

# How Good is Phase-Shift Keying for Peak-Limited Fading Channels in the Low-SNR Regime?

Wenyi Zhang and J. Nicholas Laneman

Dept. of Electrical Engineering

University of Notre Dame

wzhang1, jnl@nd.edu

## Abstract

This paper investigates the achievable information rate of phase-shift keying (PSK) over frequency non-selective Rayleigh fading channels without channel state information (CSI). The fading process exhibits general temporal correlation characterized by its spectral density function. We consider both discrete-time and continuous-time channels, and find their asymptotics at low signal-to-noise ratio (SNR). Compared to known capacity upper bounds under peak constraints, these asymptotics usually lead to negligible rate loss in the low-SNR regime for slowly time-varying fading channels. We further specialize to case studies of Gauss-Markov and Clarke's fading models.

## I. INTRODUCTION

For Rayleigh fading channels without channel state information (CSI) at low signal-to-noise ratio (SNR), the capacity-achieving input gradually tends to bursts of “on” intervals sporadically inserted into the “off” background, even under vanishing peak power constraints [1]. This highly unbalanced input usually imposes implementation challenges. For example, it is difficult to maintain carrier frequency and symbol timing during the long “off” periods. Furthermore, the unbalanced input is incompatible with linear codes, unless appropriate symbol mapping (*e.g.*,  $M$ -ary orthogonal modulation with appropriately chosen constellation size  $M$ ) is employed to match the input distribution.

This paper investigates the achievable information rate of phase-shift keying (PSK). PSK is appealing because it has constant envelope and is amenable to linear codes without additional symbol mappings. Focusing on low signal-to-noise ratio (SNR) asymptotics, we utilize a recursive training scheme to convert the original fading channel without CSI into a series of parallel sub-channels, each with estimated CSI but additional noise that remains circular complex white Gaussian. The central results in this paper are as follows. First, for a discrete-time channel whose unit-variance fading process  $\{H_d[k] : -\infty < k < \infty\}$  has a spectral density function  $S_{H_d}(e^{j\Omega})$  for  $-\pi \leq \Omega \leq \pi$ , the achievable rate is  $(1/2) \cdot \left[ (1/2\pi) \cdot \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega - 1 \right] \cdot \rho^2 + o(\rho^2)$  nats per symbol, as the average channel SNR  $\rho \rightarrow 0$ . This achievable rate is at most  $(1/2) \cdot \rho^2 + o(\rho^2)$  away from the channel capacity under peak SNR constraint  $\rho$ . Second, for a continuous-time channel whose unit-variance fading process  $\{H_c(t) : -\infty < t < \infty\}$  has a spectral density function  $S_{H_c}(j\omega)$  for  $-\infty < \omega < \infty$ , the achievable rate as the input symbol duration  $T \rightarrow 0$  is  $\left[ 1 - (1/2\pi P) \cdot \int_{-\infty}^{\infty} \log(1 + P \cdot S_{H_c}(j\omega)) d\omega \right] \cdot P$  nats per unit time, where  $P > 0$  is the envelope power. This achievable rate coincides with the channel capacity under peak envelope  $P$ .

We further apply the above results to specific case studies of Gauss-Markov fading models (both discrete-time and continuous-time) as well as a continuous-time Clarke's fading model. For discrete-time Gauss-Markov fading processes with innovation rate  $\epsilon \ll 1$ , the quadratic behavior of the achievable rate becomes dominant only for  $\rho \ll \epsilon$ . Our results, combined with previous results for the high-SNR asymptotics, suggest that coherent communication can

essentially be realized for  $\epsilon \leq \rho \leq 1/\epsilon$ . For Clarke's model, we find that the achievable rate scales sub-linearly, but super-quadratically, as  $O(\log(1/P) \cdot P^2)$  nats per unit time as  $P \rightarrow 0$ .

The remainder of this paper is organized as follows. Section II describes the channel model and the recursive training scheme. Section III deals with the discrete-time channel model, and Section IV the continuous-time channel model. Finally Section V provides some concluding remarks. Throughout the paper, random variables are in capital letters while their sample values are in small letters. All logarithms are to base  $e$ , and information units are measured in nats. Proofs and further interpretations can be found in the journal version of the paper [2].

