Asymptotic Capacity of Multi-Carrier Transmission With Frequency-Selective Fading and Limited Feedback

Yakun Sun[*]   and   Michael L. Honig[†]

∗ Formerly with the Dept. of Electrical & Computer Engineering; Northwestern University, Evanston, IL 60208. He is currently with Marvell Semiconductor Inc, 5488 Marvell Lane, Santa Clara, CA, 94085, USA; email: yakunsun@marvell.com.

† Dept. of Electrical & Computer Engineering; Northwestern University, Evanston, IL 60208; Tel/Fax: (847) 491-7803/467-3550, email: mh@ece.northwestern.edu.

Abstract

We study the capacity of multi-carrier transmission through a slow frequency-selective fading channel with limited feedback, which is used to specify channel state information. Our results are asymptotic in the number of carriers, or channels N. We first assume *i.i.d.* channel gains, and show that, for a large class of fading distributions, a uniform power distribution over an optimized subset of carriers, or *on-off* power allocation, gives the same asymptotic growth in capacity as optimal water filling, e.g., $O(\log N)$ with Rayleigh fading. Furthermore, the $O(\log N)$ growth in data rate can be achieved with a feedback rate, which grows as $O(\log^3 N)$. If the number of active channels is bounded, then the feedback rate decreases to $O(\log N)$, but the capacity grows only as $O(\log \log N)$. We then consider correlated channels modeled as a Markov process, and study the savings in feedback relative to *i.i.d.* channels. Assuming a fixed ratio of coherence bandwidth to the total available bandwidth, the ratio between minimum feedback rates with correlated and *i.i.d.* sub-channels converges to zero with N. For a sequence of Rayleigh fading channels, which satisfy a first-order autoregressive process, the ratio goes to zero as $O\left(\sqrt{\frac{\log N}{N}}\right)$. We also show that adaptive modulation, or rate control schemes, in which the transmitter selects the rate on each subchannel from a quantized set, achieves the same asymptotic growth rates in capacity and required feedback. Finally, our results are extended to cellular uplink and downlink channel models.

I. INTRODUCTION

The capacity of a slowly varying frequency-selective fading channel can generally be increased by relaying channel information from the receiver back to the transmitter. In the case of multi-carrier communications, this feedback allows the transmitter to optimize the allocation of power and data rates over the carriers, or sub-channels [1]. Given accurate channel estimates and sufficient feedback, a substantial increase in capacity can be obtained relative to the analogous scheme without feedback (e.g., Orthogonal Frequency Division Multiplexing (OFDM)).

We consider multi-carrier transmission over a frequency-selective channel with limited feedback. We assume perfect channel knowledge at the receiver, and an error-free, finite-rate feedback link. The channel is assumed to be stationary for the duration of a codeword. Our objective is to characterize the asymptotic growth in capacity as a function of the number of sub-channels N for a given feedback rate and finite transmitted power. This problem is analogous to the problem of evaluating the performance (e.g., achievable rate) of a Multi-Input/Multi-Output (MIMO) fading channel with limited feedback, which has recently attracted much interest (e.g., see [12–21]). For the OFDM channel model considered, however, the relation between the growth in capacity and feedback rate is significantly different from the capacity scaling for MIMO channels with limited feedback [16]. (See also [22], which considers quantized signatures for Code-Division Multiple Access.)

We start with an analysis of *i.i.d.* sub-channels, before considering dependent sub-channels. To limit the feedback rate, we first consider the capacity with an optimal *on-off* power allocation, in which the transmitter allocates power uniformly over an optimized set of sub-channels, i.e., with gains exceeding a threshold. Previous studies have shown that for fixed channels, the capacity achieved with this scheme is close to the capacity with the optimal (water-filling) power allocation [23–25]. Here we show that for a general class of fading distributions, both water-filling and the optimal on-off power allocation achieve the same asymptotic growth in capacity. For example, with Rayleigh

1

fading, both the optimized threshold, which maximizes the on-off capacity, and the capacity (bits per channel use) grow as $O(\log N)$. In contrast, the capacity for OFDM converges to a constant as N becomes large. Furthermore, the feedback needed to identify the optimal subset of sub-channels grows as $O(\log^3 N)$ bits per code word.

We also consider *finite-precision rate control* in which the rate for each active sub-channel is quantized, i.e., chosen from a discrete set, depending on the sub-channel gain. This is motivated, in part, by bit-loading, or adaptive modulation schemes (e.g., see [8]), in which the constellation size is varied across sub-channels. In practice, this enables the use of shorter code words to achieve a given outage probability (i.e., compared with coding across sub-channels). The optimal quantized set of sub-channel rates is determined, and it is shown that asymptotically the total data rate increases at the same rate as the capacity (e.g., $O(\log N)$ for Rayleigh fading given a feedback rate, which grows as $O(\log^3 N)$). However, as $N \to \infty$, the *absolute* loss in achievable data rate with finite-precision rate control, relative to the optimal power and rate allocation, converges to a constant, which is proportional to the transmitted power. We also consider a feedback scheme in which the number of active sub-channels is fixed at $K \le N$. In that case, the feedback rate increases as $O(\log N)$, and the capacity increases as $O(\log \log N)$. The feedback rate can be varied between $O(\log N)$ and $O(\log^3 N)$ by changing the rate at which the on-off threshold tends to infinity with N, and we characterize the corresponding asymptotic growth in capacity.

In practice, the sub-channel gains are generally correlated. Assuming that the sequence of channel gains is an ergodic process, as $N \to \infty$, the channel capacity depends only on the first-order distribution of the sub-channel gains, hence correlation among sub-channels does not affect the asymptotic growth in capacity. However, correlation among sub-channels can be exploited to reduce the amount of feedback, which is needed to achieve this growth.

An on-off power allocation is specified by a binary sequence, indicating whether the sub-channels have gains exceeding the threshold. The feedback rate, in *bits per sub-channel*, is given by the entropy rate of the sequence, which is a function of the activation threshold. We model the sequence of sub-channel gains as a Markov process. With an optimal (rate-maximizing) threshold, which increases with N, the minimum feedback (bits per subchannel) required to achieve the asymptotic capacity can be determined, and converges to zero as $N \to \infty$. The ratio between the feedback rates for correlated and *i.i.d.* sub-channels also converges to zero, and we specify the corresponding convergence rate with N. As an example, we consider a first-order autoregressive sequence of complex Gaussian sub-channels. In order to model a wireless system, which occupies a fixed total bandwidth, we constrain the number of *coherence bands*, which are contained within the N sub-channels. In that case, when the activation threshold grows as $O(\log N)$, which allows the achievable rate to increase as $O(\log N)$, the savings in feedback relative to *i.i.d.* sub-channels is $O\left(\sqrt{\frac{\log N}{N}}\right)$. With finite-precision rate control, additional bits per active sub-channel are needed to choose from a discrete set of rates. We show that adding these bits does not increase the order of the feedback rate with N.

We also extend our single-user results to multi-user uplink (multiple access) and downlink (broadcast) channel models. If the number of users K is finite, then as $N \to \infty$, the sum capacity in both scenarios simply scales as K times the single-user results. This is because the probability that multiple users want to transmit on the same sub-channel (i.e., with gain that exceeds the corresponding thresholds) goes to zero. If K also tends to infinity in

proportion with N, then we show that with Rayleigh fading the capacity per sub-channel grows as $O(\log \log K)$ with the water-filling power allocation, and with the optimal on-off power allocation with finite-precision rate control.

Related work on multi-user OFDM systems with limited feedback is presented in [26]. There it is also shown that the on-off power allocation achieves the same growth rate as the water-filling capacity. Results analogous to some of those presented here are presented in [6] for random access fading models. Namely, throughput scaling is characterized as a function of users when the random access scheme exploits multiuser diversity (analogous to transmitting on the best sub-channels in our model). That work, however, is not explicitly concerned with finite-rate feedback. Capacity scaling results for Multi-input/Multi-output (MIMO) fading channels have been presented in [2–4] without feedback, and with the optimal power allocation, corresponding to infinite-rate feedback. (See also [31], which considers the OFDM downlink with multiple transmit antennas.) Other related work on OFDM with partial channel knowledge at the transmitter has been presented in [10], [11]. Namely, [10] discusses the benefits of long-term statistical channel knowledge at the transmitter in a MIMO OFDM system, and [11] discusses power allocation in the presence of channel estimation error and the associated degradation in achievable rate.

In the next section we characterize the asymptotic capacity of a single link with water-filling and on-off power allocations, as well as the achievable rate with finite-precision rate control. The minimum feedback required for *i.i.d.* and correlated sub-channels are also compared in this section. In section III the results are extended to uplink and downlink multi-user channel models. We first consider the scenario with a finite number of users, and then an infinite number of users. Numerical results are presented in the corresponding sections. Section IV concludes the paper, and proofs of the main results are given in the appendices.

II. SINGLE-USER LINK

A. System Model and Channel Capacity

Referring to Fig. 1, which shows the system block diagram for a single-user multi-carrier link with N carriers, or sub-channels, the $N \times 1$ vector of received symbols corresponding to the mth transmitted multi-carrier symbol is given by

$$
\begin{aligned}
\mathbf{r}(m) &= \mathbf{W}\left(\mathbf{H}\mathbf{W}^{\dagger}\mathbf{s}(m) + \mathbf{n}(m)\right) \\
&= \tilde{\mathbf{H}}\mathbf{s}(m) + \tilde{\mathbf{n}}(m)
\end{aligned}
$$

where $\mathbf{s}(m) = [s_1(m), \cdots, s_N(m)]^T$ is the vector of source symbols (in the frequency domain), $\mathbf{W}$ ($\mathbf{W}^{\dagger}$) is the DFT (IDFT) matrix, $\mathbf{H}$ is the time-domain channel matrix, assumed to be circulant, and $\tilde{\mathbf{H}}$ is the associated diagonal channel matrix with (random) diagonal elements $h_1, \cdots, h_N$. The noise $\mathbf{n}(m)$ is circularly symmetric, white Gaussian with variance $\mathbf{I}_N$, hence the noise $\tilde{\mathbf{n}}(m)$ is also white Gaussian. The total transmit power is constrained to be at most $\mathcal{P}$, i.e., trace$\{E[\mathbf{s}\mathbf{s}^{\dagger}]\} \leq \mathcal{P}$, where $(\cdot)^{\dagger}$ is Hermitian transpose. The transmit power on sub-channel i is $P_i = E[|s_i|^2]$, so that $\sum_{i=1}^{N} P_i \leq \mathcal{P}$.
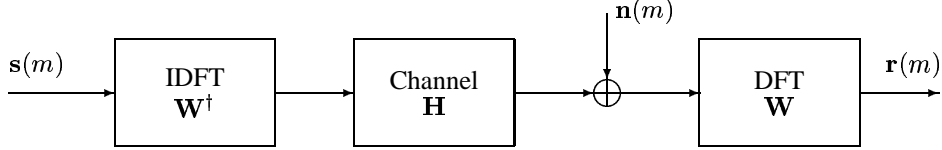
Fig. 1. System diagram of a single-user multi-carrier link.

Conditioned on the sub-channel gains, the channel capacity is given by [27]

$$C_N = \sum_{i=1}^{N} \log(1 + P_i \mu_i) \tag{1}$$

where $\mu_i = |h_i|^2$. We assume that the μ_i's are identically distributed, and denote the probability density function (pdf) of μ_i as $f_\mu(\cdot)$. Furthermore, we assume that the channel is static, and is known perfectly at the receiver, which has access to an error-free, finite-rate feedback channel. This model approximates a slowly-varying channel in which the transmitter codes across sub-channels, and the feedback specifies the power allocation and code rate during a coherence time.*

We also consider quantized rate control, which applies to the scenario in which the information bits are partitioned into sub-streams for the sub-channels, and are coded independently. For example, this corresponds to adapting the modulation format across sub-channels, as discussed in [8]. The rate for each sub-channel is then selected from a finite set of rates (perhaps only one), and the total rate is the sum of the rates across sub-channels. With this scheme an outage on one sub-channel (e.g., due to a finite-length codeword) does not affect the other sub-channels. This may be attractive for delay-sensitive applications.

Suppose that no feedback is available, so that the transmitter has no knowledge of the random sub-channel gains. In that case, the transmitter should spread the available power uniformly over all sub-channels, as in OFDM, i.e., $P_i = \frac{P}{N}$. As the number of sub-channels (or bandwidth) $N \to \infty$, $C_N \to \mathcal{P} E[\mu]$ nats, which is finite. In contrast, with unlimited feedback, the transmitter can obtain perfect channel knowledge, and the capacity with the optimal water-filling power allocation is

$$C_N^{(\text{wf})} = \sum_{i=1}^{N} \log\left(1 + (\lambda - \frac{1}{\mu_i})^+ \mu_i\right) \tag{2}$$

where the water level λ is determined by

$$\mathcal{P} = \sum_{i=1}^{N} (\lambda - \frac{1}{\mu_i})^+ \to N \int_0^\infty (\lambda - \frac{1}{x})^+ f_\mu(x) dx \tag{3}$$

as $N \to \infty$, with probability one (w.p.1).

---

*For the asymptotic (large N) analysis, we implicitly assume that the number of coherence bands increases linearly with N. Justification for this assumption for indoor wideband channels is given by the measurement study in [9].

Let $C_{N;N_a}$ denote the capacity subject to the constraint that no more than a finite number of sub-channels, N_a, are activated (i.e., $P_i > 0$). The following Lemma states that as $N \to \infty$, the optimal number of activated sub-channels $N_a \to \infty$.

Lemma 1: If the p.d.f. $f_\mu(\cdot)$ has infinite support, then for any finite N_a, there exists an $N_0 > 0$ and $M > N_a$ such that for all $N > N_0$, $C_{N;N_a}$ is achieved by activating N_a sub-channels, and $C_{N;N_a} < C_{N;M}$.

Proof: To achieve $C_{N;N_a}$, we must clearly activate a subset of N_a sub-channels with the N_a largest channel gains, denoted as $\mu_N^{(1)} \geq \cdots \geq \mu_N^{(N_a)}$. Let

$$\tilde{C}_n = \sum_{i=1}^{n} \log(1 + P_i \mu_N^{(i)}) \tag{4}$$

where $1 \leq n \leq N_a$ and $P_i > 0$ for $i = 1, \cdots, n$, so that $C_{N;N_a} = \max_n \tilde{C}_n$. As $N \to \infty$, $\mu_N^{(N_a)} \to \infty$. hence from (4) we have that

$$\tilde{C}_n = \sum_{i=1}^{n} \log\left(P_i \mu_N^{(i)}\right) + o(1) = \underbrace{\sum_{i=1}^{n} \log \mu_N^{(i)}}_{\to \infty} + \underbrace{\sum_{i=1}^{n} \log P_i}_{< \infty} + o(1). \tag{5}$$

The second term on the far right must be bounded since as $N \to \infty$, we can assume that each $P_i > \epsilon > 0$. This is because letting $P_i \to 0$ decreases the capacity on sub-channel i, and allocating the extra power to other sub-channels does not increase the asymptotic capacity, since the first channel-gain term dominates. Therefore as $N \to \infty$, $\tilde{C}_n$ increases with n, so that *all* of the best N_a sub-channels should be activated. This also implies that $C_{N;N_a}$ increases with N_a for large enough N. Note that the power assignment across active sub-channels does not affect the asymptotic growth rate of the capacity. ∎

B. Asymptotic Capacity With On-Off Power Allocation

To reduce the amount of required feedback for power and rate optimization, we consider a specific feedback method, in which the transmitter allocates equal power $\bar{P}$ across a subset of sub-channels with gains that exceed a threshold μ_0. We refer to this as "on-off" feedback. The power constraint then becomes $\sum_{i=1}^{N} P_i = \sum_{i=1}^{N} \bar{P} \, \mathbf{1}_{\mu_i \geq \mu_0} = \mathcal{P}$, or

$$\bar{P} = \frac{\mathcal{P}}{\sum_{i=1}^{N} \mathbf{1}_{\mu_i \geq \mu_0}}.$$

Optimizing the threshold gives the corresponding on-off capacity for finite N,

$$C_N^{(\text{on-off})} = \max_{\mu_0} \sum_{i=1}^{N} \log\left(1 + \bar{P}\mu_i\right) \mathbf{1}_{\mu_i \geq \mu_0} \tag{6}$$

As $N \to \infty$, the power per active sub-channel converges to $\bar{P} = \frac{\mathcal{P}}{N \bar{F}_\mu(\mu_0)}$ w.p.1, where $\bar{F}_\mu(x) = 1 - F_\mu(x)$, and $F_\mu(\cdot)$ is the c.d.f. of channel gain μ.

The capacities corresponding to both optimal on-off and water-filling power allocations approach infinity as $N \to \infty$. To study the rate at which the capacity goes to infinity in each case, we define two sequences, $\{a_N\}$ and

5

$\{b_N\}$, as being *asymptotically equivalent* if the ratio a_N/b_N converges to one as $N \to \infty$. We denote this as

$$a_N \asymp b_N \Longleftrightarrow \frac{a_N}{b_N} \to 1 \qquad \text{w.p.1} \tag{7}$$

Theorem 1: If $0 < \frac{\bar{F}_\mu(x)}{f_\mu(x)} < \infty$ for all $x > 0$, then $C_N^{(\text{wf})} \asymp C_N^{(\text{on-off})} \asymp P\mu_0^*$, where the optimal threshold μ_0^* satisfies

$$\frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0^*)}E^2[\mu|\mu > \mu_0^*] = 2\left(E[\mu|\mu > \mu_0^*] - \mu_0^*\right) \tag{8}$$

The proof is given in Appendix A. Theorem 1 states that on-off feedback is asymptotically optimal in the sense that it achieves the same asymptotic growth in capacity as water-filling.

C. Finite-Precision Rate Control

We now consider a feedback scheme, which specifies rates for each sub-channel. As previously discussed, this is motivated by bit loading schemes, which have been used with multi-carrier modulation. Here we assume that the data rates across sub-channels are chosen from a small set of rates. Specifically, we define the n channel thresholds $0 < \nu_{n,0} < \nu_{n,1} < \cdots < \nu_{n,n-1} < \infty$, where $\nu_{n,0} = \mu_0$ and $\nu_{n,n} = \infty$. If the kth sub-channel gain satisfies $\nu_{n,i} \leq \mu_k < \nu_{n,i+1}$, then the power $P_k = \bar{P} = \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}$ (independent of i) and the rate assigned to sub-channel k, $R_k = \bar{R}_i = \log\left(1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}\nu_{n,i}\right)$. If $\mu_k < \mu_0$, then $P_k = R_k = 0$. The total data rate is then $R_n^{(\text{fp})} = \sum_{k=1}^N R_k$. The following theorem gives the optimal quantization thresholds for a class of fading distributions. In what follows, we let $\bar{R}_n^{(\text{fp})}$ denote the value of $E[R_n^{(\text{fp})}]$ maximized over $\nu_{n,0}, \cdots, \nu_{n,n-1}$.

