Outage minimization for parallel fading channels with limited feedback

YuanYuan He and Subhrakanti Dey[*]

Abstract

We address an optimal power allocation problem for minimizing the outage probability for M parallel block-Nakagami-fading channels under a long-term average sum transmit power constraint with finite rate feedback of channel state information (CSI). A simulation-based optimization technique called simultaneous perturbation stochastic approximation algorithm (SPSA) is employed first to numerically derive a locally optimal power codebook. Due to the high computational complexity and long convergence time of SPSA, we make an ordering assumption on the power codebook entries and derive effective hyperplane based approximations to the channel quantization regions and present a number of low-complexity suboptimal quantized power codebook design algorithms. Unlike existing work on outage minimization for multiple-input multiple-output (MIMO) channels with limited feedback, we do not assume that identical transmission power is used for all channels within each channel quantization region. We also do not resort to a Gaussian approximation for the instantaneous mutual information in general as used in many existing work. Based on our power ordering assumption and hyperplane based approximations, we show that allocating identical power to all channels within a given channel quantization region in the limited feedback scenario is asymptotically optimal only at high average power (or average signal-to-noise ratio (SNR)) for the Rayleigh fading case, whereas for the general Nakagami case, the transmit power allocation for an individual channel within each quantized region is asymptotically proportional to the corresponding Nakagami fading parameter (severity of fading). We also present a novel diversity order result for the outage probability for the Nakagami fading case. Finally, we derive a suitable Gaussian approximation based low-complexity power allocation scheme for a large number of parallel channels, which has important applications in wideband slow-fading orthogonal frequency-division multiplexing (OFDM) systems. Extensive numerical results illustrate that only a few bits of feedback close the gap substantially in outage performance between the limited feedback case and the full instantaneous CSI at the transmitter case.

Introduction

Determining the information theoretic capacity of block-fading wireless channels has been an important area of research over the past decade. Various notions of capacity for single-user fading channels include *ergodic capacity* [1], *delay-limited capacity* [2] and *capacity versus outage probability* [3]. For delay-sensitive traffic such as voice and video, the latter two notions are rather important. In particular, the notion of outage probability signifies the probability that the capacity of a wireless channel falls below a required rate threshold. In [3], optimal power allocation for outage minimization in the case of parallel fading channels (single user) was obtained with the

assumption of full channel state information (CSI) at the transmitter. However, full CSI at the transmitter is hard to obtain in practice due to limited bandwidth in the feedback channel from the receiver to the transmitter, and it is more common to have full CSI at the receiver. This has motivated researchers over the last decade to analyze performances of wireless systems with various forms of partial CSI at the transmitter (CSIT), such as noisy CSIT, statistical CSIT and quantized CSIT. In particular, the idea of Grassmanian line packing was used to design optimal beamforming codebooks for MIMO systems in [4], whereas in a related study [5], the authors derived a lower bound on the outage performance of a multiple-antenna systems using beamforming based on quantized CSIT. More recently, in [6], maximization of expected rate over a single-input single-output slowly fading channel is investigated using optimized discrete rate

*Correspondence: s.dey@ee.unimelb.edu.au
Department of Electrical and Electronic Engineering, University of Melbourne, Vic. 3010, Melbourne, Australia

and power control with quantized CSIT. A general framework for power allocation in Gaussian vector channels with l_p norm constraints on the eigenvalues of the MIMO channel matrix was investigated in [7]. The authors of [6] have also investigate the diversity-multiplexing tradeoff in MIMO channels with quantized CSIT in [8] (see also [9]). A number of recent articles have investigated outage minimization for fading channels with limited feedback for MIMO or multi-antenna systems. Such studies include [9-13]. In particular, [12] looks at outage minimization with a finite-rate power codebook for MIMO systems. The key finding of this article (see also [8]) is that the optimal power codebook has a circular structure in that the same transmit power is allocated to the outage region and the best channel region. In order to design the optimal power codebook, it assumes however that the same transmit power (as a function of the entire channel matrix) is used in all transmit antennas. This allows the authors of [12] to reduce the finite-rate power codebook design problem to an equivalent scalar quantization problem. Even then, finding the cumulative density function for the equivalent scalar random variable requires computing multi-dimensional probability integrals which is computationally complex. Furthermore, the optimal power codebook entries are found via generic gradient search techniques which can take unreasonably long time to converge. Using a similar setting, the same authors have investigated the outage diversity behavior for multiple-antenna systems with quantized CSIT in [13] (see also [10]). In [11], the problem of outage minimization using quantized CSIT is investigated for the fading relay channel and [14] also studied the outage minimization problem for cooperative amplify-and-forward systems. In [9], a Gaussian approximation is used to capture the probability distribution of the mutual information for a MIMO system in order to study the outage behavior. Finally, many of the above results only apply to Rayleigh fading channels (where the MIMO channel matrix is assumed to have complex circularly symmetric Gaussian distributed entries). Note however that the circular nature of the optimal power codebook and some of the useful approximations developed in [10] for asymptotically large number of channel feedback bits are also relevant for our study and we duly acknowledge this fact. Our focus is however on designing practical low-complexity but sub-optimal algorithms for designing the quantized power codebook and derive theoretical properties of these power allocation schemes in order to justify the various approximations used in designing the sub-optimal schemes.

In our article, we look at an M-parallel fading channels system as introduced in [15], where one codeword spans M subchannels in one fading block and each block undergoes the same CSI, and we aim to minimize the outage probability under a sum (across all channels) long term average power constraint with quantized CSIT. Technologically, parallel fading channels constitute a useful and fundamental communication framework for various applications, for example, multiple-antenna systems after singular value decomposition or an OFDM system with frequency-selective fading [15]. Due to the unavailability of full CSIT in our framework, our model is better suited to the case of multicarrier OFDM systems, with M parallel subchannels located at nonadjacent carrier frequencies. The concept of parallel channels also extends to multiple transmission time-slots [16] and diversity available through cooperative communications such as multiple relays etc. Our results in this article are applicable to all these scenarios.

Our main contributions can be summarized as follows:

- We first formulate the above-described optimization problem and provide an simulation based iterative optimization algorithm: simultaneous perturbation stochastic approximation algorithm (SPSA), to numerically solve for the joint optimization of locally optimal channel partitions and quantized power allocation.

- Based on a power ordering assumption and a hyperplane based approximation to the basic rate achieving mutual information curve in the vector channel space, we derive a number of low-complexity suboptimal finite-rate power codebook design algorithms for outage minimization with quantized channel information—without assuming identical transmission power per channel or using a Gaussian approximation for the instantaneous mutual information in general.

- We show that in the high average power (or average SNR) regime, it is asymptotically optimal to allocate transmit power proportional to the Nakagami m fading parameter in the individual channels within each quantized region. In the Rayleigh fading case, this corresponds to allocating the same power across all channels within each quantized region (but only in the high average power regime).

- We also derive a novel diversity order result for the outage probability in the Nakagami fading case.

- Finally, we investigate the suitability of a Gaussian approximation scheme for the instantaneous mutual information in the case of a large number of independent (but not identically distributed) parallel channels, which is applicable to a slow fading broadband frequency selective channel or to a flat fast fading channel [17,18]. Note that as we will show later, although the Gaussian approximation is seen to perform poorly for a small number of parallel channels, it performs efficiently for a large number of channels (e.g., $M \geq 16$), thus having important

practical applications to such broadband
multi-carrier systems.

The organization of the article is as follows. Section
'Channel model and outage minimization' presents the
fading channel model and the typical outage problem
based on full CSIT. Section 'Optimum quantized power
control with finite-rate feedback' presents the outage min-
imization problem with quantized CSIT followed by the
modified problem formulation using the power ordering
and hyperplane based approximation. Various subopti-
mal algorithms are then presented for finding the power
codebook in the high average power regime along with
their associated theoretical properties. A new result on
the diversity order for the outage probability is then pre-
sented for the Nakagami fading case using our power
allocation algorithm based on the power ordering and
hyperplane based approximation. Section 'Large number
of channels analysis' presents a Gaussian approximation
based sub-optimal algorithm applicable to the case of a
large number of independent parallel channels. Section
'Numerical results' presents an extensive set of numer-
ical results illustrating the efficiency of our algorithms
measured by closeness of their outage performance as
compared to the full CSIT based optimal power alloca-
tion solution. Finally, Section 'Conclusions' presents some
concluding remarks and ideas for future extensions of this
study.

Channel model and outage minimization

We consider an M-parallel flat-fading channel model,
where a transmitted codeword spans M subchannels in
one fading block. For each fading block, the received
signal of each subchannel $i, i \in \{1, 2, \ldots, M\}$, can be
represented as[a]:

$$y_i = \sqrt{h_i} x_i + w_i, \tag{1}$$

where h_i is the channel power gain and x_i is the chan-
nel input symbol. The noise sequences $w_1, \ldots, w_M$ are
independent and identically distributed (i.i.d) Gaussian
random variables with zero mean and unit variance. It
is assumed that the components of channel power gain
vector $\mathbf{h} = (h_1, \ldots, h_M)$ are mutually independent, indi-
vidually i.i.d across fading blocks and ergodic and fading
is sufficiently slow so that the input symbols transmitted
over the same fading block experience the same channel
state. It is also assumed that the fading block length $N \to
\infty$ so that information theoretic results can be applied.
The individual fading distributions may not be identi-
cal. However, they (and hence the joint channel fading
distribution) are assumed to be continuous.

Given a channel realization $\mathbf{h}$, and assuming the avail-
ability of full channel state information (CSI) at the
transmitter and receiver, denote the corresponding power

allocation to the M subchannels by the vector $\mathbf{p}(\mathbf{h}) =
(p_1(\mathbf{h}), \ldots, p_M(\mathbf{h}))$. Then the maximum mutual informa-
tion of an M-parallel channel is given by [19],

$$r(\mathbf{h}, \mathbf{p}(\mathbf{h})) = \frac{1}{M} \sum_{i=1}^{M} \frac{1}{2} \log(1 + h_i p_i(\mathbf{h})), \tag{2}$$

where, the rate unit is nats per real dimension. Note that in
(2), we consider that the capacity is averaged over parallel
channels as [15] did.

Thus, the outage probability, defined as the proba-
bility that the instantaneous mutual information of the
channel is less than a pre-specified transmission rate r_0
(nats/channel use), can be expressed as

$$P_{\text{out}}(\mathbf{h}, \mathbf{p}(\mathbf{h}), r_0) = \text{Prob}\left[r(\mathbf{h}, \mathbf{p}(\mathbf{h})) < r_0\right]. \tag{3}$$

Under a long term average power constraint defined by
$E[\langle \mathbf{p}(\mathbf{h}) \rangle] \le P_{\text{av}}$, (where $\langle x \rangle$ denotes the arithmetic mean
of the vector x with length M, namely, $\langle x \rangle = \frac{1}{M} \sum_{i=1}^{M} x_i$),
the outage minimization problem can be described as

$$\min_{\mathbf{p}(\mathbf{h}) \ge 0} \text{Prob}\left[\frac{1}{M} \sum_{i=1}^{M} \frac{1}{2} \log(1 + h_i p_i(\mathbf{h})) < r_0\right]$$

$$\text{s.t. } E[\langle \mathbf{p}(\mathbf{h}) \rangle] \le P_{\text{av}}. \tag{4}$$

The optimal power allocation with full CSI at the trans-
mitter for this problem can be found explicitly by using
convex optimization techniques and was presented in
Proposition 4 of [3]. The readers are referred to [3] for
further details. Note that here P_{av} can be thought of effec-
tively as the transmitter side signal-to-noise ratio (since
noise variance has been normalized to unity). In the
following we will address the optimal power allocation
problem for outage minimization where only partial or
limited CSI is available at the transmitter. For the pur-
pose of analysis, we will assume that each channel h_i
is gamma distributed (Nakagami fading) with mean $\frac{1}{\lambda_i}$,
which probability density function (pdf) is given by

$$f(h_i) = (m_i \lambda_i)^{m_i} \frac{h_i^{m_i-1}}{\Gamma(m_i)} e^{-m_i \lambda_i h_i}, \quad h_i > 0, \tag{5}$$

where $\Gamma(.)$ is gamma function ($\Gamma(s) = \int_0^\infty t^{s-1} e^{-t} dt$)
and constant $m_i \ge 0.5$. m_i is called the fading parame-
ter. Larger values of the fading parameter m_i imply less
severe fading environments. When $m_i = 1$, the above
distribution boils down to an exponential distribution
(corresponding to Rayleigh fading) and the non fading
case corresponds to $m_i = \infty$.

Optimum quantized power control with finite-rate feedback

It is well known that having perfect CSI at both transmit-
ter and receiver is hard to satisfy in a practical system due
to bandwidth constraints on the receiver to transmitter

feedback link as well as considerable communication cost overhead. In this section, we consider designing a power allocation procedure for M-parallel flat-fading channels based on quantized vector CSI $\mathbf{h} = (h_1, \ldots, h_M)$ (in M dimensions) acquired via a no-delay and error-free feedback link with limited rate from the receiver to the transmitter.

Optimal power allocation with limited feedback strategy

We assume that the receiver can perfectly estimate the full CSI information. Given B bits of feedback, a power codebook $\mathcal{P} = \{\mathbf{P}_1, \ldots, \mathbf{P}_L\}$, where $\mathbf{P}_j = \{p_{1j}, \ldots, p_{Mj}\}$, $j = 1, \ldots, L$ of cardinality $L = 2^B$, is designed off-line purely on the basis of the statistics of $\mathbf{h}$. Note that the power levels for different channels here are distinct as opposed to [9,12] where the same transmit power was allocated to all transmit antennas in the MIMO setting. This codebook is known *a priori* by both the transmitter and the receiver. Given a channel realization $\mathbf{h}$,

- First, the receiver applies a deterministic mapping denoted as I from current instantaneous $\mathbf{h}$ information into one of L integer indices [9], where the mapping I partitions the entire M-dimensional space of $\mathbf{h}$ into L regions $\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_L$, given as $I(\mathbf{h}) = j$, if $\mathbf{h} \in \mathcal{R}_j$, $j = 1, \ldots, L$.
- Second, the receiver sends the corresponding index $j = I(\mathbf{h})$ to the transmitter via the feedback link.
- Then, the jth entry of the power codebook $\mathcal{P}$, i.e., $\mathbf{P}_j$, will be employed by the transmitter for transmission.

Therefore the key steps involved in the limited feedback design problem constitute obtaining (off-line) the jointly optimal CSI partitions and power codebook design. Our objective is to design efficient algorithms for solving this joint optimization problem of the channel partition regions and the power codebook, so as to minimize the outage probability while satisfying a long term average power constraint.

Let $\Pr(\mathcal{R}_j)$, $E[\bullet|\mathcal{R}_j]$ denote $\Pr(\mathbf{h} \in \mathcal{R}_j)$ (the probability that $\mathbf{h}$ falls in the region $\mathcal{R}_j$) and $E[\bullet|\mathbf{h} \in \mathcal{R}_j]$, respectively. Define the indicator function x_j, $j = 1, \ldots, L$ as

$$x_j = \begin{cases} 1, & \text{if } \frac{1}{M} \sum_{i=1}^{M} \frac{1}{2} \log(1 + h_i p_{ij}) < r_0, \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

Then outage minimization problem (4) with limited feedback can be formulated as

$$\min_{\mathbf{P} \geq 0, \, \mathcal{R}_j, \, \forall j} \sum_{j=1}^{L} E\left[x_j|\mathcal{R}_j\right] \Pr(\mathcal{R}_j)$$

$$\text{s.t.} \sum_{j=1}^{L} E\left[\mathbf{P}_j^{\Sigma}|\mathcal{R}_j\right] \Pr(\mathcal{R}_j) \leq P_{\text{av}}, \quad (7)$$

where $\mathbf{P}_j^{\Sigma} = \frac{1}{M} \sum_{i=1}^{M} p_{ij}$, i.e., the average of all the elements in vector $\mathbf{P}_j$. It can be easily verified that the above optimization problem satisfies the long term average power constraint with equality.

The dual problem of (7) is expressed as

$$\min_{\lambda > 0} g(\lambda) - \lambda P_{\text{av}}, \quad (8)$$

where λ is the nonnegative Lagrange multiplier associated with the long term average power constraint in Problem (7), and the Lagrange dual function $g(\lambda)$ is defined as

$$g(\lambda) = \min_{\mathbf{P}_j \geq 0, \, \mathcal{R}_j, \, \forall j} \sum_{j=1}^{L} E\left[x_j + \lambda \mathbf{P}_j^{\Sigma}|\mathcal{R}_j\right] \Pr(\mathcal{R}_j). \quad (9)$$

With a fixed λ, we can employ an iterative simulation-based optimization algorithm called the simultaneous perturbation stochastic approximation algorithm (SPSA) to find the optimal power codebook of problem (9). A step-by-step guide to an implementation of SPSA can be found in [20], which, when applied to our problem, can be summarized in the following steps.

Step 1 *Initialization and coefficient selection:* Set counter index $k = 0$. Pick initial guess of the power codebook $\hat{\mathcal{P}}_0$ and non-negative coefficients a, c, A, α and γ in the SPSA gain sequences $a_k = \frac{a}{(A+k+1)^\alpha}$ and $c_k = \frac{c}{(k+1)^\gamma}$. For guideline on choosing these coefficients see [20].

Step 2 *Generation of simultaneous perturbation vector:* Generate a p-dimensional ($p = ML$) random perturbation vector Δ_k, where each component of Δ_k are i.i.d Bernoulli ± 1 distributed with probability of $\frac{1}{2}$ for each outcome.

Step 3 *Loss function evaluations:* Obtain two measurements of the loss function $\mathcal{L}(\cdot)$ based on the simultaneous perturbation around the current power codebook $\hat{\mathcal{P}}_k$: $\mathcal{L}(\hat{\mathcal{P}}_k + c_k \Delta_k)$ and $\mathcal{L}(\hat{\mathcal{P}}_k - c_k \Delta_k)$ with c_k and Δ_k from Steps 1 and 2.

Step 4 *Gradient approximation:* Generate the simultaneous perturbation approximation to the unknown gradient $\hat{g}_k(\hat{\mathcal{P}}_k)$ given as,

$$\hat{g}_k(\hat{\mathcal{P}}_k) = \frac{\mathcal{L}(\hat{\mathcal{P}}_k + c_k \Delta_k) - \mathcal{L}(\hat{\mathcal{P}}_k - c_k \Delta_k)}{2c_k} \begin{bmatrix} \Delta_{k1}^{-1} \\ \Delta_{k2}^{-1} \\ \vdots \\ \Delta_{kp}^{-1} \end{bmatrix}$$

where Δ_{ki} is the ith component of the Δ_k vector.

Step 5 *Updating power codebook:* Use the algorithm

$$\hat{\mathcal{P}}_{k+1} = \hat{\mathcal{P}}_k - a_k \hat{g}_k(\hat{\mathcal{P}}_k)$$

to update $\hat{\mathcal{P}}_k$ to a new value $\hat{\mathcal{P}}_{k+1}$.

Step 6 *Iteration or termination:* Return to Step 2 with $k + 1$ replacing k. Terminate the algorithm if

there is little change in several successive iterations or the maximum allowable number of iterations has been reached.

Note that in the Step 3 of the SPSA which involves calculating a loss function with a given power codebook, we use the objective function of problem (9) as the loss function. And then given a power codebook, we use the nearest neighbor condition of a generalized Lloyd algorithm with a Lagrangian distortion $d(\mathbf{h}, \mathbf{P}_j) = x_j + \lambda \mathbf{P}_j^\Sigma$ to generate the optimal partition regions [21], given as, $j = 1, \ldots, L$,

$$\mathcal{R}_j = \left\{ \mathbf{h} | x_j + \lambda \mathbf{P}_j^\Sigma \leq x_i + \lambda \mathbf{P}_i^\Sigma, \forall i \neq j \right\}. \quad (10)$$

Therefore, with a given power codebook and resulting quantization regions, we can numerically calculate the loss function. We repeatedly apply Step 2 to Step 5 of SPSA until the resulting outage probability converges within a pre-specified accuracy (Step 6 of SPSA). After that, we solve the dual problem for finding the optimal λ by using a subgradient based search method, i.e., updating λ until convergence using $\lambda^{l+1} = \left[\lambda^l - \alpha^l \left(P_{av} - \sum_{j=1}^{L} E\left[\mathbf{P}_j^\Sigma | \mathcal{R}_j \right] \Pr(\mathcal{R}_j) \right) \right]^+$, where l is the iteration number, α^l is a positive scalar step sizes for the lth iteration satisfying $\sum_l \alpha^l = \infty$ and $\sum_l \alpha^{l^2} < \infty$. Due to the fact that problem (7) is not convex, in general, the optimal solution we obtain here is only locally optimal.

Figure 1 gives an example of what the structure of a locally optimal channel quantization region looks like by using SPSA with a given power codebook for $M = 2$ channels with 2 bits of feedback (i.e., four quantization regions, each color stands for a particular region), where a sufficiently large number of training samples (channel realizations for h_1, h_2) drawn from empirical distributions is used to generate the optimal partition regions according to (10). From Figure 1, we can see that, in general, it is difficult to compute the surface area (or in general volumes in higher dimensional space) of these regions which have irregular shapes. Although we can use SPSA to numerically obtain a locally optimal power codebook and partition regions, it takes a very long time to converge and is computationally highly complex especially when the number of feedback bits or the number of channels is large. In the following few sections, we therefore focus on designing sub-optimal algorithms by introducing appropriate assumptions and approximations to the quantized regions and power codebook.

Power ordering assumption and hyperplane approximation (POHPA)

Let $\mathbf{P}(\mathbf{h})$ represent the optimal power allocation strategy which maps the channel realization $\mathbf{h}$ to a power level in $\mathcal{P}$. Without loss of generality, we assume that power levels are such that $\mathbf{P}_1^\Sigma > \cdots > \mathbf{P}_L^\Sigma$ corresponding to the partition $\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_L$, then we have the following result which generalizes the circular nature of the quantized channel regions presented in [9,12] for a scalar power allocation scenario to the parallel channels case with a vector power allocation.
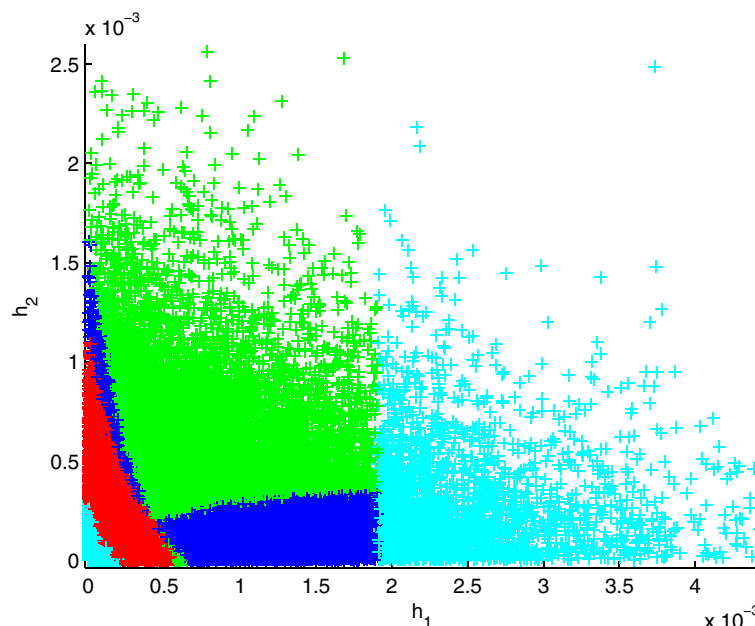

Figure 1 Structure of the optimal vector quantization regions for two channels with $B = \log_2 4$ bits of feedback.

Lemma 1. *Let* $P^*(h)$ *denotes as the minimum power level required to have no outage, i.e.,* $\frac{1}{M}\sum_{i=1}^{M}\frac{1}{2}\log\left(1+h_i p_i^*(h)\right)=r_0$. *The optimal solution satisfies:*

$$P(h)=P_j, \ \text{if} \ P_{j+1}^{\Sigma} < (P^*(h))^{\Sigma} \le P_j^{\Sigma}, \ j=1,\ldots,L-1;$$

$$P(h)=P_L, \ \text{if} \ (P^*(h))^{\Sigma} \le P_L^{\Sigma} \ \text{or} \ (P^*(h))^{\Sigma} > P_1^{\Sigma}.$$
(11)

Proof. The proof is similar to [9]. However, since it generalizes the result for a scalar power allocation in [9] to a vector power allocation case in this article, we provide a sketch of the proof (see Appendix 1). □

If the same transmit power is allocated to all transmit channels, i.e., $p_{1j} = \cdots = p_{Mj} = P_j^{\Sigma}$, the above Lemma result reduces to the case of [9,12]. From Lemma 1, we also have that there is no outage in the first $L-1$ regions and outage only occurs in the last region $\mathcal{R}_L$; the optimal partition satisfies that a channel realization $\mathbf{h} = \{h_1,\ldots,h_M\}$ either belongs to the region $\mathcal{R}_j$, where $j \in \{1,\ldots,L\}$ is the maximum index that can guarantee zero outage for it or belongs to $\mathcal{R}_L$; $\mathcal{R}_L$ includes two parts: $\left\{\mathbf{h}|(P^*(\mathbf{h}))^{\Sigma} > P_1^{\Sigma}\right\}$ (outage) and $\left\{\mathbf{h}|(P^*(\mathbf{h}))^{\Sigma} \le P_L^{\Sigma}\right\}$, denoted as $\mathcal{R}_{L,1}$ and $\mathcal{R}_{L,2}$, respectively.

From Lemma 1, we have the boundary between $\mathcal{R}_{j-1}$ and $\mathcal{R}_j$, $j = 2,\ldots,L-1$ is a hypersurface denoted as $g(h_1,\ldots,h_{M-1},\mathbf{P}_j)$, which is obtained by solving for h_M from equation $r(\mathbf{h},\mathbf{P}_j) = \frac{1}{M}\sum_{i=1}^{M}\frac{1}{2}\log(1+h_i p_{ij}) = r_0$, namely,

$$g(h_1,\ldots,h_{M-1},\mathbf{P}_j) = \frac{k - \prod_{i=1}^{M-1}(1+h_i p_{ij}) + 1}{p_{Mj}\prod_{i=1}^{M-1}(1+h_i p_{ij})}, \quad (12)$$

where $k = e^{2Mr_0} - 1$. The boundaries between $\mathcal{R}_L$ and $\mathcal{R}_1$, $\mathcal{R}_{L-1}$ are given by $g(h_1,\ldots,h_{M-1},\mathbf{P}_1)$, $g(h_1,\ldots,h_{M-1},\mathbf{P}_L)$, respectively. Let $\{r_{i1},\ldots,r_{iL}\}$ represents the quantization thresholds on h_i axes ($i = 1,\ldots,M$), from (12), it can be easily verified that $r_{ij} = \frac{k}{p_{ij}}, i \in \{1,2,\ldots,M\}, j \in \{1,2,\ldots,L\}$. Therefore, if we assume that the power levels in power codebook are in descending order, i.e., $\mathbf{P}_1 > \cdots > \mathbf{P}_L$ which means $p_{i1} > \cdots > p_{iL}, i = 1,\ldots,M$ and also implies $\mathbf{P}_1^{\Sigma} > \cdots > \mathbf{P}_L^{\Sigma}$, we can obtain $r_{i1} < \cdots < r_{iL}, i = 1,\ldots,M$, which gives a simple partition structure allowing easy numerical computation of the surface area (or volumes in higher dimensions) of the quantized regions. We call it the power ordering (PO) assumption. Figure 2 gives an example of the optimal quantization structure with the PO assumption for the $M = 2$ channels case. With this PO assumption, the area below the hypersurface $g(h_1,\ldots,h_{M-1},\mathbf{P}_1)$ defines the outage region $\mathcal{R}_{L,1}$.
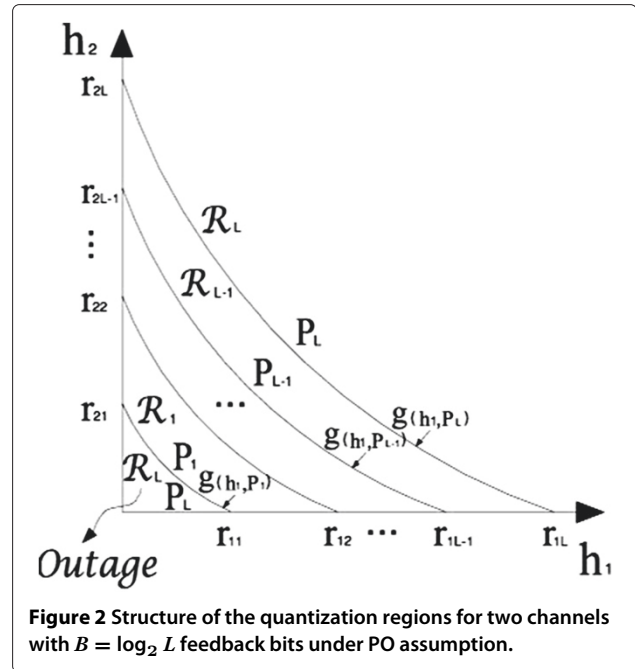

Figure 2 Structure of the quantization regions for two channels with $B = \log_2 L$ feedback bits under PO assumption.

Denoting $F(\mathbf{P}_j)$ as the probability that the channel state $(h_1,\ldots,h_M)$ lies below $g(h_1,\ldots,h_{M-1},\mathbf{P}_j)$, $j = 1,\ldots,L$, we have

$$F(\mathbf{P}_j) = \int \cdots \int_{\mathcal{R}_{L,1}\cup\mathcal{R}_1\cup\cdots\cup\mathcal{R}_{j-1}} f(h_1,\ldots,h_M)dh_1\ldots dh_M,$$
(13)

where $f(h_1,\ldots,h_M)$ is the pdf for the channel vector, given by $f(h_1,\ldots,h_M) = \prod_{i=1}^{M}(m_i\lambda_i)^{m_i}\frac{h_i^{m_i-1}}{\Gamma(m_i)}e^{-m_i\lambda_i h_i}$ for Nakagami fading (due to independence amongst the parallel channels). Thus, the probability that the channel realization $\mathbf{h} \in \mathcal{R}_j$ is $F(\mathbf{P}_{j+1}) - F(\mathbf{P}_j)$ for $j = 1,\ldots,L-1$, and $1 - F(\mathbf{P}_L) + F(\mathbf{P}_1)$ for $j = L$. The outage minimization problem with limited feedback (7) can thus be simplified as

$$\min_{\{\mathbf{P}_1 > \cdots > \mathbf{P}_{L-1} > \mathbf{P}_L \ge 0\}} P_{\text{out}} = F(\mathbf{P}_1)$$

$$\text{s.t.} \quad \sum_{j=1}^{L-1}(p_{1j}+\cdots+p_{Mj})\left(F(\mathbf{P}_{j+1}) - F(\mathbf{P}_j)\right)$$

$$+ (p_{1L}+\cdots+p_{ML})\left(1 - F(\mathbf{P}_L) + F(\mathbf{P}_1)\right) = MP_{\text{av}}.$$
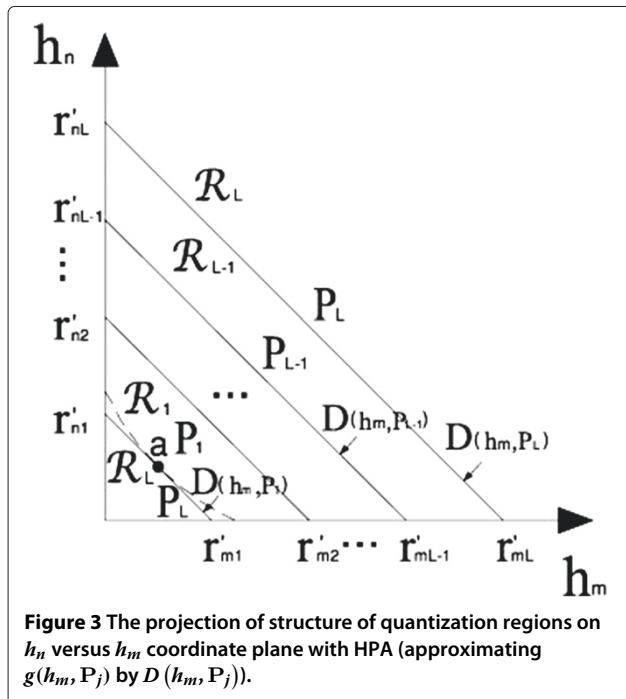(14)

Problem (14) is in general a nonlinear non-convex optimization problem. Since $g\left(h_1,\ldots,h_{M-1},\mathbf{P}_j\right), j = 1,\ldots,L$ is highly nonlinear, it is hard to obtain a closed-form expression for $F(\mathbf{P}_j)$. Although one can use numerical integrals to calculate $F(\mathbf{P}_j)$, and use randomized search techniques to find the optimum solution of problem (14), the associated computational complexity increases exponentially with the number of feedback bits and channels.

Next, we will employ another approach by deriving an approximation for $g\left(h_1, \ldots, h_{M-1}, \mathbf{P}_j\right)$, such that an analytical (approximate) closed-form expression for $F(\mathbf{P}_j)$ can be easily obtained (unlike [9] where a Gaussian distribution was used to approximate the distribution of the mutual information to evaluate an analytical expression for $F(\mathbf{P}_j)$), thus significantly reducing the computational complexity of solving problem (14). Then based on the obtained optimal power allocation using this approximation, one can use Monte Carlo simulations to evaluate the "real outage" (corresponding outage probability performance given by $F(\mathbf{P}_1)$). More details on this can be found in the *Numerical results* Section.

From (12), the projection of hypersurface $h_M = g\left(h_1, \ldots, h_{M-1}, \mathbf{P}_j\right)$, $j = 1, \ldots, L$ on any arbitrary two channel coordinate plane, i.e., h_n versus h_m, $n, m \in \{1, 2, \ldots, M\}, n \neq m$, is a curve expressed as

$$h_n = g(h_m, \mathbf{P}_j) = \frac{k - h_m p_{mj}}{p_{nj}(1 + h_m p_{mj})}. \quad (15)$$

It is easy to verify that the above curve is convex by showing that the second derivative of h_n with respect to h_m is strictly positive. This curve intersects the h_n axis and the h_m axis at quantization thresholds r_{nj} and r_{mj}, respectively. We can approximate the curve (15) by a straight line $r'_{nj}r'_{mj}$, as displayed in Figure 3, which is parallel to $r_{nj}r_{mj}$ and a tangent to the curve (15) at the intersection point 'a.'

Figure 3 The projection of structure of quantization regions on h_n versus h_m coordinate plane with HPA (approximating $g(h_m, \mathbf{P}_j)$ by $D\left(h_m, \mathbf{P}_j\right)$).

The straight line intersects h_n axis and h_m axis at point r'_{nj} and r'_{mj}, respectively. The line $r'_{nj}r'_{mj}$ is expressed as

$$h_n = D\left(h_m, \mathbf{P}_j\right) = \frac{K - h_m p_{mj}}{p_{nj}} \quad h_m \in \left[0, \frac{K}{p_{mj}}\right], \quad (16)$$

where $K = 2(\sqrt{k+1} - 1)$, $r'_{nj} = \frac{K}{p_{nj}}, r'_{mj} = \frac{K}{p_{mj}}$ (new quantization thresholds), and point 'a' is $\left(\frac{K}{2p_{mj}}, \frac{K}{2p_{nj}}\right)$. We name this approximation as the hyperplane approximation (HPA). Note that we could also use the straight line $r_{nj}r_{mj}$ to do the approximation, but simulations demonstrate that the $r'_{nj}r'_{mj}$ approximation is always better than $r_{nj}r_{mj}$. To see clearly, Figure 4 gives an example of the HPA in three-dimensional (3D) space.

Thus, with PO and HPA (POHPA), the boundaries between $\mathcal{R}_{L,1}$ and $\mathcal{R}_1$, $\mathcal{R}_1$ and $\mathcal{R}_2, \ldots, \mathcal{R}_{L-1}$ and $\mathcal{R}_{L,2}$ can be approximated as

$$D\left(h_1, \ldots, h_{M-1}, \mathbf{P}_j\right) = \frac{K - \sum_{i=1}^{M-1} h_i p_{ij}}{p_{Mj}}, j = 1, \ldots, L. \quad (17)$$

Any channel vector below $D\left(h_1, \ldots, h_{M-1}, \mathbf{P}_1\right)$ is said to be in outage. Since $D\left(h_1, \ldots, h_{M-1}, \mathbf{P}_j\right)$ is linear, an analytical closed-form approximation for $F(\mathbf{P}_j)$ can be obtained, which is denoted as $F'(\mathbf{P}_j)$. In this case, by definition we have

$$F'(\mathbf{P}_j) = \Pr\left(D(h_1, \ldots, h_{M-1}, \mathbf{P}_j) < \frac{K - \sum_{i=1}^{M-1} h_i p_{ij}}{p_{Mj}}\right)$$

$$= \Pr\left(\sum_{i=1}^{M} h_i p_{ij} < K\right). \quad (18)$$

$\sum_{i=1}^{M} h_i p_{ij}$ is a weighted sum of independent gamma random variables, and $F'(\mathbf{P}_j)$ can be treated as the cumulative distribution function (cdf) of $\sum_{i=1}^{M} h_i p_{ij}$. Thus a closed-form expression for (18) can be obtained by using any of the following two equivalent results, which however differ in their analytical derivations.

1) *Multiple infinite series representation:* This analytical expression was derived in [22],

$$F'(\mathbf{P}_j) = \frac{1}{\Gamma\left(1 + \sum_{i=1}^{M} m_i\right)} \left[\prod_{i=1}^{M} \left(\frac{m_i \lambda_i K}{p_{ij}}\right)^{m_i}\right]$$

$$\times \Phi_2^{(M)}\left(m_1, m_2, \ldots, m_M; 1 + \sum_{i=1}^{M} m_i; \right.$$

$$\left. -\frac{m_1 \lambda_1 K}{p_{1j}}, -\frac{m_2 \lambda_2 K}{p_{2j}}, \ldots, -\frac{m_M \lambda_M K}{p_{Mj}}\right), \quad (19)$$
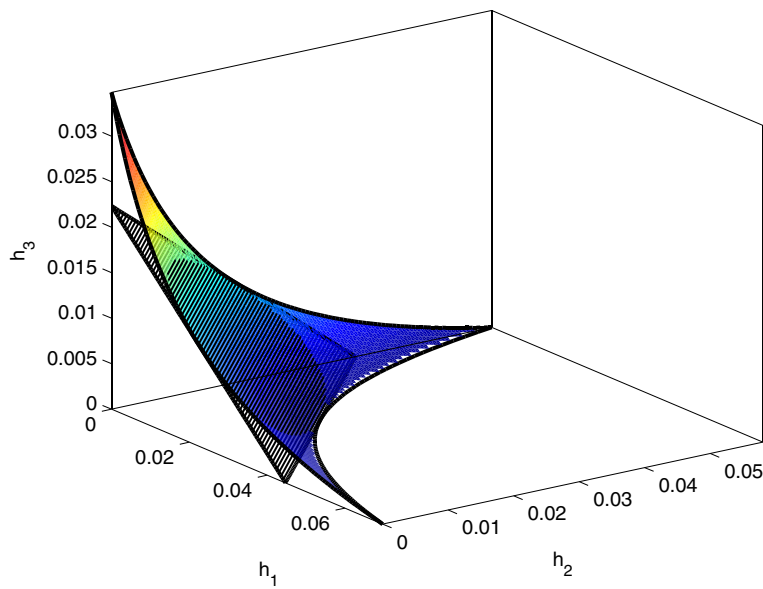
Figure 4 The projection of structure of quantization regions on h_n versus h_m coordinate plane with 3D of HPA (approximating $g(h_m, \mathbf{P}_j)$ by $D(h_m, \mathbf{P}_j)$).

where $\Phi_2^{(M)}(\ldots)$ is the confluent Lauricella multivariate hypergeometric function, involving multiple infinite sums [22]:

$$
\Phi_2^{(M)}\left(m_1, \ldots, m_M; 1 + \sum_{i=1}^{M} m_i; -\frac{m_1 \lambda_1 K}{p_{1j}}, \ldots, \right.
$$

$$
\left. -\frac{m_M \lambda_M K}{p_{Mj}}\right)
$$

$$
= \sum_{n_1=0}^{\infty} \cdots \sum_{n_M=0}^{\infty} \frac{\left[\prod_{i=1}^{M} (m_i)_{n_i}\left(-\frac{m_i \lambda_i K}{p_{ij}}\right)^{n_i} \frac{1}{n_i!}\right]}{\left(1 + \sum_{i=1}^{M} m_i\right)_{n_\tau}}, \tag{20}
$$

where $n_\tau = \sum_{i=1}^{M} n_i$ and the Pochhammer symbol is defined as $(\alpha)_k = \frac{\Gamma(\alpha+k)}{\Gamma(k)}$ [22]. Equation (19) can be numerically calculated. However, as M becomes large, computation of the multiple infinite sum may become too prohibitive to implement.

2) *Single infinite series representation:* The second result provides a simpler expression for (18) involving only a single infinite sum [23], which was proposed by Moschopoulos [24].

$$
F'(\mathbf{P}_j) = \prod_{i=1}^{M}\left(\frac{\beta_1}{\beta_{ij}}\right)^{m_i} \sum_{n=0}^{\infty} \frac{\delta_n \gamma\left(\rho + n, \frac{K}{\beta_1}\right)}{\Gamma(\rho + n)}, \tag{21}
$$

where $\rho = \sum_{i=1}^{M} m_i$, $\beta_{ij} = \frac{p_{ij}}{m_i \lambda_i}$, $\beta_1 = \min(\beta_{ij})$, $\gamma(.)$ is incomplete gamma function ($\gamma(s,x) = \int_0^x t^{s-1} e^{-t} dt$) and the coefficients δ_n are obtained recursively by

$$
\delta_{n+1} = \frac{1}{n+1} \sum_{l=1}^{n+1}\left[\delta_{n+1-l} \sum_{i=1}^{M} m_i\left(1 - \frac{\beta_1}{\beta_{ij}}\right)^l\right],
$$

$$
n = 0, 1, \ldots, \quad \delta_0 = 1. \tag{22}
$$

Special cases

- If $\rho = \sum_{i=1}^{M} m_i$ is an integer, (21) can be further simplified as [23]

$$
F'(\mathbf{P}_j) = \prod_{i=1}^{M}\left(\frac{\beta_1}{\beta_{ij}}\right)^{m_i} \sum_{n=0}^{\infty} \delta_n
$$

$$
\times \left\{1 - e^{-\frac{K}{\beta_1}} \sum_{l=0}^{\rho+n-1} \frac{\left(\frac{K}{\beta_1}\right)^l}{l!}\right\}. \tag{23}
$$

- If $M = 2$, let $\beta_2 = \max(\beta_{ij})$, and m_θ is the corresponding fading parameter for β_2, we have

$$
F'(\mathbf{P}_j) = \left(\frac{\beta_1}{\beta_2}\right)^{m_\theta} \sum_{n=0}^{\infty} \frac{(m_\theta)_n\left(1 - \frac{\beta_1}{\beta_2}\right)^n}{n!}
$$

$$
\times \frac{\gamma\left(\rho + n, \frac{K}{\beta_1}\right)}{\Gamma(\rho + n)}, \tag{24}
$$

where $(m_\theta)_{n+1}$ represents the Pochhammer symbol.

With POHPA, Problem (14) can be approximated as

$$
\min_{\{\mathbf{P}_1 > \cdots > \mathbf{P}_{L-1} > \mathbf{P}_L \geq 0\}} F'(\mathbf{P}_1)
$$

$$
\text{s.t.} \quad \sum_{j=1}^{L-1}(p_{1j} + \cdots + p_{Mj})(F'(\mathbf{P}_{j+1}) - F'(\mathbf{P}_j))
$$

$$
+ (p_{1L} + \cdots + p_{ML})\left(1 - F'(\mathbf{P}_L) + F'(\mathbf{P}_1)\right) = MP_{\text{av}}. \tag{25}
$$

It is not hard to verify that Problem (25) is still nonconvex. However, we can employ the Karush-Kuhn-Tucker (KKT) necessary conditions to achieve locally optimal solutions.

Remark 1. Note that KKT necessary conditions usually require *regularity* of a local optimum, which amounts to (in the context of Problem (25)) linear independence of the gradients of the active inequality constraints evaluated at the local optimum (see Proposition 3.3.1, p. 310 in [25]). In Problem (25), if a local optimum of the power vector satisfies $\mathbf{P}_1 > \cdots > \mathbf{P}_{L-1} > \mathbf{P}_L > 0$, then the only active inequality constraint is the average power constraint, in which case the linear independence property is trivially satisfied. In the case where the local optimum for $\mathbf{P}_L = 0$, it can be easily shown by simple linear algebra that the gradients corresponding to these two ($\mathbf{P}_L = 0$ and the average power constraint) active inequality constraints satisfy the linear independence condition.

Since regularity of a local optimum is thus established, one can now use KKT necessary conditions to obtain the following important result that can be used to design locally optimal quantized power codebooks:

Theorem 1. *Suppose* $\{p_{1j}{}^*, \ldots, p_{Mj}{}^*\}_{j=1}^L$ *be an optimum to Problem (25). Then we have*

$$
\frac{\partial F'(\mathbf{P}_j)}{\partial p_{1j}{}^*} = \cdots = \frac{\partial F'(\mathbf{P}_j)}{\partial p_{Mj}{}^*}, \quad j = 1, \ldots, L \tag{26}
$$

where $\frac{\partial F'(\mathbf{P}_j)}{\partial p_{M1}{}^*} = -\frac{\mu(F'(\mathbf{P}_2) - F'(\mathbf{P}_1))}{1 - \mu \sum_{i=1}^M (p_{i1} - p_{iL})}$, $\frac{\partial F'(\mathbf{P}_j)}{\partial p_{Mj}{}^*} = -\frac{F'(\mathbf{P}_{j+1}) - F'(\mathbf{P}_j)}{\sum_{i=1}^M (p_{i,j-1} - p_{ij})}$, $j = 2, \ldots, L-1$, *and* $\frac{\partial F'(\mathbf{P}_L)}{\partial p_{ML}{}^*} = -\frac{1 - F'(\mathbf{P}_L) + F'(\mathbf{P}_1)}{\sum_{i=1}^M (p_{i,L-1} - p_{iL})}$

Proof. See Appendix 2. □

Combining the above result with the average power constraint in (25), we have the following system of $(ML + 1)$ nonlinear equations.

$$
\begin{cases}
\sum_{j=1}^{L-1}(p_{1j} + \cdots + p_{Mj})(F'(\mathbf{P}_{j+1}) - F'(\mathbf{P}_j)) + (p_{1L} + \cdots \\
\qquad + p_{ML})(1 - F'(\mathbf{P}_L) + F'(\mathbf{P}_1)) = MP_{\text{av}}, \\[2mm]
\dfrac{\partial F'(\mathbf{P}_j)}{\partial p_{1j}{}^*} = \cdots = \dfrac{\partial F'(\mathbf{P}_j)}{\partial p_{Mj}{}^*}, \quad j = 1, \ldots, L \\[2mm]
\dfrac{\partial F'(\mathbf{P}_j)}{\partial p_{M1}{}^*} = -\dfrac{\mu(F'(\mathbf{P}_2) - F'(\mathbf{P}_1))}{1 - \mu \sum_{i=1}^M (p_{i1} - p_{iL})}, \\[2mm]
\dfrac{\partial F'(\mathbf{P}_j)}{\partial p_{Mj}{}^*} = -\dfrac{F'(\mathbf{P}_{j+1}) - F'(\mathbf{P}_j)}{\sum_{i=1}^M (p_{i,j-1} - p_{ij})}, \quad j = 2, \ldots, L-1, \\[2mm]
\dfrac{\partial F'(\mathbf{P}_L)}{\partial p_{ML}{}^*} = -\dfrac{1 - F'(\mathbf{P}_L) + F'(\mathbf{P}_1)}{\sum_{i=1}^M (p_{i,L-1} - p_{iL})}, \\[2mm]
\mathbf{P}_1 > \cdots > \mathbf{P}_L.
\end{cases} \tag{27}
$$

A solution to (27) provides a locally optimum power allocation policy $\{\mathbf{P}_j{}^*\}_{j=1}^L$. For small values of L and M, the above system of nonlinear equations can be solved by various optimization softwares. However, the complexity of solving the above set of nonlinear equations is still too high for moderately large numbers of feedback bits and channels. Therefore, we consider several low-complexity suboptimal schemes suited to special cases of high or low P_{av} as described below.

High average power approximation (HP_{av}A)

In the high average power or average SNR regime, the following result allows us to simplify the computation of the quantized power codebook. It also illustrates that using our hyperplane based approximations, it is not optimal to allocate identical power to individual channels within each quantized region in general.

Theorem 2. *For arbitrary M, in high average power (as* $P_{av} \to \infty$*), the multiple infinite series representation (19),* $F'(\mathbf{P}_j), j = 1, \ldots, L$ *can be further approximated as*

$$
F'(\mathbf{P}_j) \approx \frac{1}{\Gamma\left(1 + \sum_{i=1}^M m_i\right)} \prod_{i=1}^M \left(\frac{m_i \lambda_i K}{p_{ij}}\right)^{m_i} \tag{28}
$$

and a locally optimum power allocation scheme for (25) satisfies the following approximate relationship:

$$
\frac{m_i}{p_{ij}{}^*} \approx \frac{m_l}{p_{lj}{}^*}, \quad i, l \in \{1, 2, \ldots, M\}, \ i \neq l, \ j = 1, \ldots, L. \tag{29}
$$

Proof. See Appendix 3. □

(29) implies that in high P_{av}, for each quantization region, the power allocated to each channel asymptotically depends only on the severity of fading (represented by the parameter m).

Special cases:

- *Identical fading parameters*: If $m_1 = \cdots = m_M$, from (29), we have

$$p_{1j}{}^* \approx \cdots \approx p_{Mj}{}^*, \ j = 1, \ldots, L \qquad (30)$$

which means, in high P_{av}, with identical fading parameters for all channels, for each quantization region, the power assigned to each channel is asymptotically equal, and we call this solution as equal power per channel (EPPC).

- *Rayleigh fading*: $(m_1 = \cdots = m_M = 1)$, from Theorem 2, (28) reduces to

$$F'(\mathbf{P}_j) \approx \frac{1}{M!} \prod_{i=1}^{M} \left(\frac{\lambda_i K}{p_{ij}} \right) \qquad (31)$$

and (29) reduces to EPPC.

For the general case (29), without loss of generality, by letting $p_{ij} = \frac{m_i}{m_M} p_{Mj}$, $\forall i \in \{1, 2, \ldots, M-1\}$ and denoting p_{Mj} as p_j for simplicity, the above $ML+1$ equations system (27) can be simplified into an $L+1$ equations system only:

$$
\begin{cases}
\sum_{j=1}^{L-1} p_j(F'(p_{j+1}) - F'(p_j)) + p_L(1 - F'(p_L) + F'(p_1)) = P'_{av}, \\
\dfrac{\partial F'(p_j)}{\partial p_j} = -\dfrac{F'(p_{j+1}) - F'(p_j)}{(p_{j-1} - p_j)}, \ j = 2, \ldots, L-1 \\
\dfrac{\partial F'(p_L)}{\partial p_L} = -\dfrac{1 - F'(p_L) + F'(p_1)}{(p_{L-1} - p_L)},
\end{cases}
\qquad (32)
$$

where $P'_{av} = P_{av} \frac{M m_M}{\sum_{i=1}^{M} m_i}$.

Thus $HP_{av}A$ reduces the M-dimensional vector quantization problem into a one-dimensional scalar quantization problem, as illustrated in Figure 5, with corresponding quantization thresholds $r_1, \ldots, r_L$, where $r_j = r'_{Mj} = K/p_j$, and remarkably reduces the complexity. We call this

Figure 5 One-dimensional scalar quantization model: PFPPC.
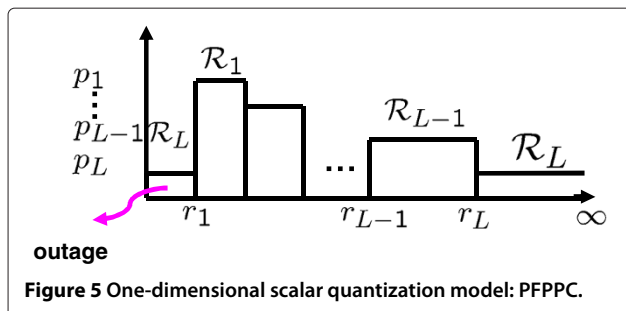
suboptimal scheme as the proportional to fading parameter per channel (PFPPC) scheme. For small values of L, the above L nonlinear equations (where one can evaluate $F'(p_j)$ using the single infinite series representation (21)) can be solved by various optimization softwares.

For large values of L (e.g., $L \geq 16$ or $B \geq 4$), one can use the so-called equal average power per region (EPPR) approximation for such a scalar quantization problem by using the mean value theorem [10]. This essentially implies that when L goes to infinity, the total average power assigned to each quantization region is asymptotically equal and the performance using this approximation is close to optimum for large number of bits of feedback. In this case, we need to solve the following set of L equations instead of (32)

$$p_j(F'(p_{j+1}) - F'(p_j)) = \frac{P'_{av}}{L}, \ j = 1, \ldots, L-1;$$

$$p_L(1 - F'(p_L) + F'(p_1)) = \frac{P'_{av}}{L}. \qquad (33)$$

which can be carried out by an iterative algorithm employing the standard bisection search method. We call this algorithm as 'PFPPC+EPPR'.

Let P_{tot}^j represents the total average power allocated to region $\mathcal{R}_j$ at PFPPC case. Then the average power constraint in (32) can be rewritten as,

$$\sum_{j=1}^{L} \frac{P_{tot}^j}{p_j} = 1. \qquad (34)$$

Since $p_1 > p_2 > \cdots > p_L$, $\sum_{j=1}^{L} P_{tot}^j = P'_{av}$, it follows that

$$p_1 > P'_{av}. \qquad (35)$$

Thus, in high the P_{av} regime $(P_{av} \to \infty)$, $r_1 = \frac{K}{p_1} \to 0$, and we have the following result which indicates that the total power allocated to the outage region is asymptotically (as $P_{av} \to \infty$) negligible thus allowing us to further simplify the quantized power codebook design method:

Lemma 2. *In the high P_{av} regime, $\lim_{r_1 \to 0} P_{tot}^{L,1} = 0$, if $\sum_{j=1}^{L} \rho^j > 1$, where $\rho = \sum_{i=1}^{M} m_i$.*

Proof. See Appendix 4. □

Therefore another effective scheme for large L is to additionally (to PFPPC) employ what we call the zero power in outage region (ZPiOR) approximation (PFPPC+ZPiOR), by letting the power level $p_L = 0$ resulting in $r_L = \frac{K}{p_L} \to \infty$, as showed in Figure 6. Thus we have the total average power allocated to outage region $\mathcal{R}_{L,1}$, $P_{tot}^{L,1} = p_L * F'(p_1) = 0$.

Figure 6 One-dimensional scalar quantization model and ZPiOR approximation: PFPPC+ZPiOR.
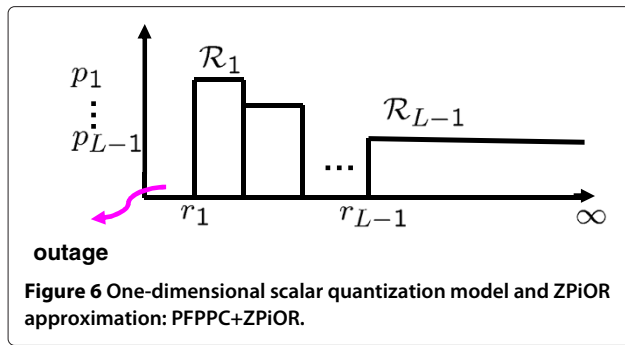
Remark 2. Note that if $\rho \geq 1$, the condition $\sum_{j=1}^{L} \rho^j > 1$ is clearly satisfied for any $L \geq 2$. For $0.5 < \rho < 1$ (which is the case of no diversity with $M = 1$ or the single channel case), one can show that there exists a finite L for which the condition $\sum_{j=1}^{L} \rho^j > 1$ is satisfied. This is easily seen by noting that the condition $\sum_{j=1}^{L} \rho^j > 1$ is equivalent to $\rho^{L+1} < 2\rho - 1$ for $\rho < 1$. It is interesting to note however that when $\rho = 0.5$ (which is the case when one has as single Nakagami channel with $m = 0.5$, the worst possible fading parameter), there is no finite value of L that can achieve $\sum_{j=1}^{L} \rho^j > 1$. Thus in high P_{av}, it is near optimal to allocate zero power to the outage region as long as $\rho \geq 1$ with any $L \geq 2$, or a single channel with $0.5 < m < 1$ and a sufficiently large L. For a single channel with $m = 0.5$, it seems that even in high P_{av}, one needs to allocate nonzero power to the outage region.

Therefore the performance of the ZPiOR approximation (PFPPC+ZPiOR) becomes asymptotically (as $L \rightarrow \infty$) close to that of the PFPPC scheme, except for the single channel case with $m = 0.5$, where one can use the EPPR approximation instead to reduce complexity.

In this case, (32) can be simplified as

$$
\begin{cases}
\sum_{j=1}^{L-1} p_j(F'(p_{j+1}) - F'(p_j)) = P'_{av} \\
\\
p_{j-1} = p_j + \dfrac{F'(p_{j+1}) - F'(p_j)}{-\dfrac{\partial F'(p_j)}{\partial p_j}} \quad j = 2, \ldots, L-1, F'(p_L) = 1
\end{cases}
\tag{36}
$$

which can be easily solved by using a standard bisection method. In fact, numerical studies illustrate (as we will see later) that the ZPiOR approximation has a near-optimum (for Problem (25)) performance for large number of quantization regions. Thus, the ZPiOR approximation achieves a better complexity-performance tradeoff than PFPPC+EPPR.

Remark 3. For the low P_{av} scenario, we can apply the ZPiOR approximation as well. This is because it is easy to verify that

$$
\frac{1}{M}(p_{1L} + \cdots + p_{ML}) < P_{av}
\tag{37}
$$

then when the average power is small ($P_{av} \rightarrow 0$), $p_{iL} \rightarrow 0, i = 1, \ldots, M$ as well, and the corresponding quantization threshold $r_{iL} \rightarrow \infty$. In this case, the region $\mathcal{R}_L$ only includes $\mathcal{R}_{L,1}$ (the outage region) and the corresponding power level $\mathbf{P}_L = 0$, thus making the ZPiOR approximation applicable. A similar observation was also made in [9].

Asymptotic behavior of outage probability

Here we briefly comment on the diversity behavior of the outage minimization algorithm using POHPA for Nakagami-m fading channels. Define the diversity gain d as

$$
d = -\lim_{P_{av} \rightarrow \infty} \frac{\log P_{out}}{\log P_{av}}.
\tag{38}
$$

Then we have the following important approximation for d, which generalizes the existing diversity order results for outage probability (e.g., [13]), that are valid for Rayleigh fading channels only.

Theorem 3. *For an arbitrary M, with $\log_2 L$ bits of quantized feedback, using the optimal power allocation employing the POHPA approximation, we have*

$$
P_{out} \approx \frac{\left(\frac{\rho L}{Mm_M}\right)^{\rho^L + \cdots + \rho} c^{\rho^{L-1} + \cdots + \rho + 1}}{P_{av}^{\rho^L + \cdots + \rho}}.
\tag{39}
$$

The diversity order can be approximated as

$$
d \approx \sum_{j=1}^{L} \left(\sum_{i=1}^{M} m_i\right)^j.
\tag{40}
$$

Proof. See Appendix 5. □

Special case: Note that for the Rayleigh fading case where $m_i = 1, \forall i = 1, 2, \ldots, M$, (40) becomes $d \approx \sum_{j=1}^{L} M^j$, which is consistent with similar results in [9,10].

Remark 4. It is possible that the result in Theorem 3 may hold with equality, rather than being an approximation for the diversity order. However, due to the various levels of approximations involved in deriving this, we are unable to prove an exact equality at this stage. This will involve computing orders of approximation errors and showing that the error goes to zero as P_{av} goes to infinity. We leave this for future study.

Large number of channels analysis

The previous algorithms can be effectively applied to find locally optimal solutions or suitable approximations for them for moderate number of parallel channels, such as $M < 10$. Once $M \geq 10$, these algorithms become computationally demanding. Given that practical wideband slow fading multi-carrier systems such as OFDM (with large number of sub-carriers), can be modeled as asymptotically large number of i.i.d parallel channels [18], one needs to find outage minimizing power allocation algorithms with limited feedback for large M. Below we provide such an algorithm using a Gaussian approximation for large M in high P_{av}.

Remark 5. According to [18], the mutually independence of large number of parallel channels can be justified by the assumptions that the number of independent propagation paths in wideband slow fading channel models increase linearly with the bandwidth and the carrier frequencies of parallel channels are sufficiently separated with the effect of the multipath spread essentially eliminated [18]. Even if the adjacent subcarriers are correlated, with subcarrier grouping technique [26], we can still have a large number of parallel independent subchannels (each comprises a number of subcarriers), such as 32 subchannels as stated in [26].

Note that in high average power, we have

$$\sum_{i=1}^{M} \log(1 + p_{ij}h_i) \approx \sum_{i=1}^{M} \log(p_{ij}h_i) = \sum_{i=1}^{M} \log\left(p_{ij}\frac{f_i}{\lambda_i}\right)$$

$$= \sum_{i=1}^{M} \log\left(\frac{p_{ij}}{\lambda_i}\right) + \sum_{i=1}^{M} \log(f_i) \qquad (41)$$

where $f_i = h_i\lambda_i$, $\frac{1}{\lambda_i}$ is the mean of channel gain h_i, and under the Nakagami fading model, the pdf of f_i is $\frac{(m_i)^{m_i}}{\Gamma(m_i)}f_i^{m_i-1}e^{-m_if_i}, \forall i$.

Thus $F(\mathbf{P}_j), j = 1, \ldots, L$ for M channels can be approximated as

$$F(\mathbf{P}_j) = \text{Prob}\left(\frac{1}{M}\sum_{i=1}^{M}\frac{1}{2}\log(1 + p_{ij}h_i) < r_0\right)$$

$$\approx \text{Prob}\left(\frac{1}{M}\sum_{i=1}^{M}\log(f_i) < s_j\right) = V(s_j) \qquad (42)$$

where $s_j = 2r_0 - \frac{1}{M}\sum_{i=1}^{M}\log\left(\frac{p_{ij}}{\lambda_i}\right) = c' - \frac{1}{M}\sum_{i=1}^{M}\log(p_{ij})$, $c' = 2r_0 + \frac{1}{M}\sum_{i=1}^{M}\log(\lambda_i)$ and the function $V(.)$ denotes the cdf of $\frac{1}{M}\sum_{i=1}^{M}\log(f_i)$. It is easy to show that the pdf of $z_i = \log(f_i)$ is $f_{z_i} = \frac{(m_i)^{m_i}}{\Gamma(m_i)}e^{-m_ie^{z_i}}e^{m_iz_i}$. Denote its mean and variance by $E[z_i]$ and $\text{Var}[z_i]$, respectively. For the Rayleigh fading case, the pdf of $z_i = \log(f_i)$ is $e^{-e^{z_i}}e^{z_i}$, which is the well known *Gumbel Distribution* with mean $E[z_i] = -r$, where r is Euler-Mascheroni constant ($r = 0.5772156649\ldots$) and variance $\text{Var}[z_i] = \frac{\pi^2}{6}$.

Note that for large M, if $m_1 = \cdots = m_M$ or in the special case of Rayleigh fading ($m_i = 1, \forall i$), z_i is i.i.d with finite mean and variance and then the Central Limit Theorem (CLT) directly applies whereby one can use a Gaussian approximation for the pdf of $\frac{1}{M}\sum_{i=1}^{M}z_i$. However, in the general case where the fading parameters m_i are different for different channels, z_i, $i = \{1, 2, \ldots, M\}$ are independent but not necessarily identically distributed. In this case, it is important to prove that $\{z_i - E[z_i]\}$ satisfies the Lindeberg condition (for a statement of this condition, see p. 262, [27]), so that a generalized CLT can be applied and a Gaussian approximation can be used for the instantaneous mutual information over parallel fading channels. Indeed, we can analytically prove the following Lemma:

Lemma 3. *The sequence $\{z_i - E[z_i]\}$ satisfies the Lindeberg condition.*

Proof. The proof of this can be found in Appendix 6. □

Therefore when the number of channels $M \to \infty$, the cdf of $\frac{1}{M}\sum_{i=1}^{M}z_i$ can still be approximated (by applying Theorem 3, Chap. VIII.4 in [27]) by a Gaussian cdf with mean and variance given by

$$\mu = \frac{1}{M}\sum_{i=1}^{M}E[z_i], \quad \sigma^2 = \frac{1}{M^2}\sum_{i=1}^{M}\text{Var}[z_i]. \qquad (43)$$

Thus, we have

$$V(s_j) \approx \int_{-\infty}^{s_j}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx = \frac{1}{2}\left[1 + \text{erf}\left(\frac{s_j - \mu}{\sigma\sqrt{2}}\right)\right]. \qquad (44)$$

The original problem (14) for a large number of channels case can be approximated as

$$\min_{\{\mathbf{P}_1 > \mathbf{P}_2 > \cdots > \mathbf{P}_L \geq 0\}} V(s_1)$$

$$\text{s.t.} \quad \sum_{j=1}^{L-1}(p_{1j} + \cdots + p_{Mj})\left(V(s_{j+1}) - V(s_j)\right)$$

$$+ (p_{1L} + \cdots + p_{ML})\left(1 - V(s_L) + V(s_1)\right) = MP_{\text{av}}. \qquad (45)$$

Using the KKT necessary conditions, we again get

$$\frac{\partial V(s_j)}{\partial p_{1j}} = \cdots = \frac{\partial V(s_j)}{\partial p_{Mj}}, j = 1, \ldots, L. \qquad (46)$$

Note that

$$\frac{\partial V(s_j)}{\partial p_{ij}} = f(s_j)\frac{\partial s_j}{\partial p_{ij}} = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{\left(c - \frac{1}{M}\sum_{i=1}^{M}\log(p_{ij}) - \mu\right)^2}{2\sigma^2}} - \frac{1}{M}\frac{1}{p_{ij}},$$
(47)

where $f(s) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(s-\mu)^2}{2\sigma^2}}$. It is easily seen that the above expression for $\frac{\partial V(s_j)}{\partial p_{ij}}$ is a monotonically increasing function of p_{ij} for all $i = 1, 2, \ldots, M$ and $j = 1, 2, \ldots, L$. Therefore, from (46), we have

$$p_{1j} = \cdots = p_{Mj}, \quad j = 1, \ldots, L.$$
(48)

The above result implies that at a local optimum, using the Gaussian approximation, the power levels (for each quantization region) for all channels are identical, which is identical to the EPPC scheme. With a slight abuse of notation, denote $p_j = p_{ij}, j = 1, \ldots, L$. The we have $s_j = c' - \log(p_j)$ and the vector quantization problem (45) can be converted into the scalar quantization problem below with quantization thresholds $s_1, \ldots, s_L$:

$$\min_{\{p_1 > \cdots > p_L \geq 0\}} V(s_1)$$

$$\text{s.t.} \quad \sum_{j=1}^{L-1} p_j(V(s_{j+1}) - V(s_j)) + p_L(1 - V(s_L) + V(s_1)) = P_{av}.$$
(49)

After employing the corresponding KKT necessary optimality conditions and simplifying, we have the system of L nonlinear equations below:

$$\begin{cases} \sum_{j=1}^{L-1} p_j(V(s_{j+1}) - V(s_j)) + p_L(1 - V(s_L) + V(s_1)) = P_{av}, \\ p_{j-1} = p_j\left(\frac{V(s_{j+1}) - V(s_j)}{f(s_j)} + 1\right), \quad j = 2, \ldots, L-1 \\ p_{L-1} = p_L\left(\frac{1 - V(s_L) + V(s_1)}{f(s_L)} + 1\right). \end{cases}$$
(50)

When L is not large, one can solve the above equations using optimization tools 1stOpt. When L is large (roughly $L \geq 16$), we can also use the EPPR approximation or the ZPiOR approximation to solve them, as discussed in the section on high P_{av} approximations. Table 1 below

Table 1 Proposed power allocation strategies

Number of channels	$M < 10$	$M \geq 10$
Optimal approach	SPSA	-
Approximation	POHPA	-
High P_{av}	PFPPC ($L \geq 16$,	GA ($L \geq 16$,
	PFPPC+ZPiOR or PFPPC+EPPR)	GA+ZPiOR or GA+EPPR)

shows the applicability of various algorithms discussed so far according to different ranges of M, L and high P_{av}, where "GA" denotes the Gaussian approximation based algorithms.

Numerical results

To numerically illustrate the performance of the designed power allocation strategies, we consider an M-parallel (independent) Nakagami block-fading channels, which characterizes a multicarrier OFDM system, with M parallel subchannels located at nonadjacent carrier frequencies. The mean value of the gamma distributed fading gain for each channel is assumed to be inversely proportional to the square of the wireless propagation distance d, and the required transmission rate is taken to be $r_0 = 0.25$ nats per channel use. Outage performance with full CSI at the transmitter is obtained with the optimal power allocation results presented in [3]. It should be noted that the results illustrate the "real outage" performance of the proposed algorithms, where the power codebook designed via the algorithms is used to obtain the average outage probability over a large number of Monte-Carlo simulated channel realizations. As a result, the average power required for a given *real* outage may not strictly be the same as the original average power based on which the power codebook is designed. However, for a given algorithm, the graphs can and should be used to determine the minimum outage probability obtainable for a given average power and vice versa.

Experiment 1. The first experiment examines the performance of POHPA. Figure 7 compares the outage performance of SPSA, an exhaustive search over the space of all possible power allocation policies implementing the power ordering (PO) approximation only, and POHPA with 1 bit feedback for two channels case ($d_1 = 40$ m, $d_2 = 60$ m, $m_1 = m_2 = 0.5$). It can be observed that when P_{av} is small, the performance of these three methods have negligible difference, while when P_{av} is large, SPSA is slightly better than PO and POHPA. Figure 8 shows the outage performance of SPSA and POHPA for a higher dimensional case (four channels case, $d_1 = 30$ m, $d_2 = 40$ m, $d_1 = 60$ m, $d_2 = 70$ m, $m_1 = m_2 = m_3 = m_4 = 1$). Again, it can be seen clearly that with identical number of feedback bits, the outage probability gap between SPSA and POHPA gradually increases as P_{av} increases. From Figures 7 and 8, it seems that in a higher dimensional space (larger number of channels), with the same number of feedback bits, the outage probability gap between the two methods is bigger than the one in a low dimensional space, especially in high P_{av}. And with the same value of M, as the number of feedback bits increases, this gap seems to decrease, as shown in Figure 8. This can be explained due to the fact that to achieve a fixed outage
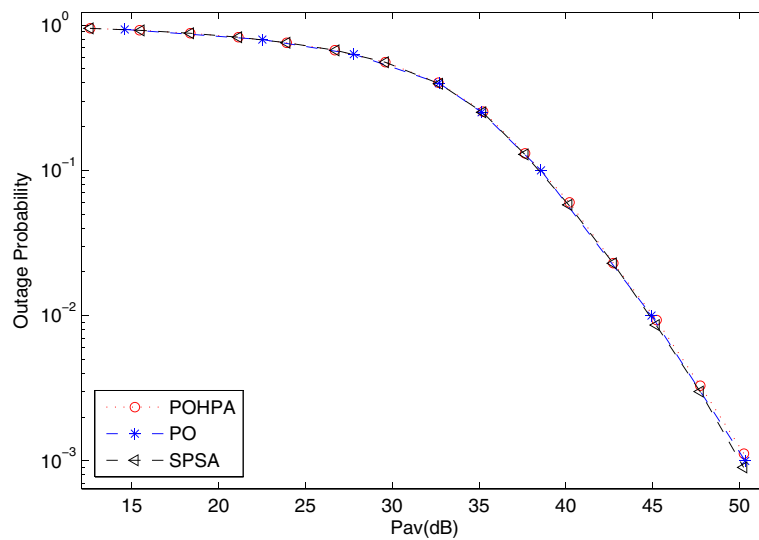
Figure 7 Outage performance comparison between SPSA, PO, and POHPA for two channels 1 bit feedback ($m_1 = m_2 = 0.5$).

probability, a larger number of feedback bits requires less P_{av}.

Experiment 2. This experiment tests the performance of suboptimal scheme HP_{av}A (namely, PFPPC). Figure 9 compares the outage performance of PFPPC approximation with its optimal case (POHPA) for two channels ($d_1 = 40$ m, $d_2 = 60$ m) with 1 bit feedback ($m_1 = m_2 = 0.5$). The striking observation in Figure 9 is that when $P_{av} \leq 32$ dB, POHPA only slightly outperforms PFPPC, while when $P_{av} > 32$ dB, the performance of PFPPC and POHPA almost overlaps each other

(i.e., the performance of PFPPC is very close to its optimum), indicating that PFPPC is an efficient near-optimal scheme for POHPA especially at high P_{av}. In addition, Figure 10 illustrate the efficiency of using PFPPC+EPPR and PFPPC+ZPiOR schemes for HP_{av}A (PFPPC) at large number of feedback bits. As shown in Figure 10, with 4 bits of feedback (16 regions), PFPPC with ZPiOR approximation (PFPPC+ZPiOR) achieves almost equivalent performance to PFPPC with EPPR approximation (PFPPC+EPPR), and both schemes are very close to their optimal case (PFPPC). This result illustrates the fact that ZPiOR can be combined with PFPPC as a computationally
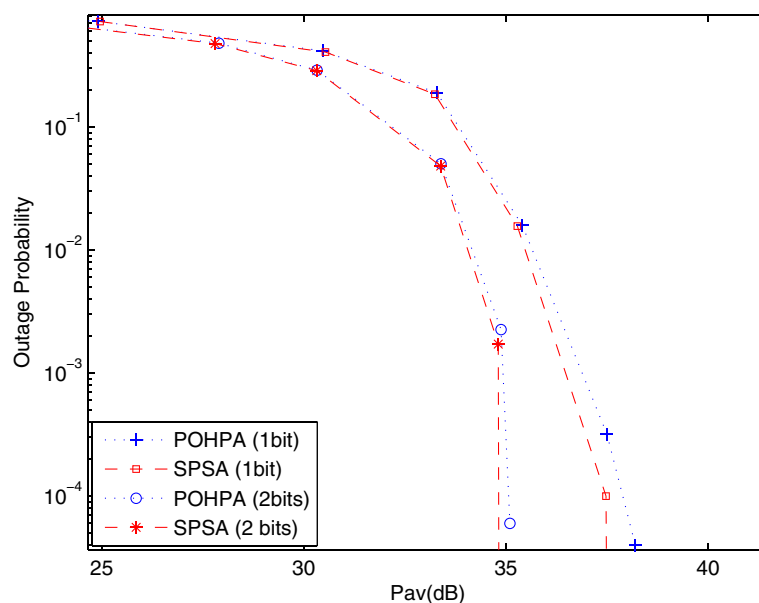
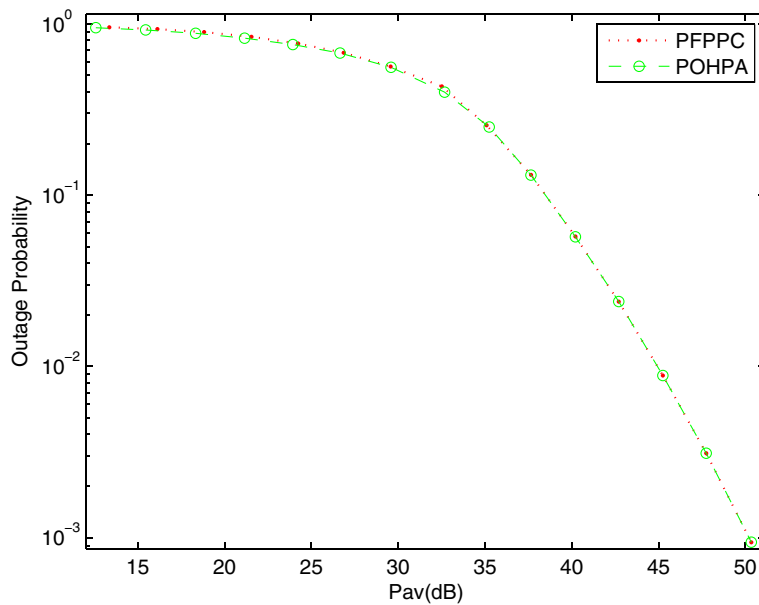**Figure 8 Outage performance comparison between SPSA and POHPA for four channels ($m_1 = m_2 = m_3 = m_4 = 1$).**

Figure 9 Outage performance of HPavA (PFPPC) scheme ($M = 2, B = 1, m_1 = m_2 = 0.5$).

simpler alternative to PFPPC+EPPR for large number of feedback bits.

Experiment 3. The third simulation, as illustrated in Figure 11 for four channels case ($\mathbf{d} = [30\,m, 40\,m, 60\,m, 70\,m]$ and $m_1 = m_2 = m_3 = m_4 = 0.5$), studies the effect of increasing the number of feedback bits on the outage performance using the proposed schemes. For comparison, the performance of the optimal power control policy with full CSI [3] is also shown. Instead of comparing the performance with the POHPA scheme, we plot the outage probabilities of its computationally efficient near-optimal schemes (PFPPC and its variants). With a small number of bits (1 bit and 2 bits) of feedback, PFPPC can be implemented by using the optimization software 1stopt, and with 4 bits of feedback ($L = 16$), we plot the performances of PFPPC+ZPiOR instead. The important observation from this figure is that the
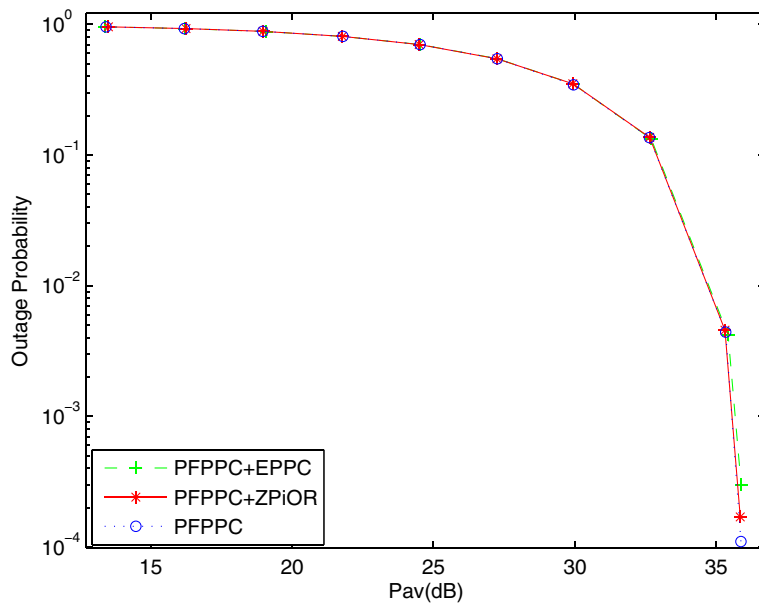
Figure 10 Outage performance of PFPPC+EPPR and PFPPC+ZPiOR schemes with 4 bits feedback for two channels and $m_1 = m_2 = 1$ ($\rho > 1$).
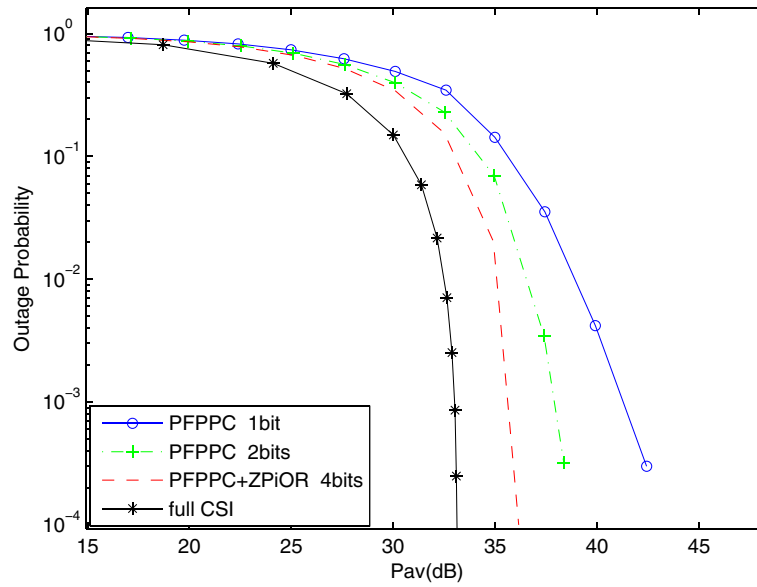
Figure 11 Effect of increasing feedback bits on outage performance for four channels ($m_1 = m_2 = m_3 = m_4 = 0.5$).

introducing one extra bit of feedback substantially reduces the gap with the full CSI performance and only a few bits of feedback can eliminate most of the gap with the full CSI performance. For example, at an outage probability of 10^{-2}, with 4 bits of feedback, there is only around 2.5 dB power loss compared to the full CSI case. This confirms that power allocation with limited feedback (only with a few feedback bits) can provide a dramatic performance advantage over no CSI (channel non-adaptive power allocation across all channels).

Figure 12 depicts the diversity behavior of the proposed outage minimization scheme POHPA using the derived bound given in (39) for four channels with $m_1 = m_2 = m_3 = m_4 = 0.5$. In high P_{av}, the outage performance of the PFPPC scheme is very close to optimum (POHPA), thus here we plot the performance of the computationally efficient PFPPC scheme instead of POHPA. We also use (28) to approximate the outage expression when the outage probability $\leq 10^{-10}$. As we can see from Figure 12, the derived bound captures the slope of outage behavior in high P_{av} extremely well.

Experiment 4. Figure 13 shows the effect of the fading parameter m on the outage performance. It depicts the outage performance with 4 bits of feedback over six Nakagami fading channels with different values of the fading parameter: $m = 0.5$, $m = 1$, and $m = 2$ (here we use identical fading parameter for each channel, i.e., $m_1 = \cdots = m_6 = m$, and $d = [20, 30, 40, 60, 70, 80]$). It can be noticed that as m increases, i.e., the fading severity decreases, significant performance gains can be easily

observed. To achieve a target outage probability 10^{-2}, 4 bits of feedback with $m = 2$ provides around 2.65 and 1.18 dB improvements over 4 bits of feedback with $m = 0.5$ and $m = 1$, respectively, as measured by the SNR gap with respect to their respective full CSI performances.

Experiment 5. Figure 14 compares the outage performance between the PFPPC scheme and the Gaussian approximation (GA) for four channels case ($m_1 = m_2 = m_3 = m_4 = 0.5$). It can be seen very clearly that the PFPPC scheme outperforms GA, the benefit of the PFPPC scheme becoming more pronounced as P_{av} increases. For instance, with the same feedback bits, at an outage probability of 10^{-3}, PFPPC with 1 bit requires roughly 7.8 dB less power than GA does; and PFPPC with 2 and 4 bits feedback provide around 5.6 and 2.3 dB power savings over GA, respectively. Even with only with 1 bit of feedback, PFPPC can achieve a better performance than GA with 2 bits of feedback in high P_{av}. These results indicate that the POHPA (with the PFPPC approximation) can achieve remarkable performance advantage over GA, especially in high P_{av}. However, when the number of channels is large, POHPA becomes computationally prohibitive. And in this case, GA is an efficient alternative, which is consistent with similar observations (for MIMO settings) in [28].

Figure 15 illustrates the outage probability over large number of channels (16 channels) using GA, with the values of distances $d_1, \ldots, d_{16}$ randomly obtained (with a uniform distribution) from the range [20 m, 100 m] and different identical fading parameters ($m_1 = \cdots = m_{16} =$
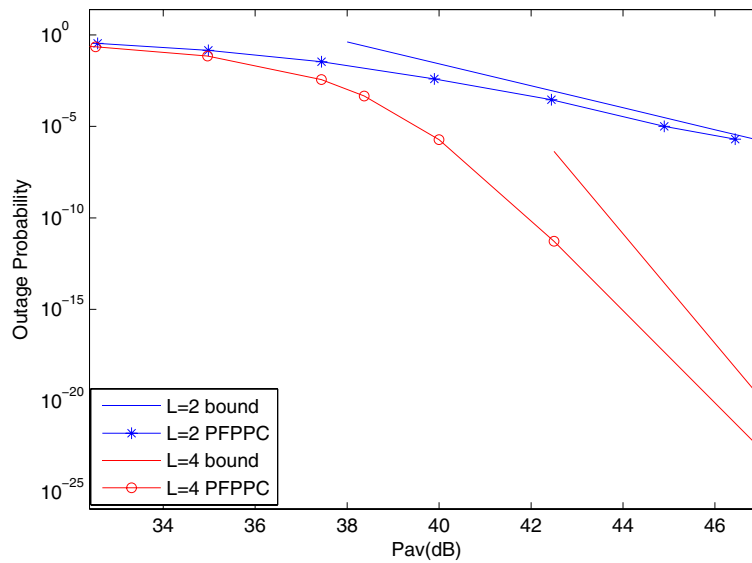
Figure 12 Bound on diversity order for $L = 2$ and $L = 4$ with PFPPC scheme ($M = 4$, $m_1 = m_2 = m_3 = m_4 = 0.5$).

m) $m = 0.5$ and $m = 2$, respectively. We again see that only a few bits of feedback are required to close the gap with the performance with full CSI. For instance, to achieve a target outage probability 10^{-2}, with $m = 2$, the power consumption gap between 10 bits of feedback (less than one bit per channel) and its full CSI based counterpart is only about 2.4 dB, while with $m = 0.5$, the gap is 2.8 dB. On the other hand, as m decreases, i.e., the fading severity increases, the outage performance of the limited feedback schemes deteriorates as expected.

Conclusions

In this article, we have derived a simulation based optimization algorithm using SPSA and presented various low-complexity sub-optimal outage minimization algorithms via optimal power allocation with finite-rate or quantized channel feedback for an M-parallel block-fading channels under a long term average power constraint. Numerical results illustrate the effectiveness of these algorithms via their outage performance in comparison with the performance of the optimal power allocation
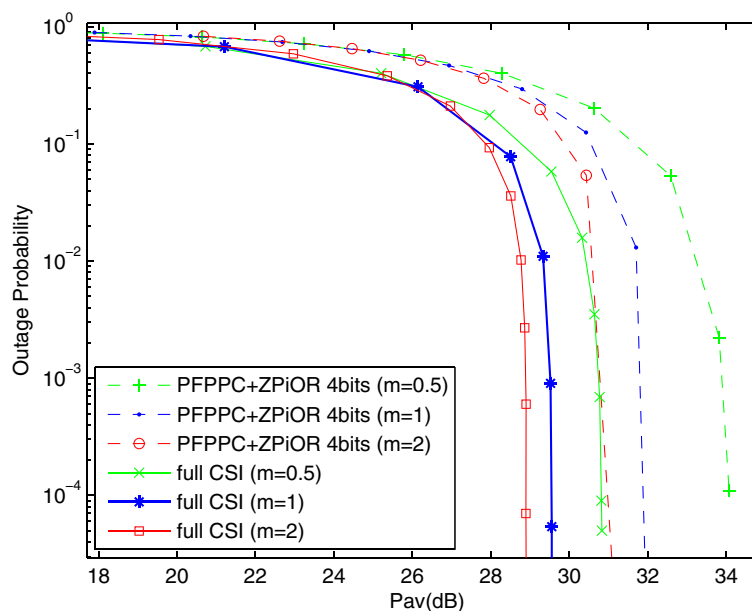

Figure 13 Effect of the fading parameter m on outage performance for six channels.
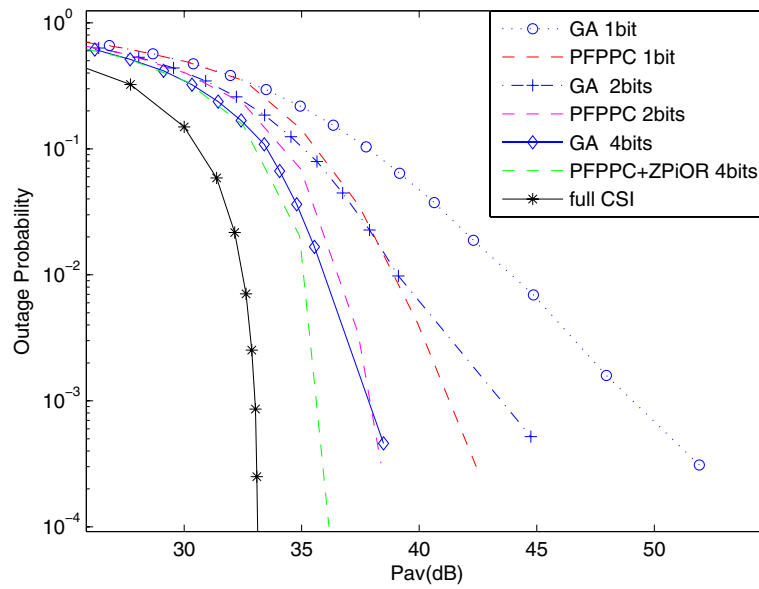
Figure 14 Outage performance comparison between POHPA and GA ($M = 4$, $B = 1, 2, 4$ bits, $m_1 = m_2 = m_3 = m_4 = 0.5$).

with full CSI, and show that only 4 bits of feedback close the gap with the outage performance of the full CSI algorithm substantially for $M = 4$ or $M = 6$. For a large number of channels ($M = 16$), our Gaussian approximation based algorithm performs approximately within 2.8 dB (SNR gap) of the full CSI based algorithm at an outage probability of 10^{-2} with less than 1 bit of (broadcast) feedback per channel when all channels undergo severe Nakagami fading with identical fading parameter $m = 0.5$. Future study includes extension of these results to correlated fading channels, consideration of noisy or erroneous feedback as investigated in [29] and quantized CSIT based power allocation to more general optimization problems such as the service-outage based power and rate allocation in [15].

Endnote

[a]A flat fading channel can be expressed as a complex-baseband model. However, according to the Appendix B.4.2 of ([19], pp. 527–528), one way to derive the capacity of a complex-baseband channel is to think of each user of the complex channel as two uses of a real channel. Thus, we only need to consider a real-baseband model, i.e., (1), and then multiply $1/2$ at the maximum mutual information with unit nats per real dimension, as in (2).

Appendix 1
Proof of Lemma 1

Proof. The proof is similar to [9,12]. For all $j, 1 \leq j \leq L - 1$, $\mathbf{P}(\mathbf{h}) = \mathbf{P}_j$, if $\mathbf{h} \in \mathcal{R}_j$, let $\mathcal{R}_j^*$ be the set of all $\mathbf{h}$ such that $\mathbf{P}_{j+1}^{\Sigma} < (\mathbf{P}^*(\mathbf{h}))^{\Sigma} \leq \mathbf{P}_j^{\Sigma}$, we need to prove that $\mathcal{R}_j^* = \mathcal{R}_j$. Assume the contrary, that $\mathcal{R}_j^* \backslash \mathcal{R}_j$ is a

non empty set ($\backslash$ denotes the set subtraction operation), i.e., if $\mathbf{h} \in \mathcal{R}_j^* \backslash \mathcal{R}_j$, then $\mathbf{h} \in \mathcal{R}_j^*$ and $\mathbf{h} \notin \mathcal{R}_j$. And we can partition the set $\mathcal{R}_j^* \backslash \mathcal{R}_j$ into two subsets $\mathcal{R}_j^- = \mathcal{R}_j^* \backslash \mathcal{R}_j \bigcap \sum_{k=1}^{j-1} \mathcal{R}_k$ and $\mathcal{R}_j^+ = \mathcal{R}_j^* \backslash \mathcal{R}_j \bigcap \sum_{k=j+1}^{L} \mathcal{R}_k$. If the set $\mathcal{R}_j^-$ has nonzero probability, then we can construct a new scheme by assigning all elements of this set to $\mathcal{R}_j$ instead. Since $\forall \mathbf{h} \in \mathcal{R}_j^-, \mathbf{P}^*(\mathbf{h})^{\Sigma} \leq \mathbf{P}_j^{\Sigma}$, such rearrangement achieves the same outage probability but with less average power due to $\mathbf{P}_j^{\Sigma} < \mathbf{P}_k^{\Sigma}, 1 \leq k \leq j - 1$, which is in contradiction with the optimality of the optimal solution $\mathcal{P}$ and $\mathcal{R}$. On the other hand, the set $\mathcal{R}_j^+$ is also an empty set, otherwise, we can easily see that this set is in outage (since $\forall \mathbf{h} \in \mathcal{R}_j^+, (\mathbf{P}^*(\mathbf{h}))^{\Sigma} > \mathbf{P}_{j+1}^{\Sigma}$), thus we have larger overall outage probability, which is also a contradiction. Therefore, we have $\mathcal{R}_j^* \subseteq \mathcal{R}_j$. We also can similarly prove $\mathcal{R}_L^* \subseteq \mathcal{R}_L$ as [9] did, omitted due to space limit. Since $\bigcup_{j=1}^{L} \mathcal{R}_j^* = \bigcup_{j=1}^{L} \mathcal{R}_j$, we can conclude that $\mathcal{R}_j^* = \mathcal{R}_j, \forall j$. □

Appendix 2
Proof of Theorem 1

Proof. We introduce μ as the Lagrange multiplier associated with the average power constraint. The Lagrangian can be written as

$$J(\mathbf{P}, \mu) = F'(\mathbf{P}_1) + \mu \left[\sum_{j=1}^{L-1} \left(\sum_{i=1}^{M} p_{ij} \right) (F'(\mathbf{P}_{j+1}) - F'(\mathbf{P}_j)) \right.$$

$$\left. + \left(\sum_{i=1}^{M} p_{iL} \right) (1 - F'(\mathbf{P}_L) + F'(\mathbf{P}_1)) - M P_{\text{av}} \right].$$
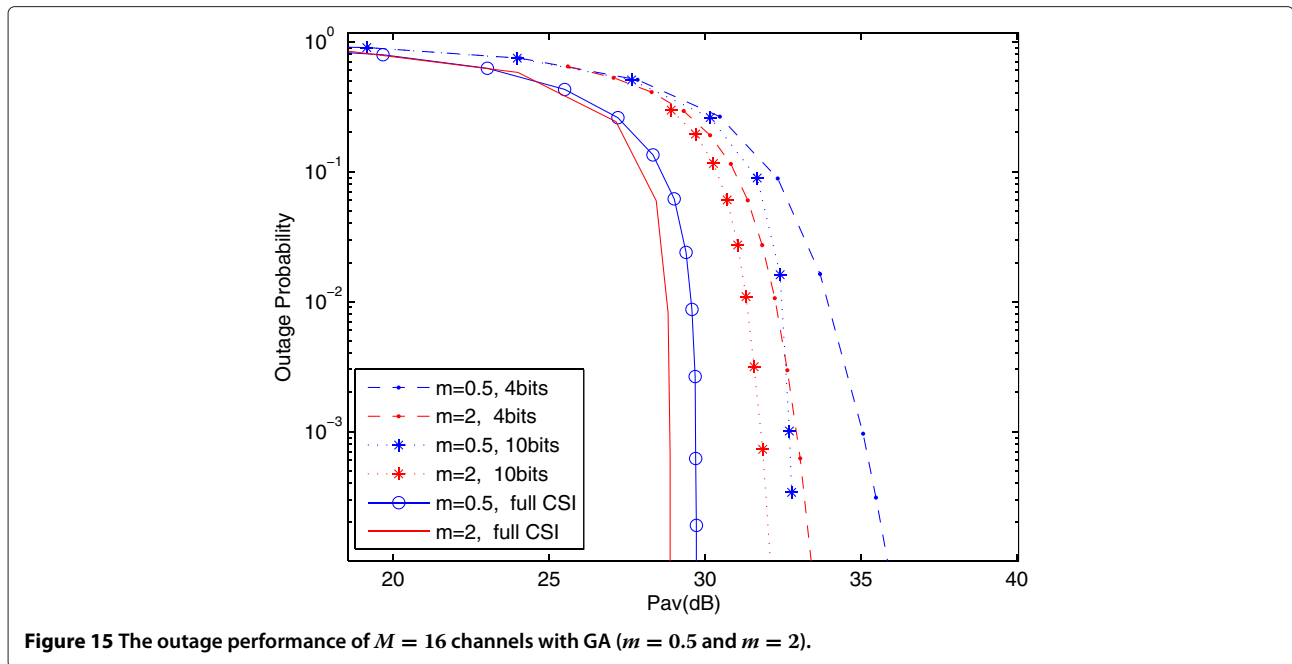
$$(51)$$

Figure 15 The outage performance of $M = 16$ channels with GA ($m = 0.5$ and $m = 2$).

Setting the first-order partial derivatives to zero (i.e., $\frac{\partial J}{\partial p_{ij}*} = 0, i = 1, \ldots, M$, for $j = 1, \ldots, L$), produces

$$\frac{\partial J}{\partial p_{i1}*} = \left(1 - \mu \sum_{i=1}^{M} p_{i1} + \mu \sum_{i=1}^{M} p_{iL}\right) \frac{\partial F'(\mathbf{P}_1)}{\partial p_{i1}*}$$

$$+ \mu(F'(\mathbf{P}_2) - F'(\mathbf{P}_1)) = 0,$$

$$\frac{\partial J}{\partial p_{ij}*} = \mu \left[\left(\sum_{i=1}^{M} p_{i,j-1} - \sum_{i=1}^{M} p_{ij}\right) \frac{\partial F'(\mathbf{P}_j)}{\partial p_{ij}*} + \left(F'(\mathbf{P}_{j+1})\right. \right.$$

$$\left. \left. - F'(\mathbf{P}_j)\right) \right] = 0, \quad 2 \leq j \leq L - 1$$

$$\frac{\partial J}{\partial p_{iL}*} = \mu \left[\left(\sum_{i=1}^{M} p_{i,L-1} - \sum_{i=1}^{M} p_{iL}\right) \frac{\partial F'(\mathbf{P}_L)}{\partial p_{ij}*} \right.$$

$$\left. + (1 - F'(\mathbf{P}_L) + F'(\mathbf{P}_1)) \right] = 0. \quad (52)$$

Since $\mu \neq 0$ (note that otherwise $\frac{\partial F'(\mathbf{P}_1)}{\partial p_{i1}*} = 0, i = 1, \ldots, M$ and since $F'(\mathbf{P}_1)$ is monotonically decreasing with p_{i1}, $\frac{\partial F'(\mathbf{P}_1)}{\partial p_{i1}*} = 0$ implies $p_{i1}* = \infty$, corresponding to infinite average power which is impossible), $F'(\mathbf{P}_2) \neq$

$F'(\mathbf{P}_1)$, thus $(1 - \mu \sum_{i=1}^{M} p_{i1} + \mu \sum_{i=1}^{M} p_{iL})) \neq 0$. Simplifying (52), we have,

$$\frac{\partial F'(\mathbf{P}_1)}{\partial p_{11}*} = \cdots = \frac{\partial F'(\mathbf{P}_1)}{\partial p_{M1}*} = -\frac{\mu(F'(\mathbf{P}_2) - F'(\mathbf{P}_1))}{1 - \mu \sum_{i=1}^{M}(p_{i1} - p_{iL})},$$

$$\frac{\partial F'(\mathbf{P}_j)}{\partial p_{1j}*} = \cdots = \frac{\partial F'(\mathbf{P}_j)}{\partial p_{Mj}*} = -\frac{F'(\mathbf{P}_{j+1}) - F'(\mathbf{P}_j)}{\sum_{i=1}^{M}(p_{i,j-1} - p_{ij})},$$

$$2 \leq j \leq L - 1,$$

$$\frac{\partial F'(\mathbf{P}_L)}{\partial p_{1L}*} = \cdots = \frac{\partial F'(\mathbf{P}_L)}{\partial p_{ML}*} = -\frac{1 - F'(\mathbf{P}_L) + F'(\mathbf{P}_1)}{\sum_{i=1}^{M}(p_{i,L-1} - p_{iL})}.$$

$$(53)$$

Therefore, finally, we have

$$\frac{\partial F'(\mathbf{P}_j)}{\partial p_{1j}*} = \cdots = \frac{\partial F'(\mathbf{P}_j)}{\partial p_{Mj}*}, j = 1, \ldots, L. \quad (54)$$

This completes the proof. □

Appendix 3
Proof of Theorem 2

Proof. In the multiple infinite series representation (19), for a sufficiently high P_{av}, we have $\left|-\frac{m_i \lambda_i K}{p_{ij}}\right| < 1$, $\forall i, j$. Thus from [30], the conditions of the convergence of the power series (20) are satisfied.

From (19), we have

$$F'(\mathbf{P}_j) = \frac{1}{\Gamma\left(1 + \sum_{i=1}^M m_i\right)} \left[\prod_{i=1}^M \left(\frac{m_i \lambda_i K}{p_{ij}}\right)^{m_i}\right]$$

$$\times \sum_{n_1=0}^\infty \cdots \sum_{n_M=0}^\infty \frac{\left[\prod_{i=1}^M (m_i)_{n_i}\left(-\frac{m_i \lambda_i K}{p_{ij}}\right)^{n_i}\frac{1}{n_i!}\right]}{\left(1 + \sum_{i=1}^M m_i\right)_{n_\tau}}$$

$$= \frac{1}{\Gamma\left(1 + \sum_{i=1}^M m_i\right)} \left[\prod_{i=1}^M \left(\frac{m_i \lambda_i K}{p_{ij}}\right)^{m_i}\right]$$

$$\times \left(1 + \sum_{\left\{n_1,\ldots,n_M \in \mathbb{Z} \mid \sum_{i=1}^M n_i = 1\right\}} T(n_1,\ldots,n_M)\right.$$

$$\times \frac{1}{\prod_{i=1}^M p_{ij}^{n_i}} + \cdots + \sum_{\left\{n_1,\ldots,n_M \in \mathbb{Z} \mid \sum_{i=1}^M n_i = \infty\right\}}$$

$$\left. \times T(n_1,\ldots,n_M) \frac{1}{\prod_{i=1}^M p_{ij}^{n_i}}\right) \quad (55)$$

where $T(n_1,\ldots,n_M) = \frac{\prod_{i=1}^M (m_i)_{n_i}(-m_i \lambda_i K)^{n_i}\frac{1}{n_i!}}{\left(1 + \sum_{i=1}^M m_i\right)_{n_\tau}}$ and $\mathbb{Z}$ is the
set of non-negative integers.

Since in the high P_{av} regime, using the approximation
$\frac{1}{\prod_{i=1}^M p_{ij}^{n_i}} \approx 0$ for $\sum_{i=1}^M n_i \geq 1$, we have

$$F'(\mathbf{P}_j) \approx \frac{1}{\Gamma\left(1 + \sum_{i=1}^M m_i\right)} \left[\prod_{i=1}^M \left(\frac{m_i \lambda_i K}{p_{ij}}\right)^{m_i}\right]. \quad (56)$$

From (56), we have, for $i = 1,\ldots,M$

$$\frac{\partial F'(\mathbf{P}_j)}{\partial p_{ij}} \approx -\frac{m_i}{p_{ij}}\frac{1}{\Gamma\left(1 + \sum_{i=1}^M m_i\right)}\prod_{i=1}^M \left(\frac{m_i \lambda_i K}{p_{ij}}\right)^{m_i}. \quad (57)$$

Finally, by substituting (57) in Theorem 1, we have

$$\frac{m_i}{p_{ij}^*} \approx \frac{m_l}{p_{lj}^*}, \quad i, l \in [1, M], i \neq l, j = 1,\ldots,L \quad (58)$$

which completes the proof of Theorem 2. □

Appendix 4
Proof of Lemma 2

Proof. In [24], Moschopoulos justified the uniform convergence of the single infinite series in (21). With PFPPC, (21) can be rewritten as

$$F'(p_j) = \prod_{i=1}^M \left(\frac{\lambda_i}{\max(\lambda_i)}\right)^{m_i} \sum_{n=0}^\infty \frac{\delta_n \gamma\left(\rho + n, \frac{Km_M \max(\lambda_i)}{p_j}\right)}{\Gamma(\rho + n)} \quad (59)$$

where $\rho = \sum_{i=1}^M m_i \geq 0.5$ ($M \geq 1, m_i \geq 0.5$) and the coefficients δ_n are obtained recursively by

$$\delta_{n+1} = \frac{1}{n+1}\sum_{l=1}^{n+1}\left[\delta_{n+1-l}\sum_{i=1}^M m_i\left(1 - \frac{\lambda_i}{\max(\lambda_i)}\right)^l\right],$$

$$\delta_0 = 1, \quad n = 0, 1, \ldots. \quad (60)$$

In high the P_{av} regime ($P_{av} \to \infty$), $r_1 = \frac{K}{p_1} \to 0$, then with PFPPC, the total average power allocated to the outage region $\mathcal{R}_{L,1}$ is

$$\lim_{r_1 \to 0} P_{tot}^{L,1} = \lim_{r_1 \to 0} p_L F'(p_1). \quad (61)$$

We have

$$\lim_{r_1 \to 0} p_L F'(p_1) = \lim_{r_1 \to 0} \frac{p_L}{p_1} p_1 F'(p_1)$$

$$= \lim_{r_1 \to 0} \frac{r_1}{r_L} K \frac{\prod_{i=1}^M \left(\frac{\lambda_i}{\max(\lambda_i)}\right)^{m_i}\sum_{n=0}^\infty \frac{\delta_n \gamma(\rho+n,r_1 m_M \max(\lambda_i))}{\Gamma(\rho+n)}}{r_1}$$

$$= \lim_{r_1 \to 0} \frac{r_1}{r_L} C \sum_{n=0}^\infty \frac{\delta_n (r_1 m_M \max(\lambda_i))^{\rho+n-1}e^{-r_1 m_M \max(\lambda_i)}}{\Gamma(\rho+n)}$$

$$= \lim_{r_1 \to 0} \frac{r_1^\rho}{r_L} C', \quad (62)$$

where $C = Km_M \max(\lambda_i)\prod_{i=1}^M \left(\frac{\lambda_i}{\max(\lambda_i)}\right)^{m_i}$ and $C' = C\frac{(m_M \max(\lambda_i))^{\rho-1}}{\Gamma(\rho)}$, and the last equality follows from the fact that when $n \geq 1$, the individual terms go to zero for any ρ, as $r_1 \to 0$.

From the proof of Theorem 3 (see below), we have

$$p_1 \approx \frac{P'_{av}{}^{\rho^{L-1}+\cdots+\rho+1}}{L^{\rho^{L-1}+\cdots+\rho+1}c^{\rho^{L-2}+\cdots+\rho+1}} = \frac{P'_{av}{}^{\sum_{i=0}^{L-1}\rho^i}}{C_1} \quad (63)$$

where $C_1 = L^{\rho^{L-1}+\cdots+\rho+1}c^{\rho^{L-2}+\cdots+\rho+1}$. Thus, we have

$$\lim_{r_1 \to 0} \frac{r_1^\rho}{r_L} C' = \lim_{P_{av} \to \infty} \frac{p_L}{p_1^\rho}C'K^{\rho-1} \approx \lim_{P_{av} \to \infty} p_L \frac{C_1^\rho C'K^{\rho-1}}{P'_{av}{}^{\sum_{i=1}^L \rho^i}}. \quad (64)$$

Since from the proof of Theorem 3, we have $p_L \approx \frac{P'_{av}}{L}$,

$$\lim_{r_1 \to 0} \frac{r_1^\rho}{r_L}C' \approx \lim_{P_{av} \to \infty} \frac{C_1^\rho C'K^{\rho-1}}{LP'_{av}{}^{\left(\sum_{i=1}^L \rho^i\right)-1}}. \quad (65)$$

Thus, if $\left(\sum_{i=1}^L \rho^i\right) - 1 > 0$, we have

$$\lim_{r_1 \to 0} \frac{r_1^\rho}{r_L}C' = 0. \quad (66)$$

This completes the proof of Lemma 2. □

Appendix 5
Proof of Theorem 3

Proof. In the high P_{av} regime, from Theorem 2, we have

$$F'(\mathbf{P}_j) \approx \frac{c}{p_j^{\sum_{i=1}^{M} m_i}} = \frac{c}{p_j^{\rho}}, \quad j = 1, \ldots, L, \tag{67}$$

since $p_{ij} = \frac{m_i}{m_M} p_{Mj} = \frac{m_i}{m_M} p_j$ using the same notation as in Section 'High average power approximation (HP_{av}A)'. Here $c = \frac{\prod_{i=1}^{M}(m_M \lambda_i K)^{m_i}}{\Gamma(1+\rho)}$ where $\rho = \sum_{i=1}^{M} m_i$. When $P_{av} \to \infty$, according to [13], all the quantization thresholds approach zero, thus the length between any two quantization thresholds approaches zero as well resulting in the property that the total average power assigned to each quantization region is asymptotically equal. Thus, we have

$$p_j(F'(p_{j+1}) - F'(p_j)) = \frac{P'_{av}}{L}, \quad j = 1, \ldots, L-1$$
$$p_L(1 - F'(p_L) + F'(p_1)) = \frac{P'_{av}}{L} \tag{68}$$

where $P'_{av} = \frac{Mm_M}{\rho} P_{av}$. Applying (67) to (68), we have for $j = 1, \ldots, L-1$,

$$p_j \left(\frac{c}{p_{j+1}^{\rho}} - \frac{c}{p_j^{\rho}} \right) \approx \frac{P'_{av}}{L}$$
$$\frac{1}{p_{j+1}^{\rho}} \approx \frac{\left(\frac{P'_{av}}{Lc} \right)}{p_j} + \frac{1}{p_j^{\rho}}. \tag{69}$$

It is clear to deduce that in the high P_{av} regime, for $M \geq 2$ and $M = 1$, $m \geq 1$, $\frac{1}{p_j^{\rho}}$ compared to $\frac{\left(\frac{P'_{av}}{Lc} \right)}{p_j}$ is negligible, thus (69) can be written as,

$$\frac{1}{p_{j+1}^{\rho}} \approx \frac{\left(\frac{P'_{av}}{Lc} \right)}{p_j}$$
$$p_j \approx p_{j+1}^{\rho} \frac{P'_{av}}{Lc}. \tag{70}$$

The same approximation can be shown to hold true for $M = 1$, $0.5 \leq m < 1$ by contradiction, details are omitted due to space restrictions. Thus, we have,

$$p_1 \approx p_2^{\rho} \frac{P'_{av}}{Lc} \approx \left(p_3^{\rho} \frac{P'_{av}}{Lc} \right)^{\rho} \frac{P'_{av}}{Lc} = p_3^{\rho^2} \left(\frac{P'_{av}}{Lc} \right)^{\rho+1}$$

$$\approx \cdots \approx p_L^{\rho^{L-1}} \left(\frac{P'_{av}}{Lc} \right)^{\rho^{L-2}+\cdots+\rho+1}. \tag{71}$$

Since $\lim_{P_{av} \to \infty} F'(p_1) = 0$, from (68) we have,

$$p_L(1 - F'(p_L)) = \frac{P'_{av}}{L}$$
$$p_L \left(1 - \frac{c}{p_L^{\rho}} \right) \approx \frac{P'_{av}}{L}. \tag{72}$$

Note that $\frac{c}{p_L^{\rho}}$ is negligible ($\to 0$) when P_{av} go to infinity, thus, we have,

$$p_L \approx \frac{P'_{av}}{L}. \tag{73}$$

Applying (73) to (71), we have

$$p_1 \approx \left(\frac{P'_{av}}{L} \right)^{\rho^{L-1}} \left(\frac{P'_{av}}{Lc} \right)^{\rho^{L-2}+\cdots+\rho+1}$$

$$= \frac{P'_{av}{}^{\rho^{L-1}+\cdots+\rho+1}}{L^{\rho^{L-1}+\cdots+\rho+1} c^{\rho^{L-2}+\cdots+\rho+1}}. \tag{74}$$

Since,

$$P_{out} = F'(\mathbf{P}_1) \approx \frac{c}{p_1^{\rho}} \approx c \left(\frac{L^{\rho^{L-1}+\cdots+\rho+1} c^{\rho^{L-2}+\cdots+\rho+1}}{P'_{av}{}^{\rho^{L-1}+\cdots+\rho+1}} \right)^{\rho}$$

$$= \frac{L^{\rho^L+\cdots+\rho} c^{\rho^{L-1}+\cdots+\rho+1}}{P'_{av}{}^{\rho^L+\cdots+\rho}}$$

$$= \frac{\left(\frac{\rho L}{Mm_M} \right)^{\rho^L+\cdots+\rho} c^{\rho^{L-1}+\cdots+\rho+1}}{P_{av}{}^{\rho^L+\cdots+\rho}}, \tag{75}$$

we have,

$$d = -\lim_{P_{av} \to \infty} \frac{\log P_{out}}{\log P_{av}}$$

$$\approx -\lim_{P_{av} \to \infty} \frac{\log \left(\left(\frac{\rho L}{Mm_M} \right)^{\rho^L+\cdots+\rho} c^{\rho^{L-1}+\cdots+\rho+1} \right)}{\log P_{av}}$$

$$+ \lim_{P_{av} \to \infty} \frac{\log P_{av}{}^{\rho^L+\cdots+\rho}}{\log P_{av}}$$

$$= 0 + \rho^L + \cdots + \rho = \sum_{j=1}^{L} \rho^j. \tag{76}$$

This completes the proof of Theorem 3. □

Appendix 6
Proof that $\{z_i - E[z_i]\}$ satisfies the Lindeberg condition

Proof. Given a random variable $f_i \sim Gamma(m_i, \beta_i)$ (where m_i is the fading parameter satisfying $\frac{1}{2} \leq m_i < \infty$) and $E[f_i] = m_i \beta_i = 1$, or $\beta_i = \frac{1}{m_i}$, we have $\mathrm{Var}(f_i) = m_i \beta_i^2 = \beta_i$. So $0 < \mathrm{Var}(f_i) \leq 2 < \infty$.

Let $z_i = \log(f_i)$, thus the moment-generating function of random variable z_i is

$$M_{z_i}(t) = E(e^{tz_i}) = E(e^{t\log(f_i)}) = E((f_i)^t)$$

$$= \int_0^\infty f_i^t \frac{f_i^{m_i-1}}{\Gamma(m_i)\beta_i^{m_i}} e^{-\frac{f_i}{\beta_i}} df_i$$

$$= \frac{\Gamma(m_i+t)\beta_i^{m_i+t}}{\Gamma(m_i)\beta_i^{m_i}} \int_0^\infty \frac{f_i^{t+m_i-1}}{\Gamma(m_i+t)\beta_i^{m_i+t}} e^{-\frac{f_i}{\beta_i}} df_i$$

$$= \frac{\Gamma(m_i+t)\beta_i^{m_i+t}}{\Gamma(m_i)\beta_i^{m_i}}$$

$$= \frac{\Gamma(m_i+t)\beta_i^t}{\Gamma(m_i)}. \tag{77}$$

Then,

$$E(z_i) = M_{z_i}^{(1)}(0)$$

$$= \frac{d\left(\frac{\Gamma(m_i+t)\beta_i^t}{\Gamma(m_i)}\right)}{dt}(0)$$

$$= \left(\frac{\Gamma(m_i+t)'\beta_i^t}{\Gamma(m_i)} + \frac{\Gamma(m_i+t)\beta_i^t\log(\beta_i)}{\Gamma(m_i)}\right)(0)$$

$$= \frac{\Gamma(m_i+t)'(0)}{\Gamma(m_i)} + \log(\beta_i)$$

$$= \frac{(\psi(m_i+t)\Gamma(m_i+t))(0)}{\Gamma(m_i)} + \log(\beta_i)$$

$$= \psi(m_i) + \log(\beta_i), \tag{78}$$

where $\psi(m_i)$ denotes the digamma function [31]. Similarly,

$$E(z_i^2) = M_{z_i}^{(2)}(0)$$

$$= \frac{d^2\left(\frac{\Gamma(m_i+t)\beta_i^t}{\Gamma(m_i)}\right)}{dt^2}(0)$$

$$= \left(\frac{\Gamma(m_i+t)''\beta_i^t}{\Gamma(m_i)} + 2\frac{\Gamma(m_i+t)'\beta_i^t\log(\beta_i)}{\Gamma(m_i)}\right.$$

$$\left. + \frac{\Gamma(m_i+t)\beta_i^t\log^2(\beta_i)}{\Gamma(m_i)}\right)(0)$$

$$= \frac{\Gamma(m_i+t)''(0)}{\Gamma(m_i)} + 2\frac{\Gamma(m_i+t)'(0)\log(\beta_i)}{\Gamma(m_i)} + \log^2(\beta_i)$$

$$= \frac{\Gamma(m_i+t)''(0)}{\Gamma(m_i)} + 2\frac{\Gamma(m_i+t)(0)\log(\beta_i)}{\Gamma(m_i)} + \log^2(\beta_i)$$

$$= \frac{\Gamma(m_i+t)''(0)}{\Gamma(m_i)} + 2\psi(m_i)\log(\beta_i) + \log^2(\beta_i)$$

$$= \frac{(\psi(m_i+t)'\Gamma(m_i+t) + \psi(m_i+t)\Gamma(m_i+t)')(0)}{\Gamma(m_i)}$$

$$+ 2\psi(m_i)\log(\beta_i) + \log^2(\beta_i)$$

$$= \frac{(\psi_1(m_i+t)\Gamma(m_i+t) + \psi(m_i+t)^2\Gamma(m_i+t))(0)}{\Gamma(m_i)}$$

$$+ 2\psi(m_i)\log(\beta_i) + \log^2(\beta_i)$$

$$= \psi_1(m_i) + \psi(m_i)^2 + 2\psi(m_i)\log(\beta_i) + \log^2(\beta_i)$$

$$= \psi_1(m_i) + (\psi(m_i) + \log(\beta_i))^2, \tag{79}$$

where $\psi_1(m_i)$ denotes the trigamma function [31].

Let $\sigma_i^2 = \text{Var}(z_i)$, then we have

$$\sigma_i^2 = E(z_i^2) - (E(z_i))^2$$

$$= \psi_1(m_i) + (\psi(m_i) + \log(\beta_i))^2 - (\psi(m_i) + \log(\beta_i))^2$$

$$= \psi_1(m_i)$$

$$= \sum_{n=0}^\infty \frac{1}{(m_i+n)^2}. \tag{80}$$

Since $\frac{1}{2} \le m_i < \infty$, we can obtain,

$$0 < \frac{1}{m_i^2} < \sigma_i^2 = \sum_{n=0}^\infty \frac{1}{(m_i+n)^2} \le \psi_1\left(\frac{1}{2}\right) = \frac{\pi^2}{2} < \infty. \tag{81}$$

Since $m_i < \infty$, there exists a large enough constant $\bar{C}$ so that $m_i \le \bar{C} < \infty$. From (81), we have

$$0 < \frac{1}{\bar{C}^2} < \sigma_i^2 \le \frac{\pi^2}{2} < \infty. \tag{82}$$

Let $X_i = z_i - E(z_i)$, $S_n = \sum_{i=1}^n X_i$, $\sigma_n^2 = \text{Var}(S_n) = \sum_{i=1}^n \sigma_i^2$, then from (82) we have

$$\sigma_n^2 = \sum_{i=1}^n \sigma_i^2 > n\frac{1}{\bar{C}^2}. \tag{83}$$

Thus, when $n \to \infty$, we have

$$\sigma_n^2 \to \infty. \tag{84}$$

For every $\epsilon > 0$, using Chebyshev's Inequality, we can obtain,

$$P(|X_i| > \epsilon\sigma_n) \le \frac{\sigma_i^2}{\epsilon^2\sigma_n^2} \le \frac{max\left(\sigma_i^2\right)}{\epsilon^2\sigma_n^2} = \frac{\frac{\pi^2}{2}}{\epsilon^2\sigma_n^2} \to 0 \quad as$$

$$n \to \infty. \tag{85}$$

Thus,

$$\lim_{n\to\infty} \frac{1}{\sigma_n^2} \sum_{i=1}^{n} E\left(X_i^2 I\{|X_i| > \epsilon\sigma_n\}\right)$$

$$= \lim_{n\to\infty} \frac{1}{\sigma_n^2} \sum_{i=1}^{n} \sigma_i^2 \frac{1}{\sigma_i^2} E(X_i^2 I\{|X_i| > \epsilon\sigma_n\})$$

$$\le \lim_{n\to\infty} \max\left(\frac{1}{\sigma_i^2} E(X_i^2 I\{|X_i| > \epsilon\sigma_n\})\right) \frac{1}{\sigma_n^2} \sum_{i=1}^{n} \sigma_i^2$$

$$= \lim_{n\to\infty} \max\left(\frac{1}{\sigma_i^2} E(X_i^2 I\{|X_i| > \epsilon\sigma_n\})\right)$$

$$\le \lim_{n\to\infty} \bar{C}^2 \max\left(E(X_i^2 I\{|X_i| > \epsilon\sigma_n\})\right) \tag{86}$$

where $I(A)$ denotes the indicator function taking the value 1 if the event A is true and taking the value 0 otherwise. Let Y_n denotes the random variable $X_i^2 I\{|X_i| > \epsilon\sigma_n\}$, then Y_n is nonzero if and only if $|X_i| > \epsilon\sigma_n$. Since from (85), we know that this event has probability approaching zero as $n \to \infty$, we can also conclude that $Y_n \xrightarrow{\mathcal{P}} 0$, where $\xrightarrow{\mathcal{P}}$ denotes convergence in probability. Since $|Y_n| \le X_i^2$ and $E(X_i^2) = \sigma_i^2 < \infty$, by applying the Dominated Convergence Theorem, we can conclude that $E(Y_n) \to 0$, namely,

$$E(X_i^2 I\{|X_i| > \epsilon\sigma_n\}) \to 0. \tag{87}$$

Applying the above result into (86), we have

$$\lim_{n\to\infty} \frac{1}{\sigma_n^2} \sum_{i=1}^{n} E(X_i^2 I\{|X_i| > \epsilon\sigma_n\}) = 0. \tag{88}$$

Thus, the Lindeberg condition holds, and X_i satisfies the Central limit theorem. □

Competing interests
The authors declare that they have no competing interests.

References
1. AJ Goldsmith, P Varaiya, Capacity of fading channels with channel side information. IEEE Trans. Inf. Theory. **43**, 1986–1992 (1997)
2. SV Hanly, DNC Tse, Multi-access fading channels-part II: delay-limited capacities. IEEE Trans. Inf. Theory. **44**, 2816–2831 (1998)
3. G Caire, G Taricco, E Biglieri, Optimum power control over fading channels. IEEE Trans. Inf. Theory. **45**(5), 1468–1489 (1999)
4. DJ Love, RW Heath, Grassmanian beamforming for multiple-input multiple-output wireless systems. IEEE Trans. Inf. Theory. **49**(10), 2735–2747 (2003)
5. KK Mukkavilli, A Sabharwal, E Erkip, B Aazhang, On beamforming with finite rate feedback in multiple-antenna systems. IEEE Trans. Inf. Theory. **49**(10), 2563–2579 (2003)
6. TT Kim, M Skoglund, On the expected rate of slowly fading channels with quantized side information. IEEE Trans. Commun. **55**(4), 820–829 (2007)
7. A Feiten, R Mathar, S Hanly, Eigenvalue-based optimum-power allocation for gaussian vector channels. IEEE Trans. Inf. Theory. **53**(6), 2304–2309 (2007)
8. TT Kim, M Skoglund, Diversity-multiplexing tradeoff in MIMO channels with partial CSIT. IEEE Trans. Inf. Theory. **53**(8), 2743–2759 (2007)
9. TT Kim, Skoglund M, in *Proc. IEEE International Conference on Communication (ICC'06)*, Partial power control for slowly fading MIMO channels, vol. 3, (Istanbul, Turkey, 2006), pp. 1362–1367
10. A Khoshnevis, A Sabharwal, in *Proc. IEEE International Conference on Communication (ICC'04)*, Performance of quantized power control in multiple antenna systems, vol. 2, (Paris, France, 2004), pp. 803–807
11. N Ahmed, MA Khojastepour, A Sabharwal, B Aazhang, Outage minimization with limited feedback for the fading relay channel. IEEE Trans. Commun. **54**(4), 659–669 (2006)
12. MA Khojastepour, G Yue, X Wang, M Madihian, Optimal power control in MIMO systems with quantized feedback. IEEE Trans. Wirel. Commun. **7**(12), 4859–4866 (2008)
13. A Khoshnevis, A Sabharwal, On the asymptotic performance of multiple antenna channels with quantized feedback. IEEE Trans. Wirel. Commun. **7**(10), 3869–3877 (2008)
14. I Krikidis, J Thompson, S McLaughlin, N Goertz, Optimization issues for cooperative amplify-and-forward systems over block-fading channels. IEEE Trans. Veh. Technol. **57**(5), 2868–2884 (2008)
15. JH Luo, R Yates, P Spasojević, Service outage based power and rate allocation for parallel fading channels. IEEE Trans. Inf. Theory. **51**(7), 2594–2611 (2005)
16. JN Laneman, E Martinian, GW Wornell, JG Apostolopoulos, Source-Channel Diversity for Parallel Channels. IEEE Trans. Inf. Theory. **51**(10), 3518–3539 (2005)
17. K Bakanoglu, S Tomasin, E Erkip, Resource allocation for the parallel relay channel with multiple relays. IEEE Trans. Wirel. Commun. **10**(3), 792–802 (2011)
18. W Zhang, U Mitra, On outage behavior of wideband slow-fading channels. (arXiv:0711.4557v1 [cs.IT], 2007), http://arxiv.org/pdf/0711.4557v1.pdf. Accessed on 28 November 2007
19. D Tse, P Viswanath, *Fundamentals of Wireless Communications* (Cambridge University Press, Cambridge, 2005)
20. JC Spall, Implementation of the simultaneous perturbation algorithm for stochastic optimization. IEEE Trans. Aerosp. Electron. Syst. **34**(3), 817–823 (1998)
21. PA Chou, T Lookabaugh, RM Gray, Entropy-constrained vector quantization. IEEE Trans. Acoust. Speech Signal Process. **31**(1), 31–42 (1989)
22. VA Aalo, T Piboongungon, GP Efthymoglou, Another look at the performance of MRC schemes in Nakagami-m fading channels with arbitrary parameters. IEEE Trans. Commun. **53**(12), 2002–2005 (2005)
23. S Nadarajah, Simplified expressions for the outage and error rate performance of DS-CDMA with MRC in Nakagami-m fading. Int. J. Electron. **96**(4), 365–369 (2009)
24. PG Moschopoulos, The distribution of the sum of independent gamma random variables. Annals Inst. Stat. Math. **37**(1), 541–544 (1985)
25. DP Bertsekas, *Nonlinear Programming* (Athena Scientific, Belmont, Massachusetts, 1999)
26. M Einhaus, O Klein, B Walke, D Bultmann, in *Proc. IWCMC'06*, Performance evaluation of joint downlink scheduling in multi-cellular OFDMA systems based on IEEE 802.16a (Vancouver, British Columbia, Canada, 2006), pp. 647-652
27. W Feller, *An Introduction to Probability Theory and Its Applications*, 3rd edn, vol. 2 (Wiley, New York, 1971), pp. 257–258
28. BM Hochwald, TL Marzetta, V Tarokh, Multiple-antenna channel hardening and its implications for rate feedback and scheduling. IEEE Trans. Inf. Theory. **50**(9), 1893–1909 (2004)

29. S Ekbatani, F Etemadi, H Jafarkhani, in *Proc. IEEE GLOBECOM 2007,* Outage behaviour of quasi-static fading channels with partial power control and noisy feedback Washington DC, 2007), pp. 1556–1560
30. H Exton, *Multiple Hypergeometric Functions and Applications* (Wiley, New York, 1976)
31. H Jeffreys, B Jeffreys, in *Methods of Mathematical Physics,* 3rd edn. The Digamma $\mathcal{F}$ and Trigamma $\mathcal{F}'$ Functions (Cambridge University Press, Cambridge, 1988), pp. 465–466