## II. CHANNEL MODEL, RECURSIVE TRAINING SCHEME, AND EFFECTIVE SNR

We consider a scalar time-selective, frequency non-selective Rayleigh fading channel, written in baseband-equivalent continuous-time form as

$$X(t) = H_c(t) \cdot S(t) + Z(t), \quad \text{for } -\infty < t < \infty, \quad (1)$$

where  $S(t) \in \mathcal{C}$  and  $X(t) \in \mathcal{C}$  denote the channel input and the corresponding output at time instant  $t$ , respectively. The additive noise  $\{Z(t) : -\infty < t < \infty\}$  is modeled as a zero-mean circular complex Gaussian white noise process with  $\mathcal{E}\{Z(s)Z^\dagger(t)\} = \delta(s-t)$ . The fading process  $\{H_c(t) : -\infty < t < \infty\}$  is modeled as a wide-sense stationary and ergodic zero-mean circular complex Gaussian process with unit variance  $\mathcal{E}\{H_c(t)H_c^\dagger(t)\} = 1$  and with spectral density function  $S_{H_c}(j\omega)$  for  $-\infty < \omega < \infty$ . Additionally, we impose a technical condition that  $\{H_c(t) : -\infty < t < \infty\}$  is mean-square continuous, so that its autocorrelation function  $K_{H_c}(\tau) = \mathcal{E}\{H_c(t+\tau)H_c^\dagger(t)\}$  is continuous for  $\tau \in (-\infty, \infty)$ .

Throughout the paper, we restrict our attention to PSK over the continuous-time channel (1). For technical convenience, we let the channel input  $S(t)$  have constant envelope  $P > 0$  and piecewise constant phase, *i.e.*,

$$S(t) = S[k] = \sqrt{P} \cdot e^{j\Theta[k]}, \quad \text{if } kT \leq t < (k+1)T,$$

for  $-\infty < k < \infty$ .<sup>1</sup> The symbol duration  $T > 0$  is determined by the reciprocal of the channel input bandwidth.<sup>2</sup>

Applying the above channel input to the continuous-time channel (1), and processing the channel output through a matched filter,<sup>3</sup> we obtain a discrete-time channel

$$X[k] = \sqrt{\rho} \cdot H_d[k] \cdot S[k] + Z[k], \quad \text{for } -\infty < k < \infty. \quad (2)$$

For the discrete-time channel (2) we can verify that

- The additive noise  $\{Z[k] : -\infty < k < \infty\}$  is circular complex Gaussian with zero mean and unit variance, *i.e.*,  $Z[k] \sim \mathcal{CN}(0, 1)$ , and is independent, identically distributed (i.i.d.) for different  $k$ .
- The fading process  $\{H_d[k] : -\infty < k < \infty\}$  is wide-sense stationary and ergodic zero-mean circular complex Gaussian, with  $H_d[k]$  being marginally  $\mathcal{CN}(0, 1)$ . We further notice that  $\{H_d[k] : -\infty < k < \infty\}$  is obtained through sampling the output of the matched filter, hence its spectral density function is

$$S_{H_d}(e^{j\Omega}) = \frac{1}{\sqrt{\int_0^T \int_0^T K_{H_c}(s-t) ds dt}} \cdot \sum_{k=-\infty}^{\infty} S_{H_c}\left(j\frac{\Omega - 2k\pi}{T}\right) \cdot \text{sinc}^2(\Omega - 2k\pi)$$

<sup>1</sup>Here we note a slight abuse of notation in this paper, that a symbol, *e.g.*,  $S$ , can be either continuous-time or discrete-time. The two cases are distinguished by different indexing notation, *e.g.*,  $S(t)$  for continuous-time and  $S[k]$  for discrete-time.

<sup>2</sup>For multipath fading channels,  $T$  should also be substantially greater than the delay spread [3], otherwise the frequency non-selective channel model (1) may not be valid. Throughout the paper we assume that this requirement is met.

<sup>3</sup>A matched filter suffers no information loss for white Gaussian channels [4]. For the fading channel (1), it is no longer optimal in general [5]. However, in this paper we still focus on the matched filter, which is common in most practical systems.

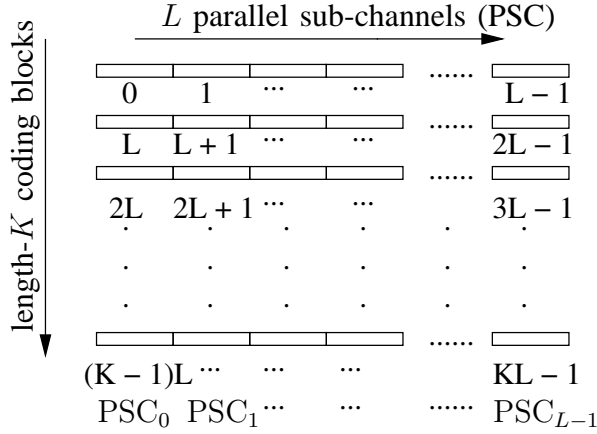


Fig. 1. Illustration of the interleaving scheme. Input symbols are encoded/decoded column-wise, and transmitted/received row-wise.

for  $-\pi \leq \Omega \leq \pi$ .