Theorem 2: If $0 < \frac{\bar{F}_\mu(x)}{f_\mu(x)} < \infty$ for all x, then $\bar{R}_n^{(\text{fp})} \asymp E[R_n^{(\text{fp})}]$ with the thresholds defined as

$$\begin{aligned} \nu_{n,0} &= \mu_0 \\ \nu_{n,i} &= \nu_{n,i-1} + \frac{\bar{F}(\mu_i) - \bar{F}(\mu_{i+1})}{f(\mu_i)} \qquad 1 \leq i \leq n-2 \\ \nu_{n,n-1} &= \nu_{n,n-2} + \frac{\bar{F}(\mu_{n-1})}{f(\mu_{n-1})} \end{aligned} \tag{9}$$

Furthermore, with this set of thresholds the increase in rate obtained by adding quantization levels is bounded as

$$E[R_n^{(\text{fp})} - R_1^{(\text{fp})}] \leq \mathcal{P} \max_x \frac{\bar{F}_\mu(x)}{f_\mu(x)} \tag{10}$$

The proof is given in Appendix B.

D. Example: Rayleigh Channel

In this section, we assume that $f_\mu(x) = e^{-x}$, which corresponds to Rayleigh fading. The condition (8) for the optimal threshold then becomes

$$\frac{P}{N}e^{\mu_0^*}(\mu_0^* + 1)^2 = 2 \tag{11}$$

and $\mu_0^* \asymp \log N$. We restate Theorem 1 as the following corollary.

Corollary 1: With Rayleigh sub-channels, $C_N^{(\text{wf})} \asymp C_N^{(\text{on-off})} \asymp P \log N$. Furthermore, the optimal number of active sub-channels $\bar{N}_a \asymp P \log^2 N$.

We remark that this result is analogous to the asymptotic growth in downlink capacity with multiuser diversity presented in [7].[†]

Using (11), for finite N the capacity and the optimal number of active sub-channels, $\bar{N}_a$, can be more accurately approximated as

$$C_N^{\text{(on-off)}} \approx \mathcal{P} \left(\log \frac{2N}{\mathcal{P}} - 2 \log \left(\log \frac{2N}{\mathcal{P}} - 2 \log \log \frac{2N}{\mathcal{P}} \right) \right) \tag{12}$$

$$\bar{N}_a \approx \mathcal{P} \left(\log \frac{2N}{\mathcal{P}} - 2 \log \left(\log \frac{2N}{\mathcal{P}} - 2 \log \left(\log \frac{2N}{\mathcal{P}} - 2 \log \log \frac{2N}{\mathcal{P}} \right) \right) \right)^2. \tag{13}$$

We note that with both water-filling and optimal on-off power allocations, the number of active channels $\bar{N}_a \to \infty$ with N, but the fraction of active sub-channels decreases to zero, i.e., $\frac{\bar{N}_a}{N} \to 0$.

For comparison, we now consider the case in which only a finite number N_a sub-channels can be activated. Let $C_N^{(N_a)}$ denote this capacity. According to Lemma 1 and the associated discussion, this capacity is achieved by activating the N_a largest sub-channels.

Theorem 3: For fixed N_a, $\lim_{N\to\infty} \left(C_N^{(N_a)} - N_a \log \log N \right) \leq N_a \log \left(\frac{\mathcal{P}}{N_a} \right)$.

The proof is in Appendix C. That is, if we bound the number of channels that can be activated, then the achievable rate grows as $O(\log \log N)$, and the associated gap to capacity increases with the Signal-to-Noise Ratio (SNR).

Given the on-off threshold μ_0, the channel thresholds, which maximize $E_\mu[R_n^{\text{(fp)}}]$ with $E[\mu] = 1$, satisfy the following recursion,

$$\begin{aligned}
\nu_{n,0} &= \mu_0 \\
\nu_{n,i} &= \nu_{n,i-1} + 1 - e^{-(\nu_{n,i+1} - \nu_{n,i})} \qquad 1 \leq i \leq n-2 \\
\nu_{n,n-1} &= \nu_{n,n-2} + 1
\end{aligned} \tag{14}$$

Note that the increment $\nu_{n,i} - \nu_{n,i-1}$ depends only on n, and not on μ_0. In what follows, we will assume that the thresholds for the preceding finite-precision rate control scheme are chosen according to (14).

Let B_N denote the *total* number of feedback bits, which is a function of N. For the preceding on-off scheme with n levels per sub-channel, $\log N$ bits are needed to specify each active channel, hence $B_N = \bar{N}_a (\log n + \log N) \asymp \mathcal{P} \log^3 N$. To reduce the amount of feedback further, we can activate a finite number of sub-channels N_a. The number of feedback bits in that case is $B_N = N_a \log N$. The following theorem specifies the asymptotic growth in the achievable data rates for each of the preceding schemes with $E[\mu] = 1$.

Theorem 4:   1) If $B_N = 0$ (no feedback) then

$$C_N \to \mathcal{P} \qquad \text{w.p.1.} \tag{15}$$

[†]The capacity grows without bound in this case because the Rayleigh distribution has unbounded support. Although the Rayleigh fading model is appropriate for wideband channels of interest (e.g., see [9]), for purposes of characterizing the maximum of N sub-channels gains, it must break down for large enough N. It follows from results in extreme statistics (e.g., [29, Theorem 2.3.1]) that the $\log N$ growth in capacity holds for $N < N_0$ provided that the Rayleigh pdf is accurate up to channel gains of $\log_e N_0$. For example, to claim that the $\log N$ growth applies for $N < 1000$, the Rayleigh pdf should be accurate for sub-channel gains up to $\log_e 1000$, or approximately nine times the mean.

2) If $B_N = N_a \log N$ (N_a active sub-channels), then $\lim_{N \to \infty} \left(R_1^{(\text{fp},N_a)} - C_N^{(N_a)} \right) = 0$, where $R_1^{(\text{fp},N_a)}$ is the achievable finite-precision rate, and the same rate is assigned to N_a active sub-channels.

3) If $B_N = \bar{N}_a \log(nN) \asymp P \log^3 N$ ($\bar{N}_a \asymp \mathcal{P} \log^2 N$ active sub-channels with optimized rate thresholds), then $R_n^{(\text{fp})} \asymp C^{(\text{on-off})}$. Furthermore,

$$C^{(\text{on-off})} - R_n^{(\text{fp})} \to 2\mathcal{P}(1 - e^{-(\nu_{n,1} - \mu_0)}) \quad \text{w.p.1} \tag{16}$$

The proof is given in Appendix D. With fixed N_a, as $N \to \infty$, the difference between the achievable rate with quantized rates across sub-channels and the expected capacity (coding across sub-channels) tends to zero. With $O(\log^3 N)$ feedback, (16) states that this difference converges to a constant. (In (16), $\nu_{n,1} - \mu_0$ depends only on n.) This is stronger than asymptotic equivalence. As an example, referring to the third item in Theorem 4, the average loss in data rate is bounded by $\mathcal{P}$ nats when the sub-channels are assigned equal rates ($n = 1$). With 6 bits/active sub-channel, this average loss decreases to $0.03\mathcal{P}$ nats. It is easy to show that as $n \to \infty$, $R_n^{(\text{fp})} \to C^{(\text{on-off})}$ w.p.1.

Figure 2 shows plots of mean data rate vs. Signal-to-Noise Ratio (SNR) $\mathcal{P}$ for a multi-carrier system with 512 Rayleigh sub-channels with $E[\mu] = 1$. Curves are shown for water-filling, optimal on-off power allocation with coding across sub-channels, and finite-precision rate control with 0, 1 and 2 bits per sub-channel. The figure shows that the capacity of the optimal on-off power allocation is very close to the water-filling capacity. As indicated in (16), the gap between the achievable finite-precision data rate and water-filling capacity increases with power. As expected, the gap between the rate with finite-rate control and the optimal on-off capacity decreases as n increases.

Figure 3 shows mean achievable data rate vs. total number of sub-channels with an SNR of 10 dB. The results are averaged over the channel distribution. Simulated results are compared with the asymptotic expression (12), which applies to water-filling and optimal on-off power allocations. The achievable data rates with finite-precision rate control are close to the asymptotic estimates. The figure shows that the capacities increase roughly as $\log N$. As predicted by Theorem 4, the gap between the achievable rate with quantized sub-channel rates and the corresponding capacity converges to a constant. Figure 3 also shows that the capacity with ten active sub-channels is accurately approximated by the asymptotic results in Theorem 3.

Figure 4 shows the mean data rate vs. number of activated sub-channels N_a with the on-off power allocation, assuming infinite- and finite-precision rate control. There are $N = 1000$ sub-channels, and the SNR= 10 dB. According to (13), the optimal number of active channels $\bar{N}_a = 63$, which is consistent with the plot. This is not very close to $\mathcal{P} \log^2 N$ in Corollary 1 because N is not large enough. There is a substantial difference between the peak achievable rate and the OFDM capacity, which corresponds to zero feedback. The latter is given by the optimal on-off rate with $N_a = N$. We observe that as the number of sub-channels increases beyond the optimal value, the data rates decrease logarithmically with N_a. Finite-precision rate control performs worse than OFDM for large N_a since the rate set with feedback is determined by the *worst* active channel. This is clearly suboptimal when nearly all sub-channels are activated.
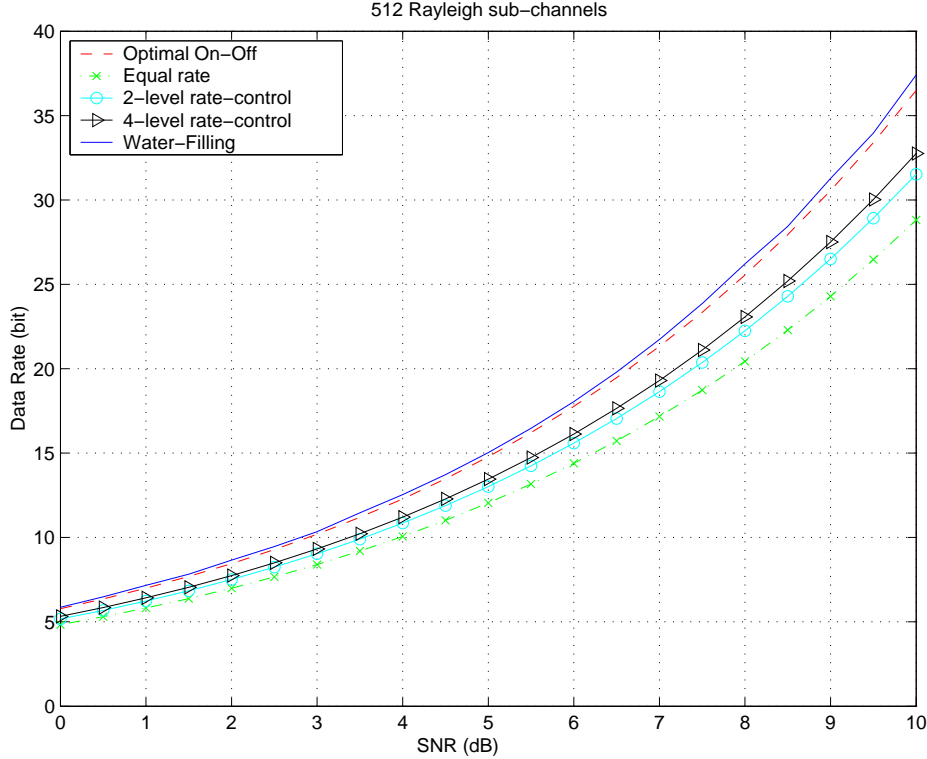
Fig. 2.   Mean channel capacity with water-filling and on-off power allocations, and achievable rates with finite-precision rate control vs. SNR.

E. Achievable Rate versus Feedback with Rayleigh Fading

Theorem 4 characterizes the growth in achievable rate with three specific feedback rates: $B_N = 0$ (no feedback), $B_N = O(\log N)$, and $B_N \asymp O(\log^3 N)$. More generally, we would like to characterize the growth in achievable rate given an arbitrary feedback rate. We can do this by varying the rate at which the on-off threshold μ_0 tends to infinity. This rate controls the feedback rate, and we can determine the associated increase in achievable data rate.

Motived by the optimal threshold condition in (11), we define the growth rate for the threshold μ_0 by choosing different values of γ in the relation

$$\frac{\mathcal{P}}{N} e^{\mu_0} \mu_0^{\gamma} = \kappa \tag{17}$$

where κ is a constant. The optimal threshold, which maximizes the growth in achievable rate, corresponds to $\gamma = 2, \kappa = 2$. In that case, μ_0 increases as $O(\log N - 2\log\log N)$, the feedback rate is $O(\log^3 N)$, and the data rate grows as $O(\log N)$. For general $\gamma > 0$, μ_0 increases as $O(\log N - \gamma\log\log N)$, and the feedback rate increases as $O(\log^{1+\gamma} N)$. Define the capacity corresponding to a specific choice of γ as $C_N^{(\gamma)}$. The following theorem characterizes the growth in capacity within different feedback regions characterized by the choice of γ in (17).
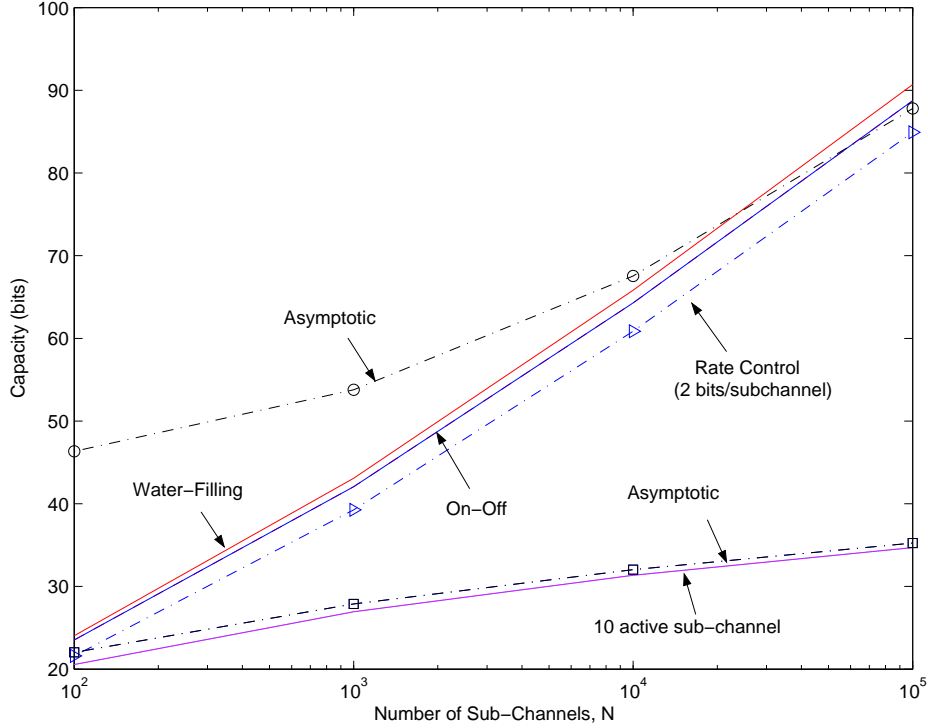
Fig. 3. Mean channel capacity vs. number of sub-channels with water-pouring and on-off power allocations. Results corresponding to both simulated systems and asymptotic estimates are shown. The SNR is 10 dB.

Theorem 5: If the feedback B_N grows as $O\left(\log^{\gamma+1} N\right)$, then the following asymptotic mean rates are achievable:

$$
E\left[C_N^{(\gamma)}\right] \asymp \begin{cases} \frac{1-\gamma}{\kappa}\mathcal{P}\log^\gamma N \log\log N & 0 \le \gamma < 1 \\ \frac{\log(1+\kappa)}{\kappa}\mathcal{P}\log N & \gamma = 1 \\ \mathcal{P}\log N & \gamma > 1 \end{cases} \tag{18}
$$

Furthermore, a finite absolute loss in achievable rate is incurred with finite-precision rate control.

The proof is given in Appendix E. According to (18), if the feedback increases faster than $O\left(\log^2 N\right)$, then the $O(\log N)$ growth in capacity can be achieved. This appears to be inconsistent with Theorem 4, which states that the feedback rate $O(\log^3 N)$ achieves the optimal on-off capacity. However, the proof of Theorem 5 shows that for $1 < \gamma < 2$ there is an *absolute* loss in capacity, which increases as $O(\log^{2-\gamma} N)$.

The differences in achievable rate shown around $\gamma = 1$ can be explained as follows. If $\gamma = 1 + \epsilon$ for small $\epsilon > 0$, then the achievable rate is the same as that obtained by taking $\gamma = 1$ and choosing a sufficiently small $\kappa > 0$. In that sense $E\left[C_N^{(1+\epsilon)}\right] \approx E\left[C_N^{(1)}\right]$. Also, if $\gamma = 1 - \epsilon$, then by choosing an appropriate κ for $\gamma = 1$, we still have $E\left[C_N^{(1-\epsilon)}\right] \approx E\left[C_N^{(1)}\right]$. Hence there is a continuous transition between the different capacity expressions corresponding to the different regions of γ shown in (18). The scenario in which only a finite number of sub-channels $\mathcal{P}/\kappa$ are activated corresponds to $\gamma = 0$. Setting $\gamma = 0$ in (18) gives an achievable rate, which grows as $O(\log\log N)$, as stated in Theorem 3.
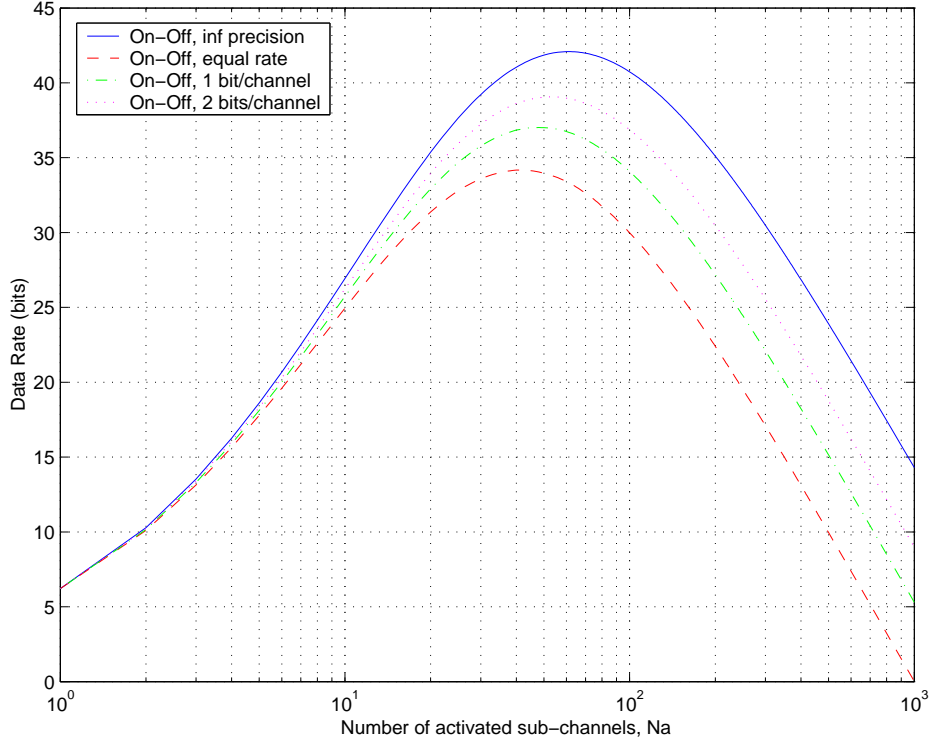
Fig. 4. Average data rate vs. the number of active sub-channels.

F. Correlated Sub-Channels

In the preceding sections, it has been shown that for a class of fading distributions, an achievable rate, which is asymptotically equivalent to the water-filling capacity, can be obtained by feeding back an optimized subset of sub-channels to activate. Since the channel capacity depends only on the first-order distribution of the sub-channel gains, that result does not depend on whether the sub-channel gains are independent or correlated. However, correlation among sub-channels can be exploited to reduce the amount of feedback needed to represent the subset of sub-channels to activate. In this section we compare the feedback needed to achieve the order-optimal growth in capacity with correlated and uncorrelated sub-channels.

Assuming on-off feedback with threshold μ_0, we let

$$\chi_i = \begin{cases} 1 & \text{if } \mu_i \geq \mu_0 \\ 0 & \text{if } \mu_i < \mu_0 \end{cases}$$

where $1 \leq i \leq N$, and define the corresponding sequence $\chi^N = (\chi_1, \chi_2, \cdots, \chi_N)$. That is, $\chi_i = 1$ if sub-channel i is active, and χ^N specifies the sequence of active sub-channels. Define the minimum feedback rate needed to represent χ^N as B_N, measured in *bits per sub-channel*. If the sequence of sub-channels is a stationary process, then as $N \to \infty$, B_N converges to the entropy rate of the sequence χ^N. If the sub-channels are *i.i.d.*, then the corresponding feedback rate $B_N^{(\text{iid})} \to H(\gamma) = -\gamma \log \gamma - (1 - \gamma) \log(1 - \gamma)$, where $\gamma = \bar{F}_\mu(\mu_0)$. If $\mu_0 \to \infty$ with

11

N, then $\gamma \to 0$, in which case $-\gamma \log \gamma \gg -(1-\gamma) \log(1-\gamma)$, so that

$$B_N^{(\text{iid})}(\mu_0) \asymp \gamma \log \frac{1}{\gamma}, \tag{19}$$

which goes to zero with N.

Suppose that the sequence of sub-channel gains $\{\mu_1, \cdots, \mu_N\}$ is a Markov process with joint second-order pdf $g(\mu_i, \mu_{i-1})$. Clearly, $\{\chi_1, \chi_2, \cdots, \chi_N\}$ is a two-state Markov chain, as shown in Fig. 5, with transition probabilities

$$q = Pr\{\chi_i = 1 | \chi_{i-1} = 1\} = \frac{1}{\gamma} \int_{\mu_0}^{\infty} \int_{\mu_0}^{\infty} g(x, y) \, dx \, dy \tag{20}$$

$$p = Pr\{\chi_i = 0 | \chi_{i-1} = 0\} = \frac{1 - (2 - q)\gamma}{1 - \gamma} \tag{21}$$

The asymptotic on-off feedback rate is then given by [27]
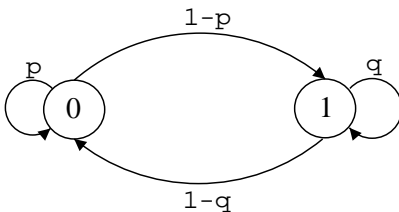

Fig. 5.   Two-state Markov chain corresponding to the sequence χ^N.

$$B_N^{(\text{corr})} = \frac{(1-p)H(q) + (1-q)H(p)}{2 - p - q} = \gamma H(q) + (1 - \gamma)H(p). \tag{22}$$

Now suppose that the data rate on each sub-channel is chosen from one of n possible rates. According to Theorem 2, the maximum achievable rate is asymptotically equivalent to the capacity, although there is an absolute loss in rate. Given a set of n thresholds $\Omega = \{\nu_{n,0}, \nu_{n,1}, \cdots, \nu_{n,n-1}\}$, the gain of each sub-channel is quantized as χ_k, where $\chi_k \in \{0, 1, \cdots, n\}$. That is, for the kth sub-channel with gain μ_k,

$$\chi_k = \begin{cases} 0 & \mu_k < \mu_0 \\ i+1 & \nu_{n,i} \leq \mu_k < \nu_{n,i+1} \end{cases}$$

where $\nu_{n,n} = \infty$. The corresponding rate for sub-channel k is then $R_k = \bar{R}_{\chi_k - 1}$ for $1 \leq \chi_k \leq n$, and $R_k = 0$ for $\chi_k = 0$.

The sequence $\chi^N = (\chi_1, \cdots, \chi_N)$, which is fed back to the transmitter, is an $(n+1)$-state Markov chain, as shown in Fig 6. Let $B_N^{(\text{fp},n)}(\mu_0)$ denote the minimum feedback rate with n-level finite-precision rate control, which is given by

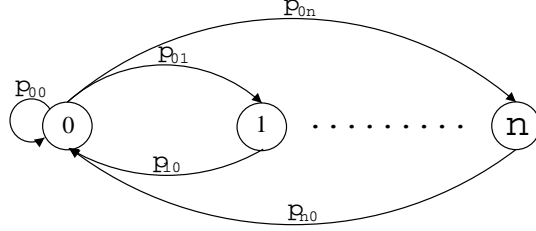$$B_N^{(\text{fp},n)}(\mu_0) = \sum_{i,j=0}^{n} \pi_i p_{ij} \log \frac{1}{p_{ij}}$$

12

Fig. 6. $(n+1)$-state Markov chain corresponding to feedback with finite-precision rate control.

where p_{ij} is the transition probability from state i to state j, and $\{\pi_i\}$ is the steady-state distribution. Generalizing q from (20) to the $(n+1)$-state Markov chain gives

$$q = Pr\{\chi_i \neq 0 | \chi_{i-1} \neq 0\} = Pr\{\mu_i \geq \mu_0 | \mu_{i-1} \geq \mu_0\} = \frac{1}{\gamma} \int_{\mu_0}^{\infty} \int_{\mu_0}^{\infty} g(x,y)dxdy.$$

We now consider the minimum feedback rate $B_N^{(\text{corr})}$ as $N \to \infty$. An objective is to use this analysis to predict the corresponding amount of feedback needed for a given wideband system that contains a finite number of coherence bands. In what follows, we will therefore assume that the ratio of coherence bandwidth to the total available bandwidth is fixed. In that case as N increases, the correlation between neighboring sub-channels increases, i.e., $\lim_{N \to \infty} q = 1$. This limit can be interpreted as keeping the total bandwidth fixed, and letting the sub-channel width tend to zero. Let

$$\theta(\mu_0) = \frac{Pr\{\mu_i \geq \mu_0\}}{Pr\{\mu_i < \mu_0 | \mu_{i-1} \geq \mu_0\}} = \frac{\gamma}{1-q}. \tag{23}$$

The asymptotic feedback rates for the optimal on-off power allocation and finite-precision rate control schemes, relative to the feedback rates for *i.i.d.* sub-channels, are given in the following theorem.

Theorem 6: If the threshold $\mu_0 \to \infty$ with N, and $\theta(\mu_0) < \infty$, then

$$\lim_{N \to \infty} \frac{B_N^{(\text{corr})}(\mu_0)}{B_N^{(\text{iid})}(\mu_0)} = \lim_{N \to \infty} \frac{B_N^{(\text{fp},n)}(\mu_0)}{B_N^{(\text{iid})}(\mu_0)} = 0$$

Furthermore, if $\theta(\mu_0) \to 0$ with N, then

$$B_N^{(\text{corr})}(\mu_0) \asymp B_N^{(\text{fp},n)} \asymp (1-q)\gamma \log \frac{1}{\gamma}. \tag{24}$$

The proof is given in Appendix F. The condition $\theta(\mu_0) \to 0$ is needed to give the expression for the feedback rate in (24), and is satisfied in many situations of interest. With this condition it is shown in the proof that

$$\frac{B_N^{(\text{corr})}(\mu_0)}{B_N^{(\text{iid})}(\mu_0)} \asymp \frac{B_N^{(\text{fp},n)}(\mu_0)}{B_N^{(\text{iid})}(\mu_0)} \asymp 1-q.$$

We also conclude that given the same sequence of thresholds with N, the growth in feedback versus N for finite-precision rate control with a finite number of rate levels n does not depend on n and the set of thresholds. Hence for large N the additional overhead needed to specify one of n data rate levels is negligible compared with the feedback needed to specify the binary on-off sequence. Furthermore, by choosing the optimal threshold set, the corresponding achievable rate is asymptotically equivalent to the capacity.

13

As an example, consider Rayleigh fading for which $f_\mu(x) = e^{-x}$. To model the correlation among sub-channels, we assume that the sequence $\{h_i\}$ is generated from a first-order autoregressive model,

$$h_i = \alpha h_{i-1} + \xi_i \tag{25}$$

where $0 \leq \alpha \leq 1$, and $\{\xi_i\}$ is sequence of *i.i.d.* complex Gaussian random variables, each with variance $1 - \alpha^2$, so that $E[|h_i|^2] = 1$. The parameter α determines the correlations between the sub-channels.[‡] Assuming that the total bandwidth is fixed, as $N \to \infty$, with fixed coherence bandwidth the correlation between neighboring sub-channels increases, so that $\alpha \to 1$.

It is straightforward to show that (25) implies that $\{\mu_1, \cdots, \mu_N\}$ is a Markov process. Conditioned on h_{i-1}, h_i is Rician distributed, and it is easy to obtain the second-order distribution for the sub-channels,

$$g(\mu_{i-1}, \mu_i) = \frac{1}{1 - \alpha^2} e^{-\frac{\mu_{i-1} + \mu_i}{1 - \alpha^2}} I_0\left(\frac{2\alpha\sqrt{\mu_{i-1}\mu_i}}{1 - \alpha^2}\right) \tag{26}$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind and zero-order. If the threshold $\mu_0 \to \infty$ as $N \to \infty$, then the Rician distribution can be approximated as a Gaussian distribution, and the transition probability q can be approximated as

$$
\begin{aligned}
q &\approx e^{\mu_0} \int_{\sqrt{\mu_0}}^\infty 2xe^{-x^2} \int_{\sqrt{\mu_0}}^\infty \frac{1}{(1 - \alpha^2)\pi} e^{-\frac{(y - \alpha x)^2}{1 - \alpha^2}} \, dy \, dx \\
&= (1 + \alpha) Q\left(\sqrt{\frac{1 - \alpha}{1 + \alpha} 2\mu_0}\right)
\end{aligned} \tag{27}
$$

where $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$. Substituting (27) into (24) gives the respective asymptotic on-off and finite-precision feedback rates. For example, if $\lim_{N \to \infty} \alpha = 1$, then choosing the (sub-optimal) threshold $\mu_0 = \frac{1}{1 - \alpha}$ gives the relative feedback gain $B_N^{(\text{corr})}/B_N^{(\text{iid})} \to 0.68$. If $(1 - \alpha)\mu_0 \to 0$, then (27) can be further approximated as

$$q \approx 1 - \sqrt{\frac{2(1 - \alpha)\mu_0}{\pi}} \tag{28}$$

and the ratio of minimum feedback rates for correlated and *i.i.d.* sub-channels converges to zero as $O\left(\sqrt{(1 - \alpha)\mu_0}\right)$.

The reduction in feedback obtained by exploiting the sub-channel correlations clearly depends on the rate at which $\alpha \to 1$ with N. Here we consider a special case, motivated by a specific definition of coherence bandwidth. Let W be the total bandwidth of the channel and W_c be the coherence bandwidth. We define the coherence bandwidth $W_c = M\Delta f$, where Δf is the width of the sub-channels, and

$$M = \max\{k : \text{cov}(\mu_i, \mu_{i+k}) \geq \rho\} \tag{29}$$

where ρ is the correlation between sub-channels separated by W_c, and $0 < \rho < 1$. The number of coherence bands spanned by the channel is assumed to be fixed, i.e., $\frac{W_c}{W} = \frac{M}{N} = \beta$ where β is a constant. Since $\text{cov}(\mu_i, \mu_{i+k}) = \alpha^{2k}$, we can write

$$\alpha = \rho^{\frac{1}{2\beta N}} = e^{-\frac{\log \frac{1}{\rho}}{2\beta N}}, \tag{30}$$

[‡] An autogressive model is proposed for Ultra-Wideband channels in [39]. This model is statistically equivalent to a power-delay profile model, which gives the same second-order statistics for the sub-channel process $\{h_i\}$.

which increases to one exponentially with N.

Corollary 2: If $\lim_{N\to\infty} \frac{\mu_0}{\frac{1}{2}\log N} \geq 1$, then the on-off feedback rate with *i.i.d.* sub-channels is

$$B_N^{(\text{iid})} \asymp e^{-\mu_0}\mu_0. \tag{31}$$

Furthermore, the on-off and finite-precision feedback rates for the autoregressive sequence of sub-channels satisfy

$$B_N^{(\text{corr})} \asymp B_N^{(\text{fp},n)} \asymp \frac{\beta_1}{\sqrt{N}}e^{-\mu_0}\mu_0^{\frac{3}{2}} \tag{32}$$

where $\beta_1 = \sqrt{\frac{\log\frac{1}{\rho}}{\pi\beta}}$.

Comparing the feedback rates for the autoregressive and *i.i.d* cases gives

$$\frac{B_N^{(\text{corr})}}{B_N^{(\text{iid})}} \asymp \frac{B_N^{(\text{fp},n)}}{B_N^{(\text{iid})}} \asymp \beta_1\sqrt{\frac{\mu_0}{N}}. \tag{33}$$

Specifically, if $\mu_0 \asymp b\log N$, where $b \geq 1/2$ is a constant, then

$$\frac{B_N^{(\text{corr})}}{B_N^{(\text{iid})}} \asymp \frac{B_N^{(\text{fp},n)}}{B_N^{(\text{iid})}} \asymp \sqrt{b}\beta_1\sqrt{\frac{\log N}{N}}. \tag{34}$$

The condition $b \geq 1/2$ is needed so that (28) is valid.

Given the total bandwidth W and coherence bandwidth W_c, the minimum feedback rate for on-off and finite-precision rate control can be estimated from (32). For example, choosing the optimal threshold μ_0^* for Rayleigh fading (i.e., $\mu_0^* \asymp \log N$) gives a minimum feedback rate of $O\left(\frac{\log^3 N}{N}\right)$ for *i.i.d.* sub-channels. For the autoregressive model, the minimum feedback rate, relative to the *i.i.d.* case, is reduced by the factor $O\left(\sqrt{\frac{\log N}{N}}\right)$. In both cases, the maximum achievable rate is asymptotically equivalent to the capacity.

Here we give a numerical example motivated by a cellular system. The channel bandwidth is 5 MHz, and the coherence bandwidth with $\rho = 0.5$ is 146 kHz [30]. We take the threshold $\mu_0 = \frac{1}{2}\log N$. (Taking $\mu_0 > \frac{1}{2}\log N$ complicates the generation of simulated results, since for finite N, active sub-channels occur relatively infrequently.) Figure 7 shows the feedback rate vs. N computed from Corollary 1, and by generating sample sequences of sub-channel gains according to (25), and encoding the on-off sequence with an arithmetic code. Results are shown for both *i.i.d.* and correlated sub-channels with the same threshold μ_0. We observe that the asymptotic bounds for both *i.i.d.* and correlated sub-channels decrease logarithmically with N, i.e., $\log B_N^{(\text{corr})}$ decreases as $-\log N$ and $\log B_N^{(\text{iid})}$ decreases as $-\frac{1}{2}\log N$. As N increases, correlated sub-channels greatly reduce the required feedback, relative to *i.i.d.* sub-channels, as indicated in (33). The variance of the simulated results with correlated channels increases with N, and the asymptotic results predict the mean behavior.

Figure 8 shows the feedback rate vs. α^2 with $N = 5000$. The feedback rate for simulated sub-channels with arithmetic encoding averaged over many realizations is also shown. The optimal threshold μ_0^* for Rayleigh fading, given by (8), is used here. As α increases, the feedback rate decreases slowly when α is small, and drops rapidly when $\alpha > 0.9$. Hence for fixed α, correlated sub-channels enable a significant decrease in the feedback rate only when α is close to one. The mean feedback rate for the simulated results follows the asymptotic curve, although the variance of the simulated results increases with α.
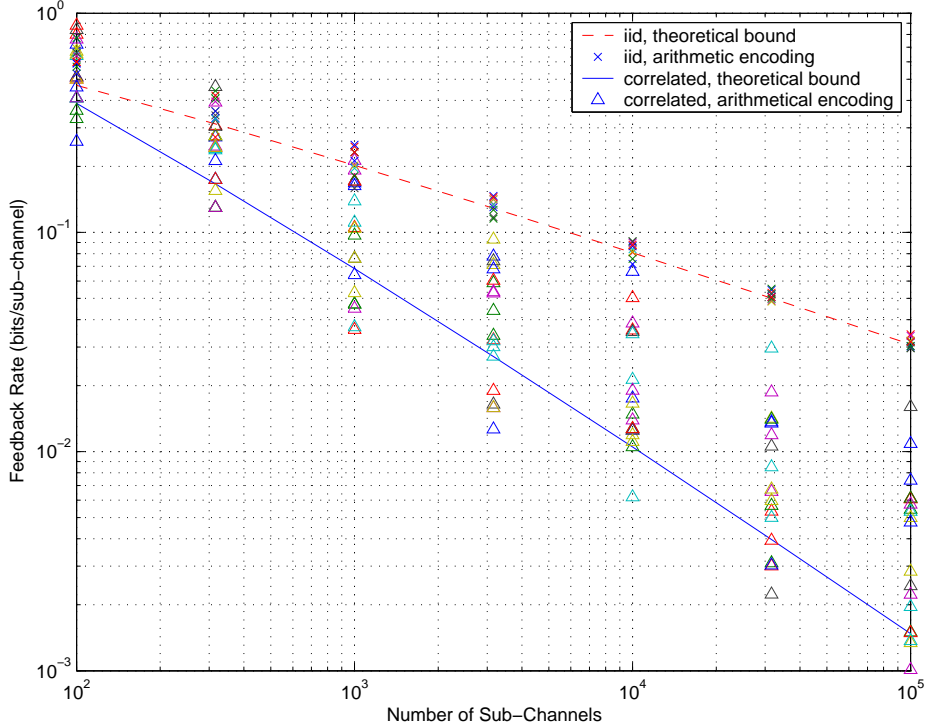
Fig. 7. Feedback rate vs. number of sub-channels N.

III. MULTIPLE USERS

In this section, the previous analysis for a single-user link is extended to multiple access and broadcast channels. Power allocation schemes for multi-user OFDM (or Orthogonal Frequency Division Multiple Access (OFDMA)) have been discussed in [8], [23], [24]. Here we characterize the asymptotic growth of the sum capacity with perfect channel knowledge at the transmitters, assuming each sub-channel can be assigned to at most one user, and present a limited feedback scheme that achieves this order-optimal growth. We consider two limits: in the first, the number of users K is fixed as $N \to \infty$, and in the second, $K \to \infty$ in proportion with N. Also, in the first limit rates are summed over sub-channels, whereas in the second limit rates are normalized by $1/N$.

A. Finite Number of Users

1) Multiple Access: The K users transmit multi-carrier signals to a single receiver. We assume N sub-channels, where the ith sub-channel gain for user k is $\mu_{k,i}$. The sub-channel gains for each user k are assumed to be *i.i.d.* with c.d.f. $F_k(\mu)$, and are independent across users. Although we initially assume that $F_k(\mu)$ can vary across users, our main results assume that the users have identical c.d.f.'s. For each user the total transmit power is constrained to be no more than $\mathcal{P}_k$, i.e., $\sum_{i=1}^{N} P_{k,i} \leq \mathcal{P}_k$, where $P_{k,i}$ is the power assigned to the ith sub-channel for user k.

The optimal power allocation, which maximizes the total (sum) capacity, is multiuser water-filling [23], [32].
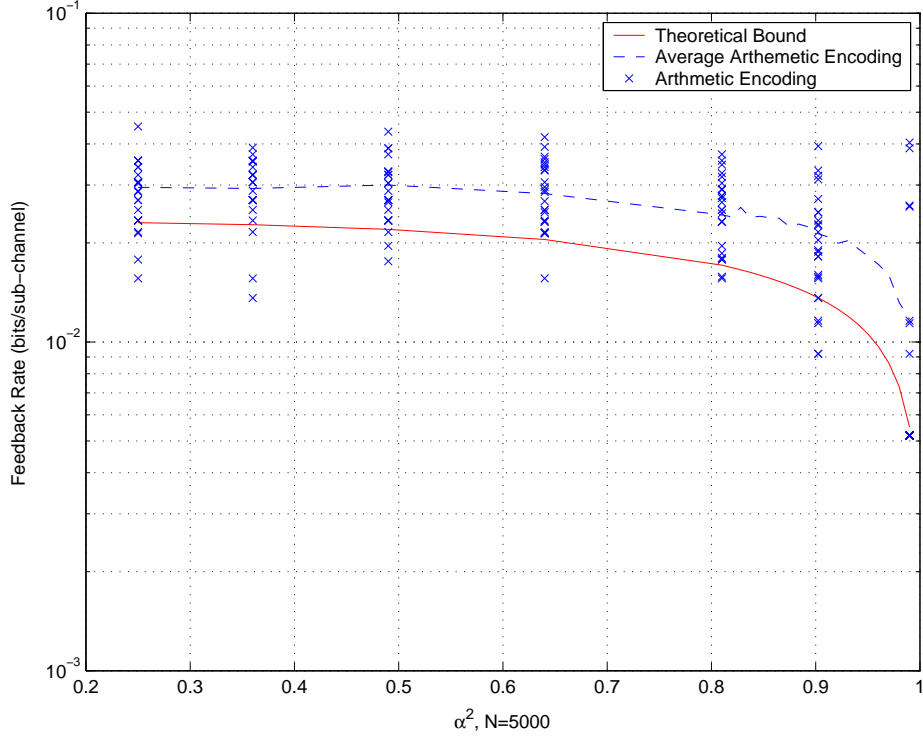
16

Fig. 8. Feedback rate vs. channel correlation, represented by the autogressive parameter α^2. There are 5000 sub-channels.

Namely,

$$P_{k,i} = \begin{cases} \left(\lambda_k - \frac{1}{\mu_{k,i}}\right)^+ & \text{if } \lambda_k \mu_{k,i} = \max_{1 \le j \le K} \lambda_j \mu_{j,i} \\ 0 & \text{otherwise} \end{cases} \tag{35}$$

where λ_k is the water-level for user k, defined by

$$\frac{\mathcal{P}_k}{N} = E\left[\left(\lambda_k - \frac{1}{\mu_{k,i}}\right)^+ \mathbf{1}_{\lambda_k \mu_{k,i} = \max_{1 \le j \le K} \lambda_j \mu_{j,i}}\right] \tag{36}$$

and the sum capacity is given by

$$C_{\text{up}}^{(\text{wf},K)} = \sum_{k=1}^{K} \sum_{i=1}^{N} E\left[(\log(\lambda_k \mu_{k,i}))^+ \mathbf{1}_{\lambda_k \mu_{k,i} = \max_j \lambda_{j,i} \mu_{j,i}}\right]. \tag{37}$$

Note that here and throughout this section, the capacity expressions are averaged over the channel distribution, and are therefore deterministic.

As for the single link, we consider on-off power allocation to limit the feedback rate, assuming the base station knows all channel gains. For each user k, the base station sets a threshold $\mu_{k,0}$ and determines the subset of sub-channels with gains that exceed the threshold. If a sub-channel gain exceeds the threshold for multiple users, then the base station exclusively assigns the subchannel to a single user, which provides the largest achievable rate associated with that sub-channel. (Note that the rate depends on the power assigned to that sub-channel.) The set

of assigned sub-channels are then fed back to each user. Each user k allocates equal power $\bar{P}_k$ across the subset of active sub-channels. Therefore the power for the ith sub-channel, assumed to be assigned to user k, is

$$P_{k,i} = \bar{P}_k \mathbf{1}_{\mu_{k,i} \geq \mu_{k,0}} \mathbf{1}_{\bar{P}_k \mu_{k,i} = \max_j \bar{P}_j \mu_{j,i}} \tag{38}$$

where $\bar{P}_k$ is determined by the power constraint

$$\frac{\mathcal{P}_k}{N} = \bar{P}_k E\left[\mathbf{1}_{\mu_{k,i} \geq \mu_{k,0}} \mathbf{1}_{\bar{P}_k \mu_{k,i} = \max_j \bar{P}_j \mu_{j,i}}\right]$$

The corresponding uplink on-off sum capacity is given by

$$C_{\text{up}}^{(\text{on-off}, K)} = \sum_{k=1}^{K} \sum_{i=1}^{N} E\left[\log(1 + \bar{P}_k \mu_{k,i})\mathbf{1}_{\mu_{k,i} > \mu_{k,0}} \mathbf{1}_{\bar{P}_k \mu_{k,i} = \max_j \bar{P}_j \mu_{j,i}}\right].$$

As before, the rate per user can be quantized by quantizing the rates across sub-channels. Here we only consider assigning the same rate to all sub-channels (one-level rate control). The achievable rate is then

$$R_{\text{up}}^{(\text{fp}, K)} = \sum_{k=1}^{K} \sum_{i=1}^{N} \log(1 + \bar{P}_k \mu_{k,0}) E\left[\mathbf{1}_{\mu_{k,i} > \mu_{k,0}} \mathbf{1}_{\bar{P}_k \mu_{k,i} = \max_j \bar{P}_j \mu_{j,i}}\right]. \tag{39}$$

For the following theorem, we assume that the users have the same channel gain distribution and the same power constraint.

Theorem 7: Given K users, as $N \to \infty$, the sum capacities corresponding to the multiuser water-filling and on-off power allocations are asymptotically equivalent, and scale linearly with the number of users, i.e.,

$$C_{\text{up}}^{(\text{on-off}, K)} \asymp C_{\text{up}}^{(\text{wf}, K)} \asymp K\mathcal{P}\mu_0^* \tag{40}$$

where μ_0^* is the optimal threshold for a single link in (8). Furthermore, $C_{\text{up}}^{(\text{on-off}, K)} - R_{\text{up}}^{(\text{fp}, K)}$ converges to a finite constant.

The proof is given in Appendix G. The uplink sum capacity with K users therefore behaves asymptotically as if the users were transmitting over non-overlapping channels. This is because as $N \to \infty$, the probability that a particular sub-channel is requested by a user tends to zero, hence the probability that more than one user requests the same sub-channel also tends to zero.[§] For large N, finite-precision rate control gives a constant loss in capacity.

2) Broadcast: The base station now transmits multi-carrier signals to the K users. We again let $\mu_{k,i}$ denote the ith channel gain for user k, and assume that the sub-channels for user k are *i.i.d.* with c.d.f. $F_k(\mu)$. The sub-channels across users are again independent. The total transmit power is constrained to be no more than $K\mathcal{P}$.

To maximize the sum capacity, the base station assigns each sub-channel to the user with the largest sub-channel gain, and then water-fills over those sub-channels [31]. The power assigned to sub-channel i is therefore

$$P_i = \left(\lambda - \frac{1}{\mu_{k_i,i}}\right)^+$$

where sub-channel i is assigned to k_i, $\mu_{k_i,i} = \max_k \mu_{k,i}$, and λ is the water level chosen to satisfy $\sum_{i=1}^{N} P_i \leq K\mathcal{P}$.

[§] The proof must show that the increase in sum rate due to multi-user diversity goes to zero.

Let $G(x) = \prod_{k=1}^{K} F_k(x)$ denote the c.d.f. for $\mu_{k_i,i}$. As $N \to \infty$, the power constraint becomes

$$\int_0^\infty \left(\lambda - \frac{1}{x} \right)^+ dG(x) = \frac{K\mathcal{P}}{N} \tag{41}$$

and the water-filling capacity per sub-channel is given by

$$C_{\text{down}}^{(\text{wf},K)} = N \int_0^\infty \log{(\lambda x)}^+ dG(x). \tag{42}$$

To limit feedback we consider the on-off power allocation with a single threshold μ_0 for all users. Each user informs the base station the indices of sub-channels, which have gains exceeding the threshold. The base station then allocates uniform power across all active sub-channels. If more than one user requests the same sub-channel, then the base station randomly selects one of those users to transmit.[¶] Hence the power on sub-channel i is $P_i = \bar{P} \mathbf{1}_{\mu_{k_i,i} \geq \mu_0}$. As $N \to \infty$, we have

$$\bar{P} = \frac{K\mathcal{P}}{N[1 - G(\mu_0)]}, \tag{43}$$

and the on-off capacity per sub-channel is given by

$$C_{\text{down}}^{(\text{on-off},K)} = NE\left[C_i \mid |\mathcal{M}_i| \geq 1\right] \Pr\{|\mathcal{M}| \geq 1\} \tag{44}$$

where $\mathcal{M}_i = \{k : \mu_{k,i} \geq \mu_0\}$.

We again consider finite-precision rate control with equal rates on all active sub-channels. The achievable rate per sub-channel is then

$$R_{\text{down}}^{(\text{fp},K)} = N \log{\left(1 + \bar{P}\mu_0\right)} \Pr\{|\mathcal{M}_i| \geq 1\}. $$

The following theorem compares the asymptotic behavior of the downlink capacity with water-filling and on-off power allocations, and the achievable rate with finite-precision rate control.

Theorem 8: Given K users with identical channel gain distributions, as $N \to \infty$, the water-filling and on-off capacities satisfy

$$C_{\text{down}}^{(\text{wf},K)} \asymp C_{\text{down}}^{(\text{on-off},K)} \asymp K\mathcal{P}\mu_0^*. \tag{45}$$

Furthermore, $C_{\text{down}}^{(\text{on-off},K)} - R_{\text{down}}^{(\text{fp},K)}$ converges to finite constant.

The proof is given in Appendix H. Theorems 7 and 8 state that the uplink and downlink capacities have the same asymptotic behavior, although the assignment of sub-channels to users is accomplished in different ways.

B. Infinite Number of Users

1) Downlink: Now we extend the results in section III-A.2 for finite number of users to the limit in which the number of users $K \to \infty$ with fixed number of users per sub-channel $\beta = \frac{K}{N}$. That is, the system size grows with fixed load. For simplicity, in this section the sub-channel gains across users are assumed to be *i.i.d.*, and the

---

[¶]Here we ignore fairness issues, which must be considered in scheduling algorithms.

sub-channel gain distribution $F_k(x) = F_\mu(x) = 1 - e^{-x}$, corresponding to Rayleigh fading. The following results rely on the limiting behavior of the extremal distribution $F_\mu^K(\cdot)$. Namely, from [29, Theorem 2.3.1]

$$\lim_{K \to \infty} F_\mu^K(a_K + b_K x) = \exp(-e^{-x}) \tag{46}$$

We remark that the extension of the following results to the more general class of sub-channel distributions, which satisfy the extremal relation (46) with the appropriate sequences a_K and b_K (see [29, Theorem 2.3.1]), does not appear to be straightforward. Here we simply note that if the results can be extended in this way, then a_K is given by $F^{-1}(1 - 1/K)$ instead of $\log K$ [29, Theorem 2.3.1].

Theorem 9: The downlink sum capacities with water-filling and on-off power allocations, and the achievable rate with finite-precision rate control satisfy

$$
\begin{aligned}
\lim_{K \to \infty} C_{\text{down}}^{(\text{wf}, K)} - C_{\text{down}}^{(\text{on-off}, K)} &= \lim_{K \to \infty} C_{\text{down}}^{(\text{on-off}, K)} - R_{\text{down}}^{(\text{fp}, K)} \\
&= \lim_{K \to \infty} R_{\text{down}}^{(\text{fp}, K)} - \log(1 + \beta \mathcal{P} a_K) = 0.
\end{aligned}
\tag{47}
$$

The proof is given in Appendix I. The downlink capacity increases as $O(\log \log K)$, which is consistent with analogous results in [31].[||]

2) Uplink: We now consider the uplink model discussed in section III-A.1, and let $K \to \infty$ with fixed $\frac{K}{N} = \beta$. The users have the same power constraint $\mathcal{P}$, and the same fading distribution, i.e., $F_k(x) = F_\mu(x)$, so that $G(x) = F_\mu^K(x)$.

With the water-filling power allocation it is easy to show that all users have the same water level, i.e., $\lambda_k = \lambda$, and (70) becomes

$$\beta \mathcal{P} = \int_0^\infty \left(\lambda - \frac{1}{x} \right)^+ dG(x). \tag{48}$$

Similarly, we can show that the asymptotic uplink capacity per sub-channel is given by

$$C_{\text{up}}^{(\text{wf}, \infty)} = \int_0^\infty [\log(\lambda x)]^+ dG(x). \tag{49}$$

From the downlink analysis in the preceding section it follows that

$$C_{\text{up}}^{(\text{wf}, \infty)} - \log(1 + \beta \mathcal{P} a_K) \to 0. \tag{50}$$

With the on-off power allocation, all users have the same threshold μ_0, so that the power constraint becomes (see (75) in Appendix F)

$$\beta \mathcal{P} = \bar{P}[1 - G(\mu_0)]$$

[||]A more precise characterization of the asymptotic behavior of the downlink on-off sum capacity has been recently presented in [33].

and the capacity per sub-channel is given by (see (77) in Appendix F)

$$
\begin{aligned}
C_{\text{up}}^{(\text{on-off},\infty)} &= \int_{\mu_0}^{\infty} \log(1 + \bar{P}x)dG(x) \\
&\geq (1 - G(\mu_0))\log(1 + \bar{P}\mu_0) = R_{\text{up}}^{(\text{fp},\infty)}.
\end{aligned}
$$

From the analysis in the preceding section it follows that

$$
C_{\text{up}}^{(\text{wf},\infty)} \geq C_{\text{down}}^{(\text{on-off},\infty)} \geq R_{\text{up}}^{(\text{fp},\infty)} \tag{51}
$$

and $R_{\text{up}}^{(\text{fp},\infty)} - \log(1 + \beta \mathcal{P}a_K) \to 0$. We restate this as the following theorem.

Theorem 10: The uplink sum capacities with water-filling and on-off power allocations, and the achievable rate with finite-precision rate control satisfy

$$
\begin{aligned}
\lim_{K\to\infty} C_{\text{up}}^{(\text{wf},K)} - C_{\text{up}}^{(\text{on-off},K)} &= \lim_{K\to\infty} C_{\text{up}}^{(\text{on-off},K)} - R_{\text{up}}^{(\text{fp},K)} \\
&= \lim_{K\to\infty} R_{\text{up}}^{(\text{fp},K)} - \log(1 + \beta \mathcal{P}a_K) = 0. \tag{52}
\end{aligned}
$$

C. Numerical Results

Figure 9 compares the preceding asymptotic results for the downlink with the analogous results obtained by numerically averaging the corresponding rate expressions over different channel realizations. For the finite-size system there are $N = 500$ *i.i.d.* sub-channels, and the power per user is 10 dB. Curves are shown for water-filling and optimized on-off power allocations, and also for equal rates assigned to active sub-channels (one-bit rate control). The capacity *per user* is shown, and as predicted by Theorem 8, is approximately constant when the number of users is small (e.g., ≤ 10). (Of course, the insensitivity to the system load when K is small becomes more pronounced as N increases.) As K increases, the curves converge, and decrease as $(\log\log K)/K$, as stated in Theorem 9.

IV. CONCLUSIONS

We have studied the benefits of relaying quantized channel gains to the transmitter, assuming multi-carrier modulation through a frequency-selective fading channel. Our results relate the asymptotic growth in capacity with number of sub-channels to the number of feedback bits per transmitted codeword. Specifically, with *i.i.d.* Rayleigh fading sub-channels, $O(\log^3 N)$ feedback bits per codeword can achieve the optimal $O(\log N)$ growth in capacity, corresponding to water-pouring with exact channel information at the transmitter. The absolute loss in achievable rate due to quantization is on the order of a constant, which increases with SNR. Reducing the feedback to $O(\log N)$ bits per codeword reduces the growth in capacity to $O(\log\log N)$. These results apply to adaptive modulation schemes, in which each active sub-channel is assigned a quantized rate with the same power, as well as coding schemes, which are applied across sub-channels.

With correlated sub-channel gains, less feedback is needed to achieve the optimal growth in achievable rate, relative to *i.i.d.* sub-channels. For the first-order autoregressive Rayleigh fading model considered, the ratio of
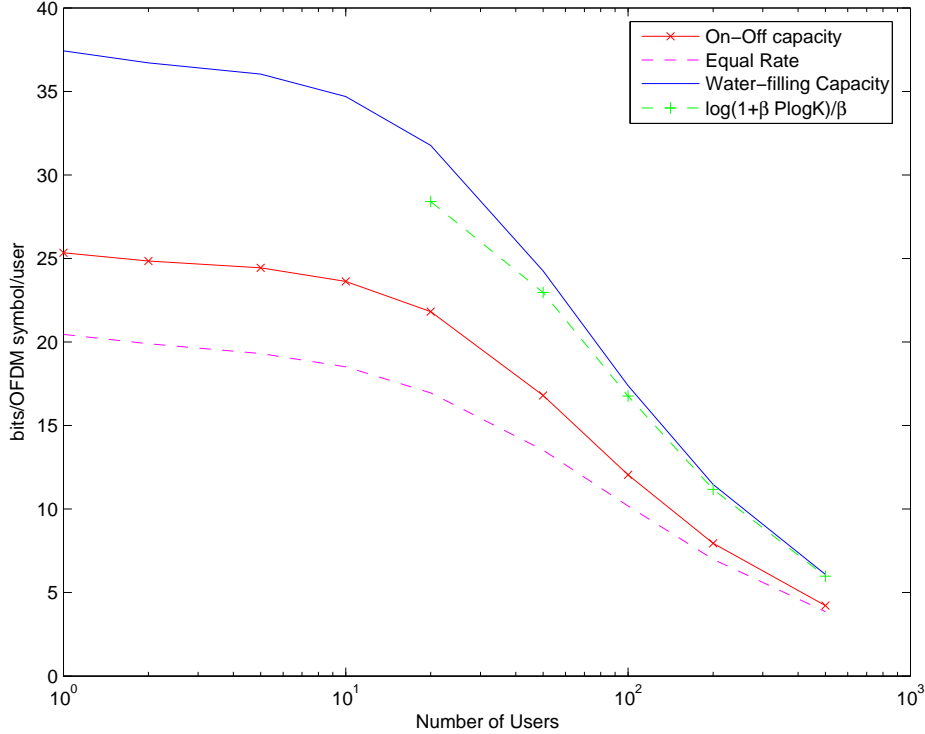
Fig. 9. Average downlink sum capacity *per user* versus number of users with on-off and water-filling power allocations. Numerical results for finite-size systems and asymptotic (large system) results are shown.

feedback for the two cases tends to zero as $O\left(\sqrt{\frac{\log N}{N}}\right)$, assuming optimized thresholds, and that the ratio of coherence bandwidth to the total available bandwidth is fixed. It is easy to verify that this order reduction in feedback also applies to a higher-order autoregressive model, although the associated constant becomes more difficult to compute.

We have also extended the preceding results to multi-user uplink and downlink channels, where each user feeds back the set of sub-channels with gains that exceed a threshold. If K is small relative to the number of sub-channels N, then the sum capacity scales linearly with the number of users K. If K and N tend to infinity with fixed ratio, then for Rayleigh fading the sum capacity per sub-channel (with limited feedback) increases as $O(\log \log K)$. These results apply to both the uplink and downlink.

All of these results assume that the feedback is used to specify an on-off power allocation, which achieves the optimal growth in capacity for a class of channel gain distributions of interest. Although this type of feedback scheme is relatively simple and practical, we have not shown that other feedback schemes cannot perform better. Namely, in general a quantized version of the received signal could be fed back to the transmitter. The capacity of such a channel model is unknown.**

**The classical information theoretic model with feedback assumes that the transmitter knows the channel (e.g., see [36], [37]).

We have also assumed that the receiver has perfect knowledge of the channel. Of course, the feedback must be determined from channel estimates. If the channel coherence time is sufficiently long, then accurate channel estimates can be obtained with a relatively small expenditure of transmitted resources, so that the results presented here apply. However, with a short coherence time, both the amount of transmitted resources devoted to channel estimation and the channel estimation error must be taken into account when computing the achievable rate. Related results for a block fading channel model are presented in [34], [35].

Finally, a natural extension of the model and results presented here is to add multiple antennas, and also multiple users with different rate requests. In those scenarios, the multi-carrier threshold feedback scheme considered here can be combined with the vector quantization schemes for MIMO channels described in [15], [38]. (Such a scheme for downlink beamforming is analyzed in [40].) The benefits of limited feedback with different user configurations (e.g., uplink and downlink) remain to be determined.

Acknowledgment: The authors thank Randall Berry and Jieying Chen for helpful discussions concerning this work, and Jieying Chen for her assistance with the proof of Theorem 9.

<center>APPENDIX</center>

A. Proof of Theorem 1

1) On-Off Power Allocation: We first determine the asymptotic behavior of the capacity with on-off feedback. With some abuse of notation, here we let C_i denote the capacity for the ith sub-channel. Given a threshold μ_0, we have

$$C_i = \log\left(1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}\mu_i\right)\mathbf{1}_{\mu_i > \mu_0},$$

which is a random variable. We first show that $C_N^{\text{on-off}} \asymp E[C_i]$ as $N \to \infty$. We have

$$
\begin{aligned}
E[C_i^2] &= \int_{\mu_0}^{\infty} \log^2\left(1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}x\right)f_\mu(x)dx \\
&\leq \log^2\left(1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}\mu_0\right)\bar{F}_\mu(\mu_0) + \frac{2\mathcal{P}}{N}\int_{\mu_0}^{\infty}\frac{\bar{F}_\mu(x)}{\bar{F}_\mu(\mu_0)}\frac{\frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}x}{1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}x}dx \\
&\leq \log^2\left(1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}\mu_0\right)\bar{F}_\mu(\mu_0) + \frac{2\mathcal{P}}{N}\cdot c_1
\end{aligned}
$$

where c_1 is a finite constant. The first inequality is obtained by integrating by parts and by using the bound $\log(1 + ax) \leq ax$. The second inequality follows from the observations that for any $x > 0$, $\frac{\frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}x}{1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}x} \leq 1$ and

$$
\begin{aligned}
E[\mu|\mu > x] - x &= \int_x^{\infty} y\frac{f_\mu(y)}{\bar{F}_\mu(y)}dy - x \\
&= \int_x^{\infty} \frac{\bar{F}_\mu(y)}{\bar{F}_\mu(x)}dy \\
&\leq \int_x^{\infty} \frac{c_2 \cdot f_\mu(y)}{\bar{F}_\mu(x)}dy < \infty
\end{aligned}
$$

where c_2 is a finite constant.

<center>23</center>

We also have

$$\begin{aligned}
E[C_i] &= \int_{\mu_0}^{\infty} \log\left(1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}x\right) f_\mu(x)dx \\
&\geq \int_{\mu_0}^{\infty} \log\left(1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}\mu_0\right) f_\mu(x)dx \\
&= \log\left(1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}\mu_0\right)\bar{F}_\mu(\mu_0).
\end{aligned}$$

Since the capacities across sub-channels are *i.i.d.*, according to the central limit theorem, the on-off capacity is Gaussian, i.e., $C \sim \mathcal{N}(NE[C_i], N\sigma^2)$ where $\sigma^2 \leq E[C_i^2]$. From the Chebyshev inequality we have that

$$Pr\left[\left|\frac{C}{E[C]} - 1\right|^2 \geq m\right] \leq \frac{N\sigma^2}{m\,(NE[C_i])^2} = \frac{1}{N\bar{F}_\mu(\mu_0)m} + O\left(\frac{1}{\mu_0^2}\right). \tag{53}$$

From Lemma 1, to achieve the capacity the expected number of active sub-channels $N\bar{F}_\mu(\mu_0) \to \infty$ as $N \to \infty$. Furthermore, the threshold $\mu_0 \to \infty$ with N, since otherwise the capacity per sub-channel would remain bounded. Hence the right side of (53) tends to zero with N, and $C \asymp E[C]$.

We now characterize the optimized threshold μ_0 as $N \to \infty$. The mean capacity is upper bounded as

$$\begin{aligned}
E[C] &= N\int_{\mu_0}^{\infty} \log\left(1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}x\right) f_\mu(x)dx \\
&= N\bar{F}_\mu(\mu_0)\int_{\mu_0}^{\infty} \log\left(1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}x\right) f_{\mu|\mu>\mu_0}(x)dx \\
&\leq \bar{C} \equiv N\log\left(1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}M(\mu_0)\right)\bar{F}_\mu(\mu_0)
\end{aligned}$$

where $M(\mu_0) \equiv E[\mu|\mu > \mu_0]$. After some manipulation we can calculate

$$\begin{aligned}
\frac{d\bar{C}}{d\mu_0} &= \frac{\frac{\mathcal{P}}{\bar{F}_\mu(\mu_0)}f_\mu(\mu_0)(2M(\mu_0)-\mu_0)}{1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}M(\mu_0)} - N\log\left(1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}M(\mu_0)\right)f_\mu(\mu_0) \\
&= \frac{\frac{\mathcal{P}f_\mu(\mu_0)}{\bar{F}_\mu(\mu_0)}}{1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}M(\mu_0)}\left[\left(M(\mu_0) - \mu_0 - \frac{1}{2}\frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}M^2(\mu_0)\right) + \frac{1}{2}\frac{\mathcal{P}M^2(\mu_0)}{N\bar{F}_\mu(\mu_0)}\frac{\mathcal{P}M(\mu_0)}{N\bar{F}_\mu(\mu_0)}\right] \\
&\quad + O\left(N\left(\frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}M(\mu_0)\right)^3 f_\mu(\mu_0)\right)
\end{aligned}$$

where $q(N) = O(g(N))$ denotes a term for which $\lim_{N\to\infty} q(N)/g(N)$ is a finite constant.

If μ_0^* satisfies (8), then

$$\begin{aligned}
\left.\frac{d\bar{C}}{d\mu_0}\right|_{\mu_0=\mu_0^*} &= \frac{\mathcal{P}\frac{f_\mu(\mu_0^*)}{\bar{F}_\mu(\mu_0^*)}}{1 + \frac{2[M(\mu_0^*)-\mu_0^*]}{M(\mu_0^*)}}\frac{2[M(\mu_0^*)-\mu_0^*]^2}{M(\mu_0^*)} + O\left(4\mathcal{P}\frac{f_\mu(\mu_0^*)}{\bar{F}_\mu(\mu_0^*)}\frac{(M(\mu_0^*)-\mu_0^*)^2}{M(\mu_0^*)}\right) \\
&= O\left(\frac{f_\mu(\mu_0^*)}{\bar{F}_\mu(\mu_0^*)}\frac{1}{M(\mu_0^*)}\right) = O\left(\frac{1}{M(\mu_0^*)}\right)
\end{aligned}$$

where the second equality follows from noting that $M(\mu_0^*) - \mu_0^* < \infty$ from (53) and $M(\mu_0^*) \geq \mu_0^* \to \infty$ as $N \to \infty$. To show that $|\max_\mu \bar{C}(\mu) - \bar{C}(\mu_0^*)| \to 0$ as $N \to \infty$, it is sufficient to show that $|d^2\bar{C}/d\mu_0^2|$ is bounded away from zero in a neighborhood of μ^*. We then have

$$\max_{\mu_0} E[C] \leq \max_{\mu_0}\bar{C} = N\log\left(1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0^*)}M(\mu_0^*)\right)\bar{F}_\mu(\mu_0^*) + O\left(\frac{1}{M(\mu_0^*)}\right) = \mathcal{P}\mu_0^* + O\left(\frac{1}{M(\mu_0^*)}\right).$$

24

Furthermore,

$$
\begin{aligned}
\max_{\mu_0} E[C] &\geq N \int_{\mu_0^*}^{\infty} \log\left(1 + \frac{\mathcal{P}}{N\bar{F}_\mu(x)}x\right) f_\mu(x)dx \\
&\geq N \log\left(1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0^*)}\mu_0^*\right) \bar{F}_\mu(\mu_0^*) \\
&= \mathcal{P}\mu_0^* + O\left(\frac{\mathcal{P}^2\mu_0^{*2}}{N\bar{F}_\mu(\mu_0^*)}\right) \\
&= \mathcal{P}\mu_0^* + O(1),
\end{aligned}
$$

so that $\lim_{N\to\infty} \frac{\max E[C]}{\mathcal{P}\mu_0^*} \to 1$, or $C^{\text{(on-off)}} \asymp \mathcal{P}\mu_0^*$.

2) Water-Filling: As for the on-off power allocation, we first show that $C \asymp E[C]$, and then show that $E[C] \asymp \mathcal{P}\mu_0^*$. The capacity for the ith sub-channel is

$$
C_i = \log(\lambda\mu_i)\mathbf{1}_{\mu_i > \frac{1}{\lambda}}.
$$

Since $\log^2(ax)$ is a concave function for $x \geq \frac{1}{a}$, we have

$$
\begin{aligned}
E[C_i^2] &= \int_{1/\lambda}^{\infty} \log^2(\lambda x) f_\mu(x)dx \leq \log^2\left(\lambda M(1/\lambda)\right) \bar{F}_\mu\left(\frac{1}{\lambda}\right) \\
&\leq \left(\lambda(M(1/\lambda) - \frac{1}{\lambda})\right)^2 \bar{F}_\mu\left(\frac{1}{\lambda}\right) = O\left(\lambda^2\bar{F}_\mu\left(\frac{1}{\lambda}\right)\right)
\end{aligned}
$$

and

$$
\begin{aligned}
E[C_i] &= \int_{1/\lambda}^{\infty} \log(\lambda x) f_\mu(x)dx = \int_{1/\lambda}^{\infty} \frac{\bar{F}_\mu(x)}{x}dx \\
&\geq \int_{1/\lambda}^{\infty} \frac{\text{constant} \cdot f_\mu(x)}{x}dx \\
&\geq O\left(\frac{\bar{F}_\mu(\frac{1}{\lambda})}{M(1/\lambda)}\right).
\end{aligned}
$$

As for the on-off case, the total capacity summed over sub-channels is Gaussian, and the Chebyshev inequality implies that

$$
\begin{aligned}
Pr\left[\left|\frac{C_i}{E[C_i]} - 1\right|^2 \geq m\right] &\leq \frac{N\sigma^2}{m(NE[C_i])^2} = O\left(\frac{(\lambda E)^2}{N\bar{F}_\mu(\frac{1}{\lambda})}\right) \\
&= O\left(\frac{(1 + \lambda(E - \frac{1}{\lambda}))^2}{N\bar{F}_\mu(\frac{1}{\lambda})}\right) = O\left(\frac{1}{N\bar{F}_\mu(\frac{1}{\lambda})}\right).
\end{aligned}
$$

From Lemma 1, the expected number of active sub-channels $N\bar{F}_\mu(\frac{1}{\lambda}) \to \infty$ as $N \to \infty$. Hence $C_i/E[C_i] \to 1$ with probability one, so that $C_i \asymp E[C_i]$.

To determine the asymptotic mean capacity we first write the power constraint (3) as

$$
\frac{\mathcal{P}}{N} = \int_{1/\lambda}^{\infty} \left(\lambda - \frac{1}{x}\right) f_\mu(x)dx = \lambda\bar{F}_\mu\left(\frac{1}{\lambda}\right) - \int_{1/\lambda}^{\infty} \frac{f_\mu(x)}{x}dx.
$$

Denote $t = M(1/\lambda) - 1/\lambda$. From Jensen's inequality, we have $\int_{1/\lambda}^{\infty} \frac{f_\mu(x)}{x}dx \geq \bar{F}_\mu(1/\lambda)/M(1/\lambda)$, so that

$$
\frac{\mathcal{P}}{N} \leq (\lambda - \frac{1}{E})\bar{F}_\mu\left(\frac{1}{\lambda}\right) = \frac{t\lambda^2}{1 + t\lambda}\bar{F}_\mu\left(\frac{1}{\lambda}\right) \leq t\lambda^2\bar{F}_\mu\left(\frac{1}{\lambda}\right). \tag{54}
$$

Let $J(x) = \frac{\bar{F}_\mu(x)(M(x)-x)}{x^2}$, and note that

$$\frac{dJ(x)}{dx} = -\frac{\bar{F}_\mu(x)}{x^2} - \frac{2}{x}J(x) < 0,$$

i.e., $J(x)$ is a decreasing function. From (8) and (54), for large N we have

$$J(\mu_0^*) = \frac{\mathcal{P}}{2N} + O\left(\frac{\mathcal{P}}{NE[\mu|\mu > \mu_0^*]}\right) < \frac{\mathcal{P}}{N} \leq J\left(\frac{1}{\lambda}\right)$$

so that $\frac{1}{\lambda} \leq \mu_0^*$.

Now let $L(\lambda) = \frac{\lambda}{aM(1/\lambda)\lambda+1-a}\bar{F}_\mu(\frac{1}{\lambda}) - \int_{1/\lambda}^\infty \frac{f_\mu(x)}{x}dx$, where $0 \leq a \leq 1$. Then we can calculate

$$\frac{dL(\lambda)}{d\lambda} = \frac{1-a}{(1+at\lambda)^2}\bar{F}_\mu\left(\frac{1}{\lambda}\right) - \frac{at^2\lambda}{(1+at\lambda)^2}f_\mu\left(\frac{1}{\lambda}\right)$$

and since $\frac{\bar{F}_\mu(x)}{f_\mu(x)} \geq m > 0$, we have

$$\frac{dL(\lambda)}{d\lambda} \geq \frac{(1-a)m - at^2\lambda}{(1+at\lambda)^2}f_\mu\left(\frac{1}{\lambda}\right) \geq 0$$

for $0 \leq \lambda \leq \frac{(1-a)m}{at^2}$. Since $L(0) = 0$, as $N \to \infty$, $\lambda \to 0$ and $a \to 1$, and we have $\frac{\lambda}{1+t\lambda}\bar{F}_\mu(\frac{1}{\lambda}) \geq \int_{1/\lambda}^\infty \frac{f_\mu(x)}{x}dx$. Hence, for large enough N

$$\frac{\mathcal{P}}{N} \geq \lambda\bar{F}_\mu\left(\frac{1}{\lambda}\right) - \frac{\lambda}{1+t\lambda}\bar{F}_\mu\left(\frac{1}{\lambda}\right) = \frac{t\lambda^2}{1+t\lambda}\bar{F}_\mu\left(\frac{1}{\lambda}\right).$$

The capacity is then upper bounded as

$$\begin{aligned}
E[C] &= N\int_{1/\lambda}^\infty \log(\lambda x)f_\mu(x)dx \leq N\int_{1/\lambda}^\infty (\lambda x - 1)f_\mu(x)dx \\
&= N\left[\lambda M(1/\lambda)\bar{F}_\mu\left(\frac{1}{\lambda}\right) - \bar{F}_\mu\left(\frac{1}{\lambda}\right)\right] = Nt\lambda\bar{F}_\mu\left(\frac{1}{\lambda}\right).
\end{aligned}$$

Hence we have

$$\frac{E[C]}{\mathcal{P}} \leq \frac{1}{\lambda} + t \leq \mu_0^*,$$

which follows because both $\frac{1}{\lambda}$ and $\mu_0^* \to \infty$ with $\frac{1}{\lambda} \leq \mu_0^*$, and $t < \infty$ (from (53)). Therefore, $E[C^{(\mathrm{wf})}] \leq \mathcal{P}\mu_0^*$. Since water-filling is optimal, we have $E[C^{(\mathrm{wf})}] \geq E[C^{(\mathrm{on\text{-}off})}]$. Therefore, $C^{(\mathrm{wf})} \asymp C^{(\mathrm{on\text{-}off})} \asymp \mathcal{P}\mu_0^*$, where μ_0^* is the solution to (8). $\blacksquare$

B. Proof of Theorem 2

We have

$$E[R_n^{(\mathrm{fp})}] = N\sum_{i=0}^{n-1} \log\left(1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}\nu_{n,i}\right)\left(\bar{F}_\mu(\nu_{n,i}) - \bar{F}_\mu(\nu_{n,i+1})\right)$$

where $\nu_{n,n} = \infty$ and $\bar{F}_\mu(\nu_{n,n}) = 0$. We can calculate

$$
\begin{aligned}
\frac{\partial E[R_n^{\text{(fp)}}]}{\partial \nu_{n,n-1}} &= N \frac{\frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)} f_\mu(\nu_{n,n-1})}{1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}\nu_{n,n-1}} \left[\frac{\bar{F}_\mu(\nu_{n,n-1})}{f_\mu(\nu_{n,n-1})} - (\nu_{n,n-1} - \nu_{n,n-2}) \right] \\
&\quad + O\left(N f_{\mu_0}(\nu_{n,n-1}) \left(\frac{\frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}(\nu_{n,n-1} - \nu_{n,n-2})}{1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}\nu_{n,n-1}} \right)^2 \right) \\
&= \frac{\frac{\mathcal{P}}{\bar{F}_\mu(\mu_0)} f_\mu(\nu_{n,n-1})}{1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}\nu_{n,n-1}} \left[\frac{\bar{F}_\mu(\nu_{n,n-1})}{f_\mu(\nu_{n,n-1})} - (\nu_{n,n-1} - \nu_{n,n-2}) \right] + O\left(N f_{\mu_0}(\nu_{n,n-1}) \left(\frac{\nu_{n,n-1} - \nu_{n,n-2}}{\nu_{n,n-1}} \right)^2 \right)
\end{aligned}
$$

If $\nu_{n,n-1} - \nu_{n,n-2} = \frac{\bar{F}_\mu(\nu_{n,n-1})}{f_\mu(\nu_{n,n-1})}$, then

$$
\begin{aligned}
\frac{\partial E[R_N]}{\partial \nu_{n,n-1}} &= O\left(N f_\mu(\nu_{n,n-1}) \left(\frac{\bar{F}_\mu(\nu_{n,n-1})}{f_\mu(\nu_{n,n-1})} \right)^2 \frac{1}{\nu_{n,n-1}^2} \right) \\
&= O\left(\frac{N\bar{F}_\mu(\nu_{n,n-1})}{\nu_{n,n-1}^2} \frac{\bar{F}_\mu(\nu_{n,n-1})}{f_\mu(\nu_{n,n-1})} \right) \to 0
\end{aligned}
$$

provided that the threshold μ_0 is chosen so that the average number of active sub-channels $N\bar{F}_\mu(\mu_0) \to \infty$ as $N \to \infty$.

Similarly, for $1 \le i \le n-2$, we can calculate

$$
\begin{aligned}
\frac{\partial E[R_n^{\text{(fp)}}]}{\partial \nu_{n,i}} &= \frac{\frac{\mathcal{P}}{\bar{F}_\mu(\mu_0)} f_\mu(\nu_{n,i})}{1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}\nu_{n,i}} \left[\frac{\bar{F}_\mu(\nu_{n,i}) - \bar{F}_\mu(\nu_{n,i+1})}{f_\mu(\nu_{n,i})} - (\nu_{n,i} - \nu_{n,i-1}) \right] \\
&\quad + O\left(N f_\mu(\nu_{n,i}) \left(\frac{\frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}(\nu_{n,i} - \nu_{n,i-1})}{1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}\nu_{n,i}} \right)^2 \right) \\
&= \frac{\frac{\mathcal{P}}{\bar{F}_\mu(\mu_0)} f_\mu(\nu_{n,i})}{1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}\nu_{n,i}} \left[\frac{\bar{F}_\mu(\nu_{n,i}) - \bar{F}_\mu(\nu_{n,i+1})}{f_\mu(\nu_{n,i})} - (\nu_{n,i} - \nu_{n,i-1}) \right] + O\left(N f_\mu(\nu_{n,i}) \left(\frac{\nu_{n,i} - \nu_{n,i-1}}{\nu_{n,i}} \right)^2 \right)
\end{aligned}
$$

and letting $\nu_{n,i} - \nu_{n,i-1} = \frac{\bar{F}_\mu(\nu_{n,i}) - \bar{F}_\mu(\nu_{n,i+1})}{f_\mu(\nu_{n,i})}$, it is easy to show that

$$
\frac{\partial E[R_n^{\text{(fp)}}]}{\partial \nu_{n,i}} = O\left(\frac{N(\bar{F}_\mu(\nu_{n,i}) - \bar{F}_\mu(\nu_{n,i+1}))}{\nu_{n,i}^2} \frac{\bar{F}_\mu(\nu_{n,i}) - \bar{F}_\mu(\nu_{n,i+1})}{f_\mu(\nu_{n,i})} \right) \to 0
$$

We can then show that $|\max_{\nu_{n,i}} R_n^{\text{(fp)}}(\nu_{n,i}) - R_n^{\text{(fp)}}(\nu_{n,i}^*)| \to 0$, where $\nu_{n,i}^*$ satisfies (9), by showing that $|\partial^2 R_n^{\text{(fp)}}/\partial \nu_{n,i}^2|$ is bounded away from zero in a neighborhood of $\nu_{n,i}^*$. Hence $\bar{R}_n^{\text{(fp)}} \asymp E[R_n^{\text{(fp)}}]$.

Furthermore, we have

$$
\begin{aligned}
E[R_n^{\text{(fp)}} - R_1^{\text{(fp)}}] &= N\left(\sum_{i=0}^{n-1}\log(1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}\nu_{n,i})(\bar{F}_\mu(\nu_{n,i}) - \bar{F}_\mu(\nu_{n,i+1})) - \log(1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}\mu_0)\bar{F}_\mu(\mu_0)\right) \\
&= N\sum_{i=1}^{n-1}\left(\log(1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}\nu_{n,i}) - \log(1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}\nu_{n,i-1})\right)\bar{F}_\mu(\nu_{n,i}) \\
&= N\sum_{i=1}^{n-1}\log\left(1 + \frac{\frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}(\nu_{n,i} - \nu_{n,i-1})}{1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}\nu_{n,i-1}}\right)\bar{F}_\mu(\nu_{n,i}) \\
&\leq N\sum_{i=1}^{n-1}\frac{\frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}(\nu_{n,i} - \nu_{n,i-1})}{1 + \frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}\nu_{n,i-1}}\bar{F}_\mu(\nu_{n,i}) \\
&\leq N\sum_{i=1}^{n-1}\frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}(\nu_{n,i} - \nu_{n,i-1})\bar{F}_\mu(\nu_{n,i}) \\
&= N\sum_{i=1}^{n-1}\frac{\mathcal{P}}{N\bar{F}_\mu(\mu_0)}(\bar{F}_\mu(\nu_{n,i}) - \bar{F}_\mu(\nu_{n,i+1}))\frac{\bar{F}_\mu(\nu_{n,i})}{f_\mu(\nu_{n,i})} \\
&\leq \mathcal{P}\frac{\bar{F}_\mu(\nu_{n,1})}{\bar{F}_\mu(\mu_0)}\max_x\frac{\bar{F}_\mu(x)}{f_\mu(x)} \\
&\leq \mathcal{P}\max_x\frac{\bar{F}_\mu(x)}{f_\mu(x)}
\end{aligned}
\tag{55}
$$

∎

C. Proof of Theorem 3

Since $\mu_N^{(1)} = \max\{\mu_1, \cdots, \mu_N\}$, and the sub-channel gains are assumed to be *i.i.d.*,

$$
\Pr\{\mu_N^{(1)} < x\} = [\Pr\{\mu < x\}]^N \to 0,
$$

i.e., the distribution is degenerate as $N \to \infty$. According to [28, Theorems 3.3 and 7.1], as $N \to \infty$,

$$
\begin{aligned}
\Pr\{\mu_N^{(1)} - \log N < x\} &\to \exp\{-e^{-x}\} \\
\Pr\{\mu_N^{(k)} - \log N < x\} &\to \exp\{-e^{-x}\}\sum_{i=0}^{k-1}\frac{e^{-ix}}{i!}.
\end{aligned}
$$

Since

$$
\frac{d}{dx}\left(\exp\{-e^{-x}\}\sum_{i=0}^{k-1}\frac{e^{-ix}}{i!}\right) = \exp\{-e^{-x}\}\frac{e^{-kx}}{(k-1)!},
$$

as $N \to \infty$, we have

$$
\begin{aligned}
E\{\mu_N^{(k)} - \log N\} &\to \int_{-\infty}^{\infty} x \exp\{-e^{-x}\} \frac{e^{-kx}}{(k-1)!} dx \\
&\qquad (z = e^{-x}, dx = \frac{dz}{z}) \\
&= -\int_0^{\infty} (\log z) \underbrace{e^{-z} \frac{z^{k-1}}{(k-1)!}}_{\Gamma \text{ distribution}} dz \\
&\geq -\log \left(\int_0^{\infty} z e^{-z} \frac{z^{k-1}}{(k-1)!} dz \right) \\
&= -\log(k-1).
\end{aligned}
\tag{56}
$$

where the inequality comes from the Jensen's Inequality since $\log(x)$ is a concave function. Also, we have

$$
E\{\mu_N^{(k)} - \log N\} \leq E\{\mu_N^{(1)} - \log N\} \to \int_{-\infty}^{\infty} x \exp\{-e^{-x}\} e^{-x} dx = c
\tag{57}
$$

where c is the Euler-Mascheroni constant. Combining (56) and (57) implies $E(\mu_N^{(k)}) \asymp \log N$.

It is easy to show that for $z \geq 1$, $0 \leq \log z < \sqrt{z}$, or $(\log z)^2 < z$. Then letting $y = \frac{1}{z}$ gives $(\log y)^2 < \frac{1}{y}$ for $0 \leq y \leq 1$. Therefore for $k \geq 2$ we have

$$
\begin{aligned}
E\{[\mu_N^{(k)} - \log N]^2\} &\to \int_{-\infty}^{\infty} x^2 \exp\{-e^{-x}\} \frac{e^{-kx}}{(k-1)!} dx \\
&= \int_0^{\infty} (\log z)^2 e^{-z} \frac{z^{k-1}}{(k-1)!} dz \\
&\leq \int_0^1 \frac{1}{z} e^{-z} \frac{z^{k-1}}{(k-1)!} dz + \int_1^{\infty} z e^{-z} \frac{z^{k-1}}{(k-1)!} dz \\
&\leq \int_0^1 e^{-z} dz + \int_0^{\infty} z e^{-z} \frac{z^{k-1}}{(k-1)!} dz \\
&= 1 - e^{-1} + k,
\end{aligned}
\tag{58}
$$

and for $k = 1$, $E\{[\mu_N^{(1)} - \log N]^2\} = 2c^2 + \frac{\pi}{6}$. Therefore, $\text{var}\left(\frac{\mu^{(k)}}{\log N}\right) \to 0$, and

$$
\lim_{N \to \infty} \frac{\mu^{(k)}}{\log N} = \lim_{N \to \infty} E\left[\frac{\mu^{(k)}}{\log N}\right] = 1
$$

in the mean square sense, which implies that

$$
\log(\mu_N^{(k)}) - \log\log N = \log\left(\frac{\mu_N^{(k)}}{\log N}\right) \to 0.
\tag{59}
$$

Therefore from Lemma 1

$$
C_N^{(N_a)} - \sum_{k=1}^{N_a} \log \mu^{(k)} \to C_N^{(N_a)} - N_a \log\log N = \sum_{i=1}^{N_a} \log P_i + o(1) \leq N_a \log\left(\frac{\mathcal{P}}{N_a}\right) + o(1).
\tag{60}
$$

29

D. Proof of Theorem 4

1) When no channel information is available at the transmitter, the power is spread uniformly across all sub-channels, so that

$$C_N = \sum_{i=1}^{N} \log\left(1 + \frac{\mathcal{P}}{N}\mu_i\right) = \sum_{i=1}^{N} \frac{\mathcal{P}}{N}\mu_i + O\left(\frac{\mu_i^2}{N^2}\right).$$

As $N \to \infty$, $C_N \to \mathcal{P}E[\mu] + O\left(\frac{E[\mu^2]}{N}\right) \to \mathcal{P}$.

2) As shown in Theorem 3, we have $\lim_{N\to\infty}\left(C_N^{(N_a)} - N_a \log\log N\right) = N_a \log\left(\frac{\mathcal{P}}{N_a}\right)$. Also, we have

$$R_1^{(\text{fp},N_a)} = N_a \log\left(1 + \frac{\mathcal{P}}{N_a}\mu_N^{(N_a)}\right) = N_a \log\left(\mu_N^{(N_a)}\right) + N_a \log\left(\frac{\mathcal{P}}{N_a}\right) + o(1).$$

From (59), it is easy to show that

$$R_1^{(\text{fp},N_a)} - N_a \log\log N \to N_a \log\left(\frac{\mathcal{P}}{N_a}\right),$$

hence $\lim_{N\to\infty}(C_N^{(N_a)} - R_1^{(\text{fp},N_a)}) = 0$.

3) With Rayleigh fading and the optimal threshold $\nu_{n,0} = \mu_0^*$, which satisfies (11), (55) becomes

$$
\begin{aligned}
E[R_n^{(\text{fp})} - R_1^{(\text{fp})}] &= N\sum_{i=1}^{n-1} \frac{\frac{\mathcal{P}}{Ne^{-\mu_0}}(\nu_{n,i} - \nu_{n,i-1})}{1 + \frac{\mathcal{P}}{Ne^{-\mu_0}}\nu_{n,i-1}}e^{-\nu_{n,i}} + O\left(\frac{Ne^{-\nu_{n,i}}}{\nu_{n,i-1}^2}\right) \\
&= N\sum_{i=1}^{n-1} \frac{\mathcal{P}}{Ne^{-\mu_0}}(\nu_{n,i} - \nu_{n,i-1})e^{-\nu_{n,i}} + O\left(\frac{1}{\mu_0}\right) \\
&\to \mathcal{P}e^{-(\nu_{n,1}-\mu_0)}
\end{aligned}
$$

Since $C^{(\text{on-off})} = R_\infty^{(\text{fp})}$, corresponding to infinite-precision rate control, and $\nu_{\infty,1} \to \mu_0$, it follows that $E[C^{(\text{on-off})} - R_n^{(\text{fp})}] = \mathcal{P}(1 - e^{-(\nu_{n,1}-\mu_0)})$. We now show that $\text{var}(C^{(\text{on-off})} - R_n^{(\text{fp})}) \to 0$ as $N \to \infty$. Namely, we have

$$
\begin{aligned}
NE^2[R_n^{(\text{fp})}] &= N\left(\log\left(1 + \frac{\mathcal{P}}{Ne^{-\mu_0}}\mu_0\right)e^{-\mu_0} + \frac{\mathcal{P}e^{-(\nu_{n,1}-\mu_0)}}{N}\right)^2 \\
&= \frac{\left(\mathcal{P}(\mu_0 + e^{-(\nu_{n,1}-\mu_0)})\right)^2}{N} \to 0
\end{aligned}
$$

and

$$\begin{aligned}
NE[R_n^{\text{(fp)}2}] &= N\sum_{i=0}^{n-1}\log^2(1+\frac{\mathcal{P}}{Ne^{-\mu_0}}\nu_{n,i})(e^{-\nu_{n,i}}-e^{-\nu_{n,i+1}}) \\
&= N\sum_{i=0}^{n-1}\left(\frac{\mathcal{P}}{Ne^{-\mu_0}}\nu_{n,i}\right)^2 e^{-\nu_{n,i}}\left(1-e^{-(\nu_{n,i+1}-\nu_{n,i})}\right)+O\left(\frac{1}{\mu_0^2}\right) \\
&= \frac{\mathcal{P}^2 e^{\mu_0}\mu_0^2}{N}\sum_{i=0}^{n-1}e^{\mu_0-\nu_{n,i}}\left(1-e^{-(\nu_{n,i+1}-\nu_{n,i})}\right)+O\left(\frac{1}{\mu_0}\right) \\
&\to 2\mathcal{P}\sum_{i=1}^{n-1}(\nu_{n,i}-\nu_{n,i-1})e^{\mu_0-\nu_{n,i}}+2\mathcal{P}(1-e^{\mu_0-\nu_{n,1}}) \\[2mm]
&= 2\mathcal{P}\left[1-e^{\mu_0-\nu_{n,1}}+\sum_{i=1}^{n-3}(\nu_{n,i}-\nu_{n,i-1})e^{\mu_0-\nu_{n,i}}+(\nu_{n,n-1}-\nu_{n,n-2})e^{\mu_0-\nu_{n,n-1}}\right. \\
&\qquad\left.+(\nu_{n,n-2}-\nu_{n,n-3})e^{\mu_0-\nu_{n,n-2}}\right] \\
&= 2\mathcal{P}\left[1-e^{\mu_0-\nu_{n,1}}+\sum_{i=1}^{n-3}(\nu_{n,i}-\nu_{n,i-1})e^{\mu_0-\nu_{n,i}}\right. \\
&\qquad\left.+e^{\mu_0-\nu_{n,n-2}}\left(1\cdot e^{\nu_{n,n-2}-\nu_{n,n-1}}+\nu_{n,n-2}-\nu_{n,n-3}\right)\right] \\
&= 2\mathcal{P}\left[1-e^{\mu_0-\nu_{n,1}}+\sum_{i=1}^{n-3}(\nu_{n,i}-\nu_{n,i-1})e^{\mu_0-\nu_{n,i}}+e^{\mu_0-\nu_{n,n-2}}\right]
\end{aligned}$$

Continuing to remove each term in the sum successively, corresponding to $i=n-3, n-2, \cdots$, and using (11), we can evaluate $NE[R_n^{\text{(fp)}2}]=2\mathcal{P}$. Therefore, $N\text{var}(R_n^{\text{(fp)}})=2\mathcal{P}$, and since the second moment is independent of the number of quantization levels, $N\text{var}(C^{\text{(on-off)}})=2\mathcal{P}$.

We also need to evaluate

$$\begin{aligned}
NE[R_n^{\text{(fp)}}C^{\text{(on-off)}}] &= N\sum_{i=0}^{n-1}\int_{\nu_{n,i}}^{\nu_{n,i+1}}\log(1+\frac{\mathcal{P}}{N}e^{\mu_0}x)\log(1+\frac{\mathcal{P}}{N}e^{\mu_0}\nu_{n,i})e^{-x}dx \\
&= N\sum_{i=0}^{n-1}\log^2(1+\frac{\mathcal{P}}{N}e^{\mu_0}\nu_{n,i})(e^{-\nu_{n,i}}-e^{-\nu_{n,i+1}}) \\
&\quad+N\sum_{i=0}^{n-1}\log(1+\frac{\mathcal{P}}{N}e^{\mu_0}\nu_{n,i})\int_{\nu_{n,i}}^{\nu_{n,i+1}}\log\left(1+\frac{\frac{\mathcal{P}}{N}e^{\mu_0}(x-\nu_{n,i})}{1+\frac{\mathcal{P}}{N}e^{\mu_0}\nu_{n,i}}\right)e^{-x}dx \\
&= 2\mathcal{P}+N\sum_{i=0}^{n-1}\log(1+\frac{\mathcal{P}}{N}e^{\mu_0}\nu_{n,i})\int_{\nu_{n,i}}^{\nu_{n,i+1}}\frac{\mathcal{P}}{N}e^{\mu_0}(x-\nu_{n,i})e^{-x}dx \\
&= 2\mathcal{P}+O\left(\frac{1}{\mu_0}\right)
\end{aligned}$$

and

$$\begin{aligned}
NE[R_n^{\text{(fp)}}]E[C^{\text{(on-off)}}] &= N\left(\log\left(1+\frac{\mathcal{P}}{Ne^{-\mu_0}}\mu_0\right)e^{-\mu_0}+\frac{\mathcal{P}}{N}\right)\left(\log\left(1+\frac{\mathcal{P}}{Ne^{-\mu_0}}\mu_0\right)e^{-\mu_0}+\frac{\mathcal{P}e^{-(\nu_{n,1}-\mu_0)}}{N}\right) \\
&\to 0
\end{aligned}$$

Combining this gives

$$\sigma^2 = N\text{var}(R_n^{\text{(fp)}}) + N\text{var}(C^{\text{(on-off)}}) - 2N\left(E[R_n^{\text{(fp)}}C^{\text{(on-off)}}] - E[R_n^{\text{(fp)}}]E[C^{\text{(on-off)}}]\right) \to 0,$$

hence the loss in achievable rate converges to a constant $\mathcal{P}(1 - e^{-(\nu_{n,1} - \mu_0)})$ w.p.1. ∎

E. Proof of Theorem 5

Before looking at the three different regions, we first bound the capacity for a given threshold μ_0. The mean capacity can be written as

$$
\begin{aligned}
E[C_N] &= \int_{\mu_0}^{\infty} N\log\left(1 + \frac{\mathcal{P}}{Ne^{-x}}x\right)e^{-x}dx \\
&= N\log\left(1 + \frac{\mathcal{P}}{Ne^{-\mu_0}}\mu_0\right)e^{-\mu_0} + \int_{\mu_0}^{\infty}\frac{\mathcal{P}e^{\mu_0 - x}}{1 + \frac{\mathcal{P}}{Ne^{-\mu_0}}x}dx
\end{aligned}
$$

From Jensen's inequality, we have

$$
\begin{aligned}
E[C_N] &\geq N\log\left(1 + \frac{\mathcal{P}}{Ne^{-\mu_0}}\mu_0\right)e^{-\mu_0} + \frac{\mathcal{P}}{1 + \frac{\mathcal{P}}{Ne^{-\mu_0}}E[x|x > \mu_0]} \\
&= N\log\left(1 + \frac{\mathcal{P}}{Ne^{-\mu_0}}\mu_0\right)e^{-\mu_0} + \frac{\mathcal{P}}{1 + \frac{\mathcal{P}}{Ne^{-\mu_0}}(\mu_0 + 1)}
\end{aligned}
\tag{61}
$$

Furthermore,

$$
\begin{aligned}
E[C_N] &\leq N\log\left(1 + \frac{\mathcal{P}}{Ne^{-\mu_0}}\mu_0\right)e^{-\mu_0} + \frac{\mathcal{P}}{1 + \frac{\mathcal{P}}{Ne^{-\mu_0}}\mu_0}\int_{\mu_0}^{\infty}e^{\mu_0 - x}dx \\
&= N\log\left(1 + \frac{\mathcal{P}}{Ne^{-\mu_0}}\mu_0\right)e^{-\mu_0} + \frac{\mathcal{P}}{1 + \frac{\mathcal{P}}{Ne^{-\mu_0}}\mu_0}
\end{aligned}
\tag{62}
$$

Case 1. $\gamma < 1$: Since $\frac{\mathcal{P}}{Ne^{-\mu_0}}\mu_0^{\gamma} = \kappa$, we have from (61-62)

$$E\left[C_N^{(\gamma)}\right] = N\log\left(1 + \frac{\mathcal{P}}{Ne^{-\mu_0}}\mu_0\right)e^{-\mu_0} + O\left(\frac{1}{\mu_0^{1-\gamma}}\right)$$

Therefore

$$\frac{E\left[C_N^{(\gamma)}\right]}{\frac{\mathcal{P}(1-\gamma)}{\kappa}\mu_0^{\gamma}\log\mu_0} = \frac{\log\left(1 + \kappa\mu_0^{1-\gamma}\right)}{\log\mu_0^{1-\gamma}} \to 1.$$

Furthermore, we can expand (17) as

$$\mu_0 = \log\frac{\kappa N}{\mathcal{P}} - \gamma\log\left(\log\frac{\kappa N}{\mathcal{P}} - \gamma\log\left(\log\frac{\kappa N}{\mathcal{P}} - \cdots\right)\right)
\tag{63}$$

which gives

$$\lim_{N\to\infty}\frac{E\left[C_N^{(\gamma)}\right]}{\frac{1-\gamma}{\kappa}\mathcal{P}\log^{\gamma}N\log\log N} = 1.$$

Case 2. $\gamma = 1$: Now we have

$$
\begin{aligned}
E\left[C_N^{(\gamma)}\right] &= N\log\left(1 + \frac{\mathcal{P}}{Ne^{-\mu_0}}\mu_0\right)e^{-\mu_0} + \frac{\mathcal{P}}{1 + \kappa} + O\left(\frac{1}{\mu_0^{1-\gamma}}\right) \\
&= \frac{\log(1 + \kappa)}{\kappa}\mathcal{P}\mu_0 + \frac{\mathcal{P}}{1 + \kappa} + O\left(\frac{1}{\mu_0^{1-\gamma}}\right)
\end{aligned}
$$

32

so that

$$\lim_{N \to \infty} \frac{E\left[C_N^{(\gamma)}\right]}{\frac{\log(1+\kappa)}{\kappa} \mathcal{P} \log N} = 1$$

Case 3. $\gamma > 1$: In this case,

$$
\begin{aligned}
E\left[C_N^{(\gamma)}\right] &= N \log\left(1 + \frac{\mathcal{P}}{Ne^{-\mu_0}}\mu_0\right) e^{-\mu_0} + \mathcal{P} + O\left(\frac{1}{\mu_0^{\gamma-1}}\right) \\
&= \frac{\mathcal{P}\mu_0^{\gamma}}{\kappa} \log\left(1 + \frac{\kappa}{\mu_0^{\gamma-1}}\right) + \mathcal{P} + O\left(\frac{1}{\mu_0^{\gamma-1}}\right) \\
&= \mathcal{P}\mu_0 + O\left(\mu_0^{2-\gamma}\right)
\end{aligned}
$$

Therefore,

$$\lim_{N \to \infty} \frac{E\left[C_N^{(\gamma)}\right]}{\mathcal{P}\mu_0} = \lim_{N \to \infty} \frac{E\left[C_N^{(\gamma)}\right]}{\mathcal{P} \log N} = 1$$

Given the threshold μ_0, the average number of active sub-channels is

$$\bar{N}_a = Ne^{-\mu_0} = \frac{\mathcal{P}}{\kappa}\mu_0^{\gamma}.$$

If the rates on all active sub-channels transmit are equal, then from (61) the loss in the achievable data rate from the capacity is bounded by

$$E[C_N^{(\gamma)} - R_N] \leq \frac{\mathcal{P}}{1 + \frac{\mathcal{P}}{Ne^{-\mu_0}}\mu_0} \leq \frac{\mathcal{P}}{1 + \kappa\mu_0^{1-\gamma}} \leq \mathcal{P}.$$

The feedback required for equal-rate transmission is therefore $\bar{N}_a \log N = O\left(\log^{1+\gamma} N\right)$. ∎

F. Proof of Theorem 6

With the on-off power allocation, as $\mu_0 \to \infty$, $\gamma \to 0$, and from (21), we have

$$p = 1 - \frac{\gamma}{1-\gamma}(1-q) \to 1$$

so that

$$H(p) = -(1-p)\log(1-p) - p\log p \to -(1-p)\log(1-p) + p(1-p).$$

Combining this with (19) and (22) gives

$$
\begin{aligned}
\frac{B_N^{(\text{corr})}}{B_N^{(\text{iid})}} &= -\frac{H(q)}{\log \gamma} + \frac{\gamma(1-q)\log(\frac{\gamma}{1-\gamma}(1-q)) + p\gamma(1-q)}{\gamma \log \gamma} \\
&= \frac{q\log q + (1-q)\log(1-q)}{\log \gamma} + 1 - q + (1-q)\frac{\log(1-q)}{\log \gamma} + \frac{(1-q)(-\log(1-\gamma)+p)}{\log \gamma} \\
&\to 1 - q + 2(1-q)\frac{\log(1-q)}{\log \gamma} + \frac{(1-q)(p+\gamma) + q\log q}{\log \gamma}
\end{aligned}
$$

If $\theta(\mu_0) = \frac{\gamma}{1-q} < \infty$, then $\frac{\log(1-q)}{\log \gamma} < \infty$ and $\frac{B_N^{(\text{corr})}}{B_N^{(\text{iid})}} \to 0$, since $q \to 1$.

33

If $\theta(\mu_0) \to 0$, then $\frac{\log(1-q)}{\log\gamma} \to 0$, and

$$
\begin{aligned}
\frac{B_N^{(\text{corr})}}{(1-q)B_N^{(\text{iid})}} &= 1 + 2\frac{\log(1-q)}{\log\gamma} + \frac{\frac{(1-q)(p+\gamma)+q\log q}{\log\gamma}}{1-q} \\
&= 1 + 2\frac{\log(1-q)}{\log\gamma} + \frac{p+\gamma+q}{\log\gamma} \to 1
\end{aligned}
$$

which implies

$$
\frac{B_N^{(\text{corr})}(\mu_0)}{B_N^{(\text{iid})}(\mu_0)} \asymp 1 - q \tag{64}
$$

where q, defined in (20), is a function of μ_0. Therefore $B_N^{(\text{corr})} \asymp (1-q)\gamma\log\frac{1}{\gamma}$.

With n-level rate control

$$
\begin{aligned}
\frac{B_N^{(\text{fp},n)}}{B_N^{(\text{iid})}} &\asymp \sum_{i=0}^n \frac{\pi_i \sum_{j=0}^n p_{ij}\log p_{ij}}{\gamma\log\gamma} \\
&= \frac{\sum_{j=0}^n \pi_0 p_{0j}\log p_{0j} + \sum_{i=1}^n \pi_i \sum_{j=0}^n p_{ij}\log p_{ij}}{\gamma\log\gamma}
\end{aligned} \tag{65}
$$

Applying the log sum inequality [27, Thm 2.7.1] gives

$$
\sum_{j=1}^n p_{ij}\log p_{ij} \geq \sum_{j=1}^n p_{ij}\log\frac{\sum_{j=1}^n p_{ij}}{n-1} = (1-p_{i0})\log\frac{1-p_{i0}}{n-1}, \tag{66}
$$

and since $\gamma\log\gamma < 0$,

$$
\frac{\sum_{j=0}^n \pi_0 p_{0j}\log p_{0j}}{\gamma\log\gamma} \leq \frac{\pi_0 p_{00}\log p_{00} + \pi_0(1-p_{00})\log[(1-p_{00})/(n-1)]}{\gamma\log\gamma}. \tag{67}
$$

As for the binary case, we have $p_{00} = \frac{1-2\gamma+\gamma q}{1-\gamma} = p, \pi_0 = 1-\gamma$, and

$$
\frac{\sum_{j=0}^n \pi_0 p_{0j}\log p_{0j}}{\gamma\log\gamma} \leq \frac{(1-q)(p+\log(1-p)-\log(n-1))}{\log\gamma}. \tag{68}
$$

Also,

$$
\begin{aligned}
\sum_{i\neq 0}\pi_i\sum_j p_{ij}\log p_{ij} &= \sum_{i\neq 0}\pi_i\sum_{j\neq 0} p_{ij}\log p_{ij} + \sum_{i\neq 0}\pi_i p_{i0}\log p_{i0} \\
&\geq \sum_{i\neq 0}\pi_i(1-p_{i0})\log\frac{1-p_{i0}}{n-1} + \sum_{i\neq 0}\pi_i p_{i0}\log p_{i0} \\
&= -\sum_{i\neq 0}\pi_i H(p_{i0}) - \sum_{i\neq 0}\pi_i(1-p_{i0})\log(n-1) \\
&\geq -(1-\pi_0)H(\sum_{i\neq 0}\frac{\pi_i}{1-\pi_0}p_{i0}) - ((1-\pi_0)-(\pi_0-\pi_0 p_{00}))\log(n-1) \\
&= -\gamma H(\sum_{i\neq 0}\frac{\pi_i}{\gamma}p_{i0}) - \gamma q\log(n-1) \\
&= -\gamma H(1-q)\log\gamma - \gamma q\log(n-1).
\end{aligned} \tag{69}
$$

Similar to the derivation of (64), combining (65)-(69) assuming $\theta(\mu_0) \to 0$, we have

$$
\begin{aligned}
\frac{B_N^{(\text{fp},n)}}{B_N^{(\text{iid})}} &\leq \frac{(1-q)(p+\log(1-p))}{\log\gamma} - \frac{H(1-q)}{\log\gamma} - \frac{\log(n-1)}{\log\gamma} \\
&= 1 - q + 2(1-q)\frac{\log(1-q)}{\log\gamma} + (1-q)\frac{p+q+\gamma}{\log\gamma} + \frac{\log(n-1)}{\log\gamma}.
\end{aligned}
$$

34

This implies $\frac{B_N^{(\mathrm{fp},n)}}{(1-q)B_N^{(\mathrm{iid})}} \leq 1$, and since $\frac{B_N^{(\mathrm{fp},n)}}{(1-q)B_N^{(\mathrm{iid})}} \geq \frac{B_N^{(\mathrm{corr})}}{(1-q)B_N^{(\mathrm{iid})}} \to 1$, we have $\frac{B_N^{(\mathrm{fp},n)}}{B_N^{(\mathrm{iid})}} \asymp 1 - q$, or $B_N^{(\mathrm{fp},n)} \asymp$ $(1-q)\gamma \log \frac{1}{\gamma}$. $\blacksquare$

G. Proof of Theorem 7

First, we characterize the asymptotic capacity with the water-filling power allocation. We rewrite (36) as

$$
\begin{aligned}
\frac{\mathcal{P}_k}{N} &= \int_0^\infty \cdots \int_0^\infty \left(\lambda_k - \frac{1}{x_k}\right)^+ \left(\prod_{j \neq k} \mathbf{1}_{x_j < \frac{\lambda_k x_k}{\lambda_j}}\right) d\left(\prod_{j=1}^K F_j(x_j)\right) \\
&= \int_0^\infty \left(\lambda_k - \frac{1}{x_k}\right)^+ \left(\prod_{j \neq k} F_j(\frac{\lambda_k x_k}{\lambda_j})\right) dF_k(x_k) \qquad (70) \\
&= \lambda_k \int_1^\infty \left(1 - \frac{1}{y}\right) \left(\prod_{j \neq k} F_j(\frac{y}{\lambda_j})\right) dF_k(\frac{y}{\lambda_k})
\end{aligned}
$$

As $N \to \infty$, $\lambda_j \to 0$, so that $F_j(\frac{1}{\lambda_j}) \to 1$ with finite users K. Therefore,

$$
\prod_{j \neq k} F_j\left(\frac{1}{\lambda_j}\right) \leq \frac{\frac{\mathcal{P}_k}{N}}{\int_0^\infty \left(\lambda_k - \frac{1}{x_k}\right)^+ dF_k(x_k)} \leq 1 \qquad (71)
$$

so that

$$
\frac{\mathcal{P}_k}{N \int_0^\infty \left(\lambda_k - \frac{1}{x_k}\right)^+ dF_k(x_k)} \to 1. \qquad (72)
$$

Denote the sum capacity with water-filling as $C_{\mathrm{up}}^{(\mathrm{wf},K)} = \sum_{k=1}^K C_{\mathrm{up},k}^{(\mathrm{wf},K)}$, where $C_{\mathrm{up},k}^{(\mathrm{wf},K)}$ is the capacity for user k. We have

$$
\begin{aligned}
C_{\mathrm{up},k}^{(\mathrm{wf},K)} &= \sum_{i=1}^N E\left[(\log(\lambda_k \mu_{k,i}))^+ \mathbf{1}_{\lambda_k \mu_{k,i} = \max_j \lambda_{j,i}\mu_{j,i}}\right] \\
&= N\lambda_k \int_1^\infty \log y \left(\prod_{j \neq k} F_j(\frac{y}{\lambda_j})\right) dF_k(\frac{y}{\lambda_k}).
\end{aligned}
$$

Following the same argument for (72) gives

$$
\frac{C_{\mathrm{up},k}^{(\mathrm{wf},K)}}{N \int_0^\infty (\log(\lambda_k x_k))^+ dF_k(x_k)} \to 1. \qquad (73)
$$

Given (72) and (73), we can simply apply the results for a single link to the multi-user uplink channel, i.e.,

$$
C_{\mathrm{up}}^{(\mathrm{wf},K)} \asymp \sum_{k=1}^K \mathcal{P}_k \mu_{k,0}^*
$$

where $\mu_{k,0}^*$ is the optimal threshold for the on-off power allocation for user k, assuming no other users are present. If the users have identical channel gain distributions and power constraints, $F_k(x) = F_\mu(x)$ and $\mathcal{P}_k = \mathcal{P}$, $1 \leq k \leq K$, then

$$
C_{\mathrm{up}}^{(\mathrm{wf},K)} \asymp K C^{(\mathrm{wf})}. \qquad (74)
$$

With the on-off power allocation, we have

$$
\begin{aligned}
\frac{\mathcal{P}_k}{N} &= E\left[\bar{P}_k \mathbf{1}_{\mu_{k,i} \geq \mu_{k,0}} \mathbf{1}_{\bar{P}_k \mu_{k,i} = \max_j \bar{P}_j \mu_{j,i}}\right] \\
&= \int_0^\infty \cdots \int_0^\infty \bar{P}_k \mathbf{1}_{x_k \geq \mu_{k,0}} \left(\prod_{j \neq k} \mathbf{1}_{x_j < \frac{\bar{P}_k x_k}{\bar{P}_j}}\right) d\left(\prod_{j=1}^K F_j(x_j)\right) \\
&= \bar{P}_k \int_{\mu_{k,0}}^\infty \left(\prod_{j \neq k} F_j(\frac{\bar{P}_k x_k}{\bar{P}_j})\right) dF_k(x_k)
\end{aligned}
\tag{75}
$$

It is difficult to solve for the power analytically, so here we only consider the case where the users have the same channel distribution and power constraint, i.e., $F_k(x) = F_\mu(x)$ and $\mathcal{P}_k = \mathcal{P}$, $1 \leq k \leq K$. From symmetry we must have $\bar{P}_k = \bar{P}_j = \bar{P}$ and $\mu_{k,0} = \mu_{j,0} = \mu_0$.

As $N \to \infty$, $\mu_0 \to \infty$, and since $F_\mu(\frac{\bar{P}_k \mu_{k,0}}{\bar{P}_j}) = F_\mu(\mu_0)$, we have $\prod_{j \neq k} F_\mu(\frac{\bar{P}_k \mu_{k,0}}{\bar{P}_j}) = F_\mu^{K-1}(\mu_0) \to 1$, and from (75) we have

$$
\frac{\bar{P}}{\frac{\mathcal{P}}{N F_\mu(\mu_0)}} \to 1.
\tag{76}
$$

The sum capacity with the on-off power allocation is then given by

$$
\begin{aligned}
C_{\text{up}}^{(\text{on-off},K)} &= \sum_{k=1}^K \sum_{i=1}^N E\left[\log(1 + \bar{P}\mu_{k,i}) \mathbf{1}_{\mu_{k,i} > \mu_{k,0}} \mathbf{1}_{\bar{P}_k \mu_{k,i} = \max_j \bar{P}_j \mu_{j,i}}\right] \\
&= KN \int_{\mu_0}^\infty \log(1 + \bar{P}x) F_\mu^{K-1}(x) dF_\mu(x).
\end{aligned}
\tag{77}
$$

Hence

$$
\frac{C_{\text{up}}^{(\text{on-off},K)}}{KN \int_{\mu_0}^\infty \log(1 + \bar{P}x) dF_\mu(x)} \to 1,
\tag{78}
$$

and combining with (76) and (77) gives

$$
C_{\text{up}}^{(\text{on-off},K)} \asymp K C^{(\text{on-off})} \asymp K \mathcal{P} \mu_0^*.
$$

If all active sub-channels are assigned the same rate, then the achievable rate is

$$
\begin{aligned}
R_{\text{up}}^{(\text{fp},K)} &= \sum_{k=1}^K \sum_{i=1}^N \log(1 + \bar{P}\mu_0) E\left[\mathbf{1}_{\mu_{k,i} > \mu_{k,0}} \mathbf{1}_{\bar{P}_k \mu_{k,i} = \max_j \bar{P}_j \mu_{j,i}}\right] \\
&= KN \log(1 + \bar{P}\mu_0) \int_{\mu_0}^\infty F_\mu^{K-1}(x) dF_\mu(x).
\end{aligned}
$$

The loss in the data rate due to rate quantization is therefore bounded as

$$
K F_\mu^{K-1}(\mu_0) \left(C_N^{(\text{on-off})} - R_1^{(\text{fp})}\right) \leq C_{\text{up}}^{(\text{on-off},K)} - R_{\text{up}}^{(\text{fp},K)} \leq K \left(C_N^{(\text{on-off})} - R_1^{(\text{fp})}\right),
$$

which implies that

$$
C_{\text{up}}^{(\text{on-off},K)} - R_{\text{up}}^{(\text{fp},K)} = K \left(C_N^{(\text{on-off})} - R_1^{(\text{fp})}\right) \leq K \mathcal{P} \max_x \frac{\bar{F}_\mu(x)}{f_\mu(x)}.
$$

∎

H. Proof of Theorem 8

If all users have the same channel gain distribution $F_\mu(x)$, then $G(x) = F_\mu^K(x)$. Substituting in (41), and noting that $F_\mu(\frac{1}{\lambda}) \to 1$ gives

$$\frac{\mathcal{P}}{N} = \int_0^\infty \left(\lambda - \frac{1}{x}\right)^+ dF_\mu(x) \tag{79}$$

and

$$C_{\text{down}}^{(\text{wf},K)} \asymp KN \int_0^\infty (\log(\lambda x))^+ dF_\mu(x) \asymp K\mathcal{P}\mu_0^*, \tag{80}$$

where μ_0^* is the solution to (8), so that the capacity scales linearly with number of users K.

With the on-off power allocation, from (44) we have that

$$C_{\text{down}}^{(\text{on-off},K)} = \frac{1 - F_\mu^K(\mu_0)}{1 - F_\mu(\mu_0)} \cdot C_N^{(\text{on-off})} \tag{81}$$

and the power per active sub-channel is

$$\bar{P} = \frac{K\mathcal{P}}{N(1 - G(\mu_0))} \to \frac{\mathcal{P}}{N(1 - F_\mu(\mu_0))}.$$

As $\mu_0 \to \infty$ and $F_\mu(\mu_0) \to 1$, we have

$$\frac{C_{\text{down}}^{(\text{on-off},K)}}{C_N^{(\text{on-off})}} = \frac{1 - F_\mu^K(\mu_0)}{1 - F_\mu(\mu_0)} \to K$$

so that

$$C_{\text{down}}^{(\text{wf},K)} \asymp C_{\text{down}}^{(\text{on-off},K)} \asymp K\mathcal{P}\mu_0^*.$$

With one-level rate control the achievable rate is given by

$$\begin{aligned}
R_{\text{down}}^{(\text{fp},K)} &= N\log\left(1 + \bar{P}\mu_0\right)\Pr\{\mu_{k,i} \ge \mu_0 \text{ for some } k\} \\
&= N[1 - F_\mu^K(\mu_0)]\log\left(1 + \bar{P}\mu_0\right) \\
&= \frac{1 - F_\mu^K(\mu_0)}{1 - F_\mu(\mu_0)} R_1^{(\text{fp})}.
\end{aligned}$$

Therefore the loss incurred is

$$C_{\text{down}}^{(\text{on-off},K)} - R_{\text{down}}^{(\text{fp},K)} = \frac{1 - F_\mu^K(\mu_0)}{1 - F_\mu(\mu_0)}(C_N^{(\text{on-off})} - R_1^{(\text{fp})}) = O(1).$$

For Rayleigh fading this loss is $C_{\text{down}}^{(\text{on-off},K)} - R_{\text{down}}^{(\text{fp},K)} = K\mathcal{P}$. ∎

I. Proof of Theorem 9

From (81), the on-off capacity per sub-channel is given by

$$C_{\text{down}}^{(\text{on-off},\infty)} = \lim_{K\to\infty} \frac{1 - F_\mu^K(\mu_0)}{1 - F_\mu(\mu_0)} \int_{\mu_0}^\infty \log(1 + \bar{P}x)f_\mu(x)dx$$

The total transmit power satisfies $\mathcal{P}_{\text{tot}} \leq K\mathcal{P}$, and

$$\beta\mathcal{P} = \frac{\mathcal{P}_{\text{tot}}}{N} = \bar{P}[1 - G(\mu_0)]. \tag{82}$$

The activation threshold μ_0 should be no less than the threshold with a single user, so that $\mu_0 \to \infty$ as $N \to \infty$. Let $G(x) = F_\mu^K(x)$. It is straightforward to show that

$$C_{\text{down}}^{(\text{on-off},K)} = [1 - G(\mu_0)]\log(1 + \bar{P}\mu_0) + O\left([1 - G(\mu_0)]\frac{\bar{P}}{1 + \bar{P}\mu_0}(E_\mu[\mu|\mu > \mu_0] - \mu_0)\right), \tag{83}$$

and letting $K \to \infty$, we define

$$\underline{C} = C_{\text{down}}^{(\text{on-off},\infty)} = [1 - G(\mu_0)]\log(1 + \bar{P}\mu_0). \tag{84}$$

We therefore have

$$\begin{aligned}
\underline{C} &= [1 - G(\mu_0)]\log\left(1 + \frac{\beta\mathcal{P}}{1 - G(\mu_0)}\mu_0\right) \\
&= [1 - G(\mu_0)]\log[1 - G(\mu_0) + \beta\mathcal{P}\mu_0) - (1 - G(\mu_0))\log(1 - G(\mu_0)] \\
&= [1 - G(\mu_0)]\log(\beta\mathcal{P}\mu_0) + \underbrace{[1 - G(\mu_0)]\log\left(1 + \frac{1 - G(\mu_0)}{\beta\mathcal{P}\mu_0}\right)}_{o(1)} - \underbrace{[1 - G(\mu_0)]\log[1 - G(\mu_0)]}_{\text{finite}}
\end{aligned}$$
$$\tag{85}$$

Maximizing $\underline{C}$ is therefore equivalent to maximizing $[1 - G(\mu_0)]\log(\beta\mathcal{P}\mu_0)$.

We first show that

$$\lim_{K \to \infty} \frac{\underline{C}}{\log\log K} \leq 1 \tag{86}$$

by following the approach in [33]. Namely, we write the threshold as

$$\mu_K = \sigma^2 \log K + x_K, \tag{87}$$

and observe that

$$\lim_{K \to \infty} G(\mu_0) = \begin{cases} 0, & \text{if } x_K \to -\infty, \\ \exp(-e^{-x_0}), & \text{if } x_K \to x_0, \\ 1, & \text{if } x_K \to \infty. \end{cases}$$

To maximize $\underline{C}$, we therefore want $x_K \to -\infty$ with $x_K = o(\log K)$, so that μ_K grows as $O(\log K)$. This choice of x_K achieves equality in (86), whereas if these conditions are not satisfied, then the inequality is strict.

We now show that the sequence x_K in (87) can be chosen so that

$$\left|[1 - G(\mu_0)]\log(\beta\mathcal{P}\mu_0) - \log(\beta\mathcal{P}\sigma^2\log K)\right| \to 0. \tag{88}$$

From (85) and (86) this choice of x_K maximizes $\underline{C}$. We first write

$$[1 - G(\mu_0)]\log(\beta\mathcal{P}\mu_0) - \log(\beta\mathcal{P}\sigma^2\log K) = \log\left(\frac{\mu_0}{\sigma^2\log K}\right) - G(\mu_0)\log(\beta\mathcal{P}) - G(\mu_0)\log\mu_0, \tag{89}$$

and note that $\mu_0/(\sigma^2 \log K) \to 1$ to achieve the bound in (86), and $G(\mu_0)\log(\beta\mathcal{P}) \to 0$. Hence we must select x_K so that $G(\mu_0)\log(\beta\mu_0) = (1 - e^{-\mu_K/\sigma^2})\log(\beta\mu_0) \to 0$. It can be shown that this is accomplished by taking $x_K = -\sigma^2 \log\log K$.

With one-level rate control,

$$R_{\text{down}}^{(\text{fp},\infty)} = \lim_{K \to \infty} [1 - G(\mu_0)]\log\left(1 + \frac{\beta\mathcal{P}}{1 - G(\mu_0)}\mu_0\right) = \underline{C}.$$

Therefore $R_{\text{down}}^{(\text{fp},\infty)} \to C_{\text{down}}^{(\text{fp},\infty)}$, i.e., the asymptotic on-off capacity can be achieved with one-level rate control.

If all users have linear receivers, then the optimal power allocation is to water-fill over the maximum sub-channel gains across users. That is, the power constraint (41) becomes

$$\beta\mathcal{P} = \int_0^\infty \left(\lambda - \frac{1}{x}\right)^+ dG(x) = \lambda\left(1 - G\left(\frac{1}{\lambda}\right)\right) - \int_{1/\lambda}^\infty \frac{dG(x)}{x}.$$

For any $\epsilon > 0$, we have

$$\int_{1/\lambda}^\infty \frac{dG(x)}{x} = \int_{1/\lambda}^{2/\epsilon} \frac{dG(x)}{x} + \int_{2/\epsilon}^\infty \frac{dG(x)}{x}$$

$$\leq \lambda G\left(\frac{2}{\epsilon}\right) + \frac{\epsilon}{2} < \epsilon$$

when $K \geq \frac{\log\left(\frac{\epsilon}{2\lambda}\right)}{\log F_\mu(2/\epsilon)}$, hence $\int_{1/\lambda}^\infty \frac{dG(x)}{x} \to 0$, and $\beta\mathcal{P} \to \lambda\left(1 - G\left(\frac{1}{\lambda}\right)\right)$. Furthermore, since $\lambda \geq \beta\mathcal{P}$, $G\left(\frac{1}{\lambda}\right) \to 0$ and $\lambda \to \beta\mathcal{P}$.

Let $X_i = \max\{\mu_{j,i}\}$, i.e., the maximum channel gain over users for sub-channel i. For large enough K, the capacity with water-filling power allocation is upper bounded as

$$C_{\text{down}}^{(\text{wf},K)} \leq \int_{\frac{1}{\beta\mathcal{P}}}^\infty \log(\beta\mathcal{P}x)dG(x) \leq \log\left(\beta\mathcal{P}E[X_i]\right) = \bar{C}^{(\text{wf})}.$$

It is easy to show that $\bar{C}^{(\text{wf})} - \log(1 + \beta\mathcal{P}\sigma^2 \log K) \to 0$ as $K \to \infty$. Because $C_{\text{down}}^{(\text{wf},\infty)} \geq C_{\text{down}}^{(\text{on-off},\infty)}$, we have

$$\lim_{K \to \infty}\left(C_{\text{down}}^{(\text{wf},K)} - \log(1 + \beta\mathcal{P}\sigma^2 \log K)\right) = \lim_{K \to \infty}\left(C_{\text{down}}^{(\text{on-off},K)} - \log(1 + \beta\mathcal{P}\sigma^2 \log K)\right)$$

$$= \lim_{K \to \infty}\left(R_{\text{down}}^{(\text{fp},K)} - \log(1 + \beta\mathcal{P}\sigma^2 \log K)\right) = 0.$$

∎

REFERENCES

[1] R. G. Gallager, *Information Theory and Reliable Communication*, New York: Wiley, 1968.

[2] G. J. Foschini and M. J. Gans, "Capacity When Using Diversity at Transmit and Receive Sites and the Rayleigh-Faded Matrix Channel in Unknown at the Transmitter", *WINLAB Workshop on Wireless Information Network*, New Brunswick, NJ, March, 1996.

[3] E. Telatar, "Capacity of Multi-Antenna Gaussian Channels", *European Transactions on Telecommunications*, 10(6):585-595, Nov/Dec 1999

[4] C. Chuah, D. Tse, J. Kahn and R. Valenzuela, "Capacity Scaling in MIMO Wireless Systems under Correlated Fading", *IEEE Transactions on Information Theory*, 48(3):637-650, March, 2002.

[5] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications", in *Conference Record of the International Conference on Communications (ICC)*, pp. 331-335, Seattle, WA, June, 1995.

[6] X. Qin and R. Berry, "Distributed Approaches for Exploiting Multiuser Diversity in Wireless Networks", *IEEE Transactions on Information Theory*, 52(2):392-413, Feb., 2006.

[7] P. Viswanath, D. Tse and R. Laroia, "Opportunistic Beamforming Using Dumb Antennas" *IEEE Transactions on Information Theory*, 48(6):1277-1294, June 2002.

[8] C. Y. Wong, R. S. Cheng, K. B. Letaief and R. D. Murch, "Multiuser OFDM with Adaptive Subcarrier, Bit, and Power allocation", *IEEE Journal on Selected Areas in Communications*, 17(10):1747-1758, Oct., 1999.

[9] U. Schuster, and H. Bolcskei, "Ultrawideband Channel Modeling on the Basis of Information-Theoretic Criteria", *IEEE Trans. on Wireless Communications*, 6(7):2464-2475, July, 2007.

[10] P. Xia, S. Zhou and G. B. Giannakis, "Adaptive MIMO-OFDM based on Partial Channel State Information", *IEEE Transactions on Signal Processing*. 52(1):202-213, Jan. 2004.

[11] Y. Yao and G. B. Giannakis, "Rate-Maximizing Power Allocation in OFDM Based on Partial Channel Knowledge", *IEEE Transactions on Wireless Communications*. 4(3):1073-1083, May 2005.

[12] A. Narula, M. J. Lopez, M. D. Trott and G. W. Wornell, "Efficient Use of Side Information in Multiple-Antenna Data Transmission over Fading Channels", *IEEE Transactions on Communications*, 16(8):1423-1436, Oct., 1998.

[13] D. J. Love and R. W. Heath, Jr. Grassmannian beamforming for multiple-input multiple-output wireless systems. *IEEE Transactions on Information Theory*, 49(10):2735–2745, October 2003.

[14] K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On Beamforming With Finite Rate Feedback In Multiple Antenna Systems", *IEEE Transactions on Information Theory*, 49(10):2562-2579, Oct., 2003.

[15] D. J. Love, R. W. Heath, Jr., W. Santipach, and M. L. Honig. What is the value of limited feedback for MIMO channels? *IEEE Communications Magazine*, 42(10):54–59, October 2004.

[16] W. Santipach and M. L. Honig. Asymptotic performance of MIMO wireless channels with limited feedback. In *Proc. IEEE Military Commun. Conf. (MILCOM)*, pages 141–146, Boston, MA, October 2003.

[17] V. K. N. Lau, Y. J. Liu, and T.-A. Chen. Capacity of memoryless channels and block-fading channels with designable cardinality-contrained channel state feedback. *IEEE Transactions on Information Theory*, 50(9):2038–2049, September 2004.

[18] S. Zhou, Z. Wang, and G. B. Giannakis. Quantifying the power-loss when transmit-beamforming relies on finite rate feedback. *IEEE Transactions on Wireless Communications*, 4(4):1948–1957, July 2005.

[19] P. Xia, S. Zhou, and G. B. Giannakis. Multiantenna adaptive modulation with beamforming based on bandwidth-constrained feedback. *IEEE Transactions on Communications*, 53(3):526–536, March 2005.

[20] J. C. Roh and B. D. Rao. Transmit beamforming in multiple antenna systems with finite rate feedback: A VQ-based approach. *IEEE Transactions on Information Theory*, 52(3):1101-1112, March 2006.

[21] S. Sanayei and A. Nosratinia. Capacity of MIMO channels with antenna selection. *IEEE Transactions on Information Theory*, vol. 53, no. 11, Nov. 2007, pp. 4356-4362.

[22] W. Santipach and M. L. Honig. Signature optimization for CDMA with limited feedback. *IEEE Transactions on Information Theory*, 51(10):3475–3492, October 2005.

[23] J Jang and K.-B Lee, "Transmit power adaptation for multiuser OFDM systems", *IEEE Journal on Selected Areas in Communications*, 21(2): 171-178, Feb. 2003.

[24] L. Hoo, B. Halder, J. Tellado, and J. Cioffi. Multiuser transmit optimization for multicarrier broadcast channels: Asymptotic FDMA capacity region and algorithms. *IEEE Transactions on Communications*, 52(6):933–946, June 2004.

[25] W. Yu and J. Cioffi. Constant-Power Waterfilling: Performance Bound and Low-Complexity Implementation. *IEEE Transactions on Communications*, 54(1):23–28, Jan. 2006.

[26] S. Sanayei and A. Nosratinia. Opportunistic downlink transmission with limited feedback. *IEEE Transactions on Information Theory*, vol. 53, no. 11, Nov. 2007, pp. 4363-4372.

[27] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 1991, John Wiley & Sons, Inc.

[28] E. Castillo, *Extreme Value Theory in Engineering*, 1998, Academic Press, New York.

[29] J. Galambos, *Asymptotic Theory of Extreme Order Statistics*, 1987, Robert E. Krieger Publishing Company, Florida.

[30] T. S. Rappaport, *Wireless Communications*, 1999, Prentice-Hall, PTR.

[31] H. Bi and M. L. Honig, "Forward Link Capacity Scaling with Linear Receivers and Multiple Transmit Antennas", IEEE Globecom Conf., San Francisco, Ca., December 2003.

[32] R. S. Cheng and S. Verdu, "Gaussian Multiaccess Channels with ISI: Capacity Region and Multiuser Water-Filling", *IEEE Transactions on Information Theory*, 39(3): 773-785, May, 1993.

[33] J. Chen, R. Berry, and M. L. Honig, "Large System Performance of Downlink OFDMA with Limited Feedback", *Proc. International Symposium on Information Theory*, Seattle, Washington, July 2006.

[34] M. Agarwal and M. L. Honig. Wideband fading channel capacity with training and partial feedback. In *Proc. Allerton Conference on Communication, Control and Computing*, Monticello, IL, September 2005.

[35] W. Santipach and M. L. Honig, "Capacity of Beamforming with Limited Training and Feedback", *Proc. International Symposium on Information Theory*, Seattle, Washington, July 2006.

[36] T. Cover and S. Pombra, "Gaussian Feedback Capacity", *IEEE Transactions on Information Theory*, Vol. 35, pp. 37-43, Jan. 1989.

[37] Y-H Kim, "Feedback Capacity of the First-Order Moving Average Gaussian Channel", *IEEE Transactions on Information Theory*, to appear.

[38] M. Sharif and B. Hassibi. On the capacity of MIMO broadcast channel with partial side information. *IEEE Transactions on Information Theory*, 51(2):506–522, February 2005.

[39] S.J. Howard, and K. Pahlavan, "Autoregressive Modeling of Wide-Band Indoor Radio Progagation", *IEEE Trans. on Comm.*, Vol. 40, No. 9, pp 1540-1552, Sept., 1992.

[40] S. Sanayei and A. Nosratinia. Opportunistic beamforming with limited feedback. *IEEE Transactions on Wireless Communications*, vol. 6, no. 8, August 2007 pp. 2765 - 2771.