.

[20] T. Cormen, C. Leiserson, R. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed. Cambridge, MA: MIT Press, 2001.

[21] E. Balas and E. Zemel, "An algorithm for large zero-one knapsack problem," *Oper. Res.*, vol. 28, pp. 1130–1154, 1980.

[22] D. Pisinger, "Budgeting with bounded multiple-choice constraints," *Europ. J. Oper. Res.*, vol. 129, pp. 471–480, 2001.

[23] J. Campello and J. Cioffi, "Optimal discrete loading," ANSI subcommittee T1E.4 contribution 1998-341, Dallas, TX, Nov. 1998.

[24] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.