- The channel input  $\{S[k] : -\infty < k < \infty\}$  is always on the unit circle. In the sequel, we will further restrict it to be complex proper [6], *i.e.*,  $\mathcal{E}\{S^2[k]\} = [\mathcal{E}\{S[k]\}]^2$ . The simplest such input is quadrature phase-shift keying (QPSK); by contrast, binary phase-shift keying (BPSK) is not complex proper.
- The average channel SNR is given by

$$\rho = \frac{P}{T} \cdot \left( \int_0^T \int_0^T K_{H_c}(s-t) ds dt \right) > 0. \quad (3)$$

Throughout the paper, we assume that the realization of the fading process  $\{H_c(t) : -\infty < t < \infty\}$  is not directly available to the transmitter or the receiver, but its statistical characterization in terms of  $S_{H_c}(j\omega)$  is precisely known at both the transmitter and the receiver.

We employ a recursive training scheme to communicate over the discrete-time channel (2). By interleaving the transmitted symbols as illustrated in Figure 1 (cf. [7]), the recursive training scheme effectively converts the original non-coherent channel into a series of parallel sub-channels, each with estimated receive CSI but additional noise that remains i.i.d. circular complex Gaussian. The interleaving scheme decomposes the channel into  $L$  parallel sub-channels (PSC). The  $l$ th ( $l = 0, 1, \dots, L-1$ ) PSC sees all the inputs  $S[k \cdot L + l]$  of (2) for  $k = 0, 1, \dots, K-1$ . These  $L$  PSCs suffer correlated fading, and this correlation is exactly what we seek to exploit using recursive training. Although some residual correlation remains within each PSC among its  $K$  symbols, due to the ergodicity of the channel (2), this correlation vanishes as the interleaving depth  $L \rightarrow \infty$ . In practical systems with finite  $L$ , if necessary, we may utilize an additional interleaver for each PSC to make it essentially memoryless.

We make a slight abuse of notation in the sequel. Since all the PSCs are viewed as memoryless, when describing a PSC we can simply suppress the internal index  $k$  among its  $K$  coding symbols, and only indicate the PSC index  $l$  without loss of generality. For example,  $H_d[l]$  actually corresponds to any  $H_d[k \cdot L + l]$ , for  $k = 0, 1, \dots, K-1$ .

The recursive training scheme performs channel estimation and demodulation/decoding in an alternating manner. To initialize transmission, PSC 0, the first parallel sub-channel, transmits pilots rather than information symbols to the receiver. Based upon the received pilots in PSC 0, the receiver predicts  $H_d[1]$ , the fading coefficient of PSC 1, and proceeds to demodulate and decode the transmitted symbols in PSC 1 coherently. If the rate of PSC 1 does not exceed the corresponding channel mutual information, then information theory ensures that, as the coding block length  $K \rightarrow \infty$ , there always exist codes that have arbitrarily small decoding error probability. Hence the receiver can, at least in principle, form an error-free reconstruction

of the transmitted symbols in PSC 1, which then effectively become “fresh” pilots to facilitate the prediction of  $H_d[2]$  and subsequent coherent demodulation/decoding of PSC 2. Alternating the estimation-demodulation/decoding procedure repeatedly, all the PSCs are reliably decoded one after another.

By induction, let us consider PSC  $l$ , assuming that the inputs  $\{S[i] : i = 0, 1, \dots, l-1\}$  of the previous PSCs have all been successfully reconstructed at the receiver. Since the channel inputs are always on the unit circle, the receiver can compensate for their phases in the channel outputs, and the resulting observations become

$$\underbrace{e^{-j\Theta[i]} \cdot X[i]}_{X'[i]} = \sqrt{\rho} \cdot H_d[i] + \underbrace{e^{-j\Theta[i]} \cdot Z[i]}_{Z'[i]} \quad \text{for } i = 0, 1, \dots, l-1.$$

Since zero-mean circular complex Gaussian distributions are invariant under rotation, the rotated noise  $Z'[i]$  remains i.i.d. zero-mean unit-variance circular complex Gaussian. Then we can utilize standard linear prediction theory (e.g., [8]) to obtain the one-step minimum mean-square error (MMSE) prediction of  $H_d[l]$  defined as

$$\hat{H}_d[l] = \mathcal{E} \{H_d[l] \mid \{X'[i] : i = 0, 1, \dots, l-1\}\}. \quad (4)$$

The estimate  $\hat{H}_d[l]$  and the estimation error  $\tilde{H}_d[l] = H_d[l] - \hat{H}_d[l]$  are jointly circular complex Gaussian distributed as  $\mathcal{CN}(0, 1 - \sigma^2[l])$  and  $\mathcal{CN}(0, \sigma^2[l])$ , respectively, and are uncorrelated and further independent. Here  $\sigma^2[l]$  denotes the mean-square prediction error. The channel equation of PSC  $l$  can then be written as

$$\begin{aligned} X[l] &= \sqrt{\rho} \cdot H_d[l] \cdot S[l] + Z[l] \\ &= \sqrt{\rho} \cdot \hat{H}_d[l] \cdot S[l] + \underbrace{\sqrt{\rho} \cdot \tilde{H}_d[l] \cdot S[l]}_{\tilde{Z}[l]} + Z[l], \end{aligned} \quad (5)$$

where the effective noise  $\tilde{Z}[l]$  is circular complex Gaussian, and is independent of both the channel input  $S[l]$  and the estimated fading coefficient  $\hat{H}_d[l]$ . Thus, the channel (5) becomes a coherent Gaussian channel with fading and receive CSI  $\hat{H}_d[l]$ , with effective SNR

$$\rho[l] = \frac{1 - \sigma^2[l]}{\sigma^2[l] \cdot \rho + 1} \cdot \rho. \quad (6)$$

In the paper we mainly focus on the ultimate performance limit without delay constraints, which is achieved as the interleaving depth  $L \rightarrow \infty$ . Under mild technical conditions, the one-step MMSE prediction error sequence  $\{\sigma^2[l] : l = 0, 1, \dots\}$  converges to the limit [9, Chap. XII, Sec. 4]

$$\sigma_\infty^2 \stackrel{\text{def}}{=} \lim_{l \rightarrow \infty} \sigma^2[l] = \frac{1}{\rho} \cdot \left\{ \exp \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \left( 1 + \rho \cdot S_{H_d}(e^{j\Omega}) \right) d\Omega \right\} - 1 \right\}. \quad (7)$$

Consequently the effective SNR (6) sequence  $\{\rho[l] : l = 0, 1, \dots\}$  converges to

$$\rho_\infty \stackrel{\text{def}}{=} \lim_{l \rightarrow \infty} \rho[l] = \frac{1 - \sigma_\infty^2}{\sigma_\infty^2 \cdot \rho + 1} \cdot \rho. \quad (8)$$

We are mainly interested in evaluating the mutual information of the induced channel (5) at the limiting effective SNR  $\rho_\infty$  as the actual channel SNR  $\rho \rightarrow 0$ . This low-SNR channel analysis is facilitated by the explicit second-order expansion formulas of the channel mutual information at low SNR [10]. Applying [10, Theorem 3] to the induced channel<sup>4</sup> (5) at  $\rho_\infty$ , we have

$$R \stackrel{\text{def}}{=} \lim_{l \rightarrow \infty} I(S[l]; X[l] \mid \hat{H}_d[l]) = \rho_\infty - \rho_\infty^2 + o(\rho^2) \quad \text{as } \rho \rightarrow 0. \quad (9)$$

<sup>4</sup>Note that [10, Theorem 3] is only applicable to complex proper channel inputs, as we have assumed in the channel model.

### III. ASYMPTOTIC CHANNEL MUTUAL INFORMATION AT LOW SNR

As shown in (9), the asymptotic channel mutual information depends on the limiting effective SNR (8), which further relates to the limiting one-step MMSE prediction error (7). The following theorem evaluates the asymptotic behavior of the channel mutual information (9).

*Theorem 3.1:* For the discrete-time channel (2), as  $\rho \rightarrow 0$ , its induced channel (5) achieves the rate

$$R = \frac{1}{2} \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega - 1 \right] \cdot \rho^2 + o(\rho^2), \quad (10)$$

if the integral  $(1/2\pi) \cdot \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega$  exists.

Theorem 3.1 states that for PSK at low SNR, the achievable channel mutual information vanishes quadratically with SNR. This is consistent with [11] [12]. Furthermore, it is of particular interest to compare the asymptotic expansion (10) with several previous results.

#### A. Comparison with a Capacity Upper Bound

For the discrete-time channel (2), PSK with constant SNR  $\rho$  is a particular peak-limited channel input. The capacity per unit energy of channel (2) under a peak SNR constraint  $\rho$  is [1]

$$\dot{C} = 1 - \frac{1}{2\pi\rho} \cdot \int_{-\pi}^{\pi} \log(1 + \rho \cdot S_{H_d}(e^{j\Omega})) d\Omega, \quad (11)$$

achieved by on-off keying (OOK) in which each “on” or “off” symbol corresponds to an infinite number of channel uses, and the probability of choosing “on” symbols vanishes. Such “bursty” channel inputs are in sharp contrast to PSK. From (11), an upper bound to the channel capacity can be derived as [1]

$$C \leq U(\rho) \stackrel{\text{def}}{=} \frac{1}{2} \cdot \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega \cdot \rho^2. \quad (12)$$

Comparing (10) and (12), we notice that the penalty of using PSK instead of the bursty capacity-achieving channel input is at most  $(1/2) \cdot \rho^2 + o(\rho^2)$ . For fast time-varying fading processes, this penalty can be relatively significant. For instance, if the fading process is memoryless, *i.e.*,  $S_{H_d}(e^{j\Omega}) = 1$  for  $-\pi \leq \Omega \leq \pi$ , then  $(1/2\pi) \cdot \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega - 1 = 0$ , implying that no information can be transmitted using PSK over a memoryless fading channel. Fortunately, for slowly time-varying fading processes, the integral  $(1/2\pi) \cdot \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega$  is typically much greater than 1, as we will illustrate in the sequel.

#### B. Comparison with the High-SNR Channel Behavior

From (10) and (12), it can be observed that  $(1/2\pi) \cdot \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega$  is a fundamental quantity associated with a fading process at low SNR. This is in contrast to the high-SNR regime, where a fundamental quantity is [13]

$$\sigma_{\text{pred}}^2 \stackrel{\text{def}}{=} \exp \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log S_{H_d}(e^{j\Omega}) d\Omega \right\}.$$

The quantity  $\sigma_{\text{pred}}^2$  is the one-step MMSE prediction error of  $H_d[0]$  given its entire noiseless past  $\{H_d[-1], H_d[-2], \dots\}$ . When  $\sigma_{\text{pred}}^2 > 0$  the process is said to be regular; and when  $\sigma_{\text{pred}}^2 = 0$  it is said to be deterministic, that is, the entire future  $\{H_d[0], H_d[1], \dots\}$  can be exactly reconstructed (in the mean-square sense) by linearly combining the entire past  $\{H_d[-1], H_d[-2], \dots\}$ . It has been established in [13] [14] that, for regular fading processes,

$$C = \log \log \rho - 1 - \gamma + \log \frac{1}{\sigma_{\text{pred}}^2} + o(1) \quad \text{as } \rho \rightarrow \infty, \quad (13)$$

where  $\gamma = 0.5772\dots$  is Euler's constant, and for deterministic fading processes,

$$\frac{C}{\log \rho} \rightarrow \frac{1}{2\pi} \cdot \mu \left( \left\{ \Omega : S_{H_d}(e^{j\Omega}) = 0 \right\} \right) \quad \text{as } \rho \rightarrow \infty, \quad (14)$$

where  $\mu(\cdot)$  denotes the Lebesgue measure on the interval  $[-\pi, \pi]$ .

It is then an interesting issue to investigate the relation between  $(1/2\pi) \cdot \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega$  and  $\sigma_{\text{pred}}^2$ . However, there is no explicit relation between these two quantities. We can explicitly construct deterministic fading processes that lead to arbitrarily small quadratic coefficient in (10), as well as almost memoryless fading processes that lead to divergent quadratic coefficient in (10).

### C. Case Study: Discrete-Time Gauss-Markov Fading Processes

In this subsection, we apply Theorem 3.1 to analyze a specific class of discrete-time fading processes, namely, the discrete-time Gauss-Markov fading processes. The fading process in the channel model can be described by a first-order auto-regressive (AR) evolution equation of the form

$$H_d[k+1] = \sqrt{1-\epsilon} \cdot H_d[k] + \sqrt{\epsilon} \cdot V[k+1], \quad (15)$$

where the innovation sequence  $\{V[k] : -\infty < k < \infty\}$  consists of i.i.d.  $\mathcal{CN}(0, 1)$  random variables, and  $V[k+1]$  is independent of  $\{H_d[i] : -\infty < i \leq k\}$ . The innovation rate  $\epsilon$  satisfies  $0 < \epsilon \leq 1$ .

The spectral density function  $S_{H_d}(e^{j\Omega})$  for such a process is

$$S_{H_d}(e^{j\Omega}) = \frac{\epsilon}{(2-\epsilon) - 2\sqrt{1-\epsilon} \cdot \cos \Omega}, \quad -\pi \leq \Omega \leq \pi. \quad (16)$$

Hence

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} S_{H_d}^2(e^{j\Omega}) d\Omega = \frac{\epsilon^2}{2\pi} \int_{-\pi}^{\pi} \frac{1}{\left( (2-\epsilon) - 2\sqrt{1-\epsilon} \cdot \cos \Omega \right)^2} d\Omega = \frac{2}{\epsilon} - 1.$$

Applying Theorem 3.1, we find that for the discrete-time Gauss-Markov fading model,

$$R = \left( \frac{1}{\epsilon} - 1 \right) \cdot \rho^2 + o(\rho^2) \quad \text{as } \rho \rightarrow 0. \quad (17)$$

For practical systems in which the fading processes are underspread [3], the innovation rate  $\epsilon$  typically ranges from  $1.8 \times 10^{-2}$  to  $3 \times 10^{-7}$  [15]. So the  $(1/2) \cdot \rho^2 + o(\rho^2)$  rate penalty of PSK with respect to optimal, peak-limited signaling may be essentially negligible at low SNR.

Due to the simplicity of the discrete-time Gauss-Markov fading model, we are able to carry out a non-asymptotic analysis to gain more insight. Applying (16) to (7), the steady-state limiting channel prediction error is

$$\sigma_{\infty}^2 = \frac{(\rho-1) \cdot \epsilon + \sqrt{(\rho-1)^2 \cdot \epsilon^2 + 4\rho\epsilon}}{2\rho}. \quad (18)$$

Further applying (18) to (8), we can identify the following three qualitatively distinct operating regimes of the induced channel (5) for small  $\epsilon \ll 1$ :

- The quadratic regime: For  $\rho \ll \epsilon$ ,  $\sigma_{\infty}^2 \approx 1 - \rho/\epsilon$ ,  $\rho_{\infty} \approx \rho^2/\epsilon$ ;
- The linear regime: For  $\epsilon \ll \rho \ll 1/\epsilon$ ,  $\sigma_{\infty}^2 \approx \sqrt{\epsilon/\rho}$ ,  $\rho_{\infty} \approx \rho$ ;
- The saturation regime: For  $1/\epsilon \ll \rho$ ,  $\sigma_{\infty}^2 \approx \epsilon$ ,  $\rho_{\infty} \approx 1/\epsilon$ .

Figure 2 illustrates these three regimes for  $\epsilon = 10^{-4}$ . The different slopes of  $\rho_{\infty}$  on the log-log plot are clearly visible for the three regimes. The linear regime covers roughly 80 dB, from -40 dB to +40 dB, in this particular example.

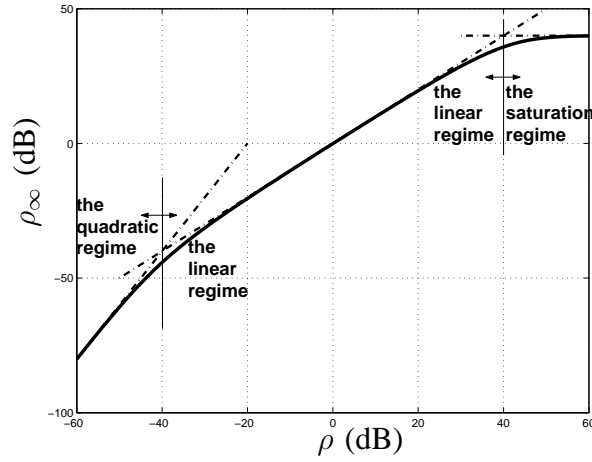


Fig. 2. Case study of the discrete-time Gauss-Markov fading model: Illustration of the three operating regimes,  $\epsilon = 10^{-4}$ .

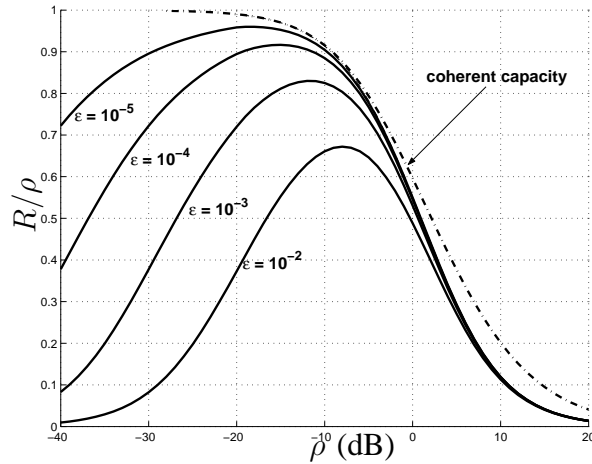


Fig. 3. Normalized rate  $R/\rho$  vs. SNR for recursive training with QPSK on the discrete-time Gauss-Markov fading channel. As a comparison, the dashed-dot curve is the channel capacity (normalized by SNR) with perfect receive CSI, achieved by circular complex Gaussian inputs.

An interesting observation is that the two SNR thresholds dividing the three regimes are determined by a single parameter  $\epsilon$ , which happens to be the one-step MMSE prediction error  $\sigma_{\text{pred}}^2$  for the discrete-time Gauss-Markov fading process. The  $1/\epsilon$  threshold dividing the linear and the saturation regimes coincides with that obtained in [15], where it is obtained for circular complex Gaussian inputs with nearest-neighbor decoding. In this paper we investigate PSK, which results in a penalty in the achievable rate at high SNR. More specifically, it can be shown that the achievable rate for  $\rho \gg 0$  behaves like  $(1/2) \cdot \log \min\{\rho, 1/\epsilon\} + O(1)$  [16].

A further observation relevant to low-SNR system design is that, the  $\epsilon$  threshold dividing the quadratic and the linear regimes clearly indicates when the low-SNR asymptotic channel behavior becomes dominant. Since the innovation rate  $\epsilon$  for underspread fading processes is typically small, we essentially have a low-SNR coherent channel above  $\rho = \epsilon$ . This suggests that there may be an “optimal” SNR at which the low-SNR coherent capacity limit is the most closely approached. Figure 3 plots the normalized achievable rate  $R/\rho$  vs. SNR, in which the achievable rate  $R$  is numerically evaluated for the induced channel (5) using QPSK. Although all the curves vanish rapidly below the threshold  $\rho = \epsilon$ , for certain  $\rho > \epsilon$ , the normalized achievable rate can be reasonably close to 1. For example, taking  $\epsilon = 10^{-4}$ , the “optimal” SNR is  $\rho \approx -15$  dB, and the corresponding normalized achievable rate is above 0.9, *i.e.*, more than 90% of the low-SNR coherent capacity limit is achieved.

#### IV. FILLING THE GAP TO CAPACITY BY WIDENING THE INPUT BANDWIDTH

In Section III we have investigated the achievable information rate of the discrete-time channel (2), which is obtained from the continuous-time channel (1) as described in Section II. The symbol duration  $T$  there is a fixed system parameter. In this section we will show that, if we are allowed to reduce  $T$ , *i.e.*, widen the input bandwidth, then the recursive training scheme using PSK achieves an information rate that is asymptotically consistent with the channel capacity of the continuous-time channel (1) under peak envelope  $P$ . More specifically, we have the following theorem.

*Theorem 4.1:* For the continuous-time channel (1) with envelope  $P > 0$ , as the symbol duration  $T \rightarrow 0$ , its induced channel (5) achieves

$$\lim_{T \rightarrow 0} \frac{R}{T} = \left[ 1 - \frac{1}{P} \cdot \frac{1}{2\pi} \int_{-\infty}^{\infty} \log(1 + P \cdot S_{H_c}(j\omega)) d\omega \right] \cdot P. \quad (19)$$

Again we compare the asymptotic achievable rate (19) to a capacity upper bound based upon the capacity per unit energy. For the continuous-time channel (1), the capacity per unit energy under a peak envelope constraint  $P > 0$  is [1]

$$\dot{C} = 1 - \frac{1}{2\pi P} \cdot \int_{-\infty}^{\infty} \log(1 + P \cdot S_{H_c}(j\omega)) d\omega, \quad (20)$$

and the related capacity upper bound (measured per unit time) is [1]

$$C \leq U(P) \stackrel{\text{def}}{=} \left[ 1 - \frac{1}{2\pi P} \cdot \int_{-\infty}^{\infty} \log(1 + P \cdot S_{H_c}(j\omega)) d\omega \right] \cdot P. \quad (21)$$

Comparing (19) and (21), it is surprising to notice that these two quantities coincide. Recalling that in Section III we have noticed a  $(1/2) \cdot \rho^2 + o(\rho^2)$  rate penalty in discrete-time channels, we conclude that widening the input bandwidth eliminates this penalty and essentially results in an asymptotically capacity-achieving scheme in the wideband regime.<sup>5</sup>

The channel capacity of continuous-time peak-limited wideband fading channels (19) was originally obtained in [17]. However, in [17] the capacity is achieved by frequency-shift keying (FSK), which is bursty in frequency. In our Theorem 4.1, we show that the capacity is also achievable if we employ recursive training and PSK, which is bursty in neither time nor frequency.

After some manipulations of (19), we further obtain the following

- As  $P \rightarrow 0$ ,

$$\frac{\lim_{T \rightarrow 0}(R/T)}{P^2} \rightarrow \frac{1}{2} \cdot \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{H_c}^2(j\omega) d\omega, \quad (22)$$

if the above integral exists.

- As  $P \rightarrow \infty$ ,

$$\frac{\lim_{T \rightarrow 0}(R/T)}{P} \rightarrow 1. \quad (23)$$

In the sequel we will see that (22) and (23) are useful for asymptotic analysis.

<sup>5</sup>Again, the same caveat as in footnote 2 applies.

### A. Case Study: The Continuous-Time Gauss-Markov Fading Model

In this subsection, we apply Theorem 4.1 to analyze a continuous-time Gauss-Markov fading process. Such a process has autocorrelation function

$$K_{H_c}(\tau) = (1 - \epsilon_c)^{|\tau|/2},$$

where the parameter  $0 < \epsilon_c \leq 1$  characterizes the channel variation, analogously to  $\epsilon$  for the discrete-time case in Section III. The spectral density function of the process is

$$S_{H_c}(j\omega) = \frac{|\log(1 - \epsilon_c)|}{\omega^2 + (\log(1 - \epsilon_c))^2/4}.$$

Applying Theorem 4.1, we find that the recursive training scheme using PSK with a wide bandwidth asymptotically achieves an information rate

$$\lim_{T \rightarrow 0} \frac{R}{T} = P - \frac{|\log(1 - \epsilon_c)|}{2} \cdot \left( \sqrt{1 + \frac{4P}{|\log(1 - \epsilon_c)|}} - 1 \right) \quad (24)$$

$$= \frac{1}{|\log(1 - \epsilon_c)|} \cdot P^2 + o(P^2) \quad \text{as } P \rightarrow 0. \quad (25)$$

### B. Case Study: Clarke's Fading Model

In this subsection, we apply Theorem 4.1 to analyze Clarke's fading process. Such a process is usually characterized by its spectral density function [18]

$$S_{H_c}(j\omega) = \begin{cases} \frac{2}{\omega_m} \cdot \frac{1}{\sqrt{1 - (\omega/\omega_m)^2}}, & \text{if } |\omega| \leq \omega_m \\ 0, & \text{otherwise,} \end{cases}$$

where  $\omega_m$  is the maximum Doppler frequency.

Applying Theorem 4.1, we find that

$$\lim_{T \rightarrow 0} \frac{R}{T} = \begin{cases} \frac{\omega_m}{\pi} \cdot \left\{ \log \frac{\omega_m}{P} - \sqrt{1 - (2P/\omega_m)^2} \cdot \log \frac{\omega_m \cdot [1 + \sqrt{1 - (2P/\omega_m)^2}]}{2P} \right\}, & \text{if } P \leq \omega_m/2 \\ \frac{\omega_m}{\pi} \cdot \left\{ \log \frac{\omega_m}{P} + \sqrt{(2P/\omega_m)^2 - 1} \cdot \arctan \sqrt{(2P/\omega_m)^2 - 1} \right\}, & \text{if } P > \omega_m/2. \end{cases} \quad (26)$$

For large  $P$ , the asymptotic behavior of (26) is consistent with (23). For small  $P$ , however, the integral in (22) diverges, hence the asymptotic behavior of (26) scales super-quadratically with  $P$ . After some manipulations of (26), we find that

$$\lim_{T \rightarrow 0} \frac{R}{T} = \frac{2}{\pi\omega_m} \cdot \log \frac{1}{P} \cdot P^2 + O(P^2) \quad \text{as } P \rightarrow 0. \quad (27)$$

## V. CONCLUDING REMARKS

For fading channels that exhibit temporal correlation, a key to enhancing communication performance is efficiently exploiting the implicit CSI embedded in the fading process. From the preceding developments in this paper, we see that a recursive training scheme, which performs channel estimation and demodulation/decoding in an alternating manner, accomplishes this job reasonably well, especially when the channel fading varies slowly. The main idea of recursive training is to repeatedly use decisions of previous information symbols as pilots, and to ensure the reliability of these decisions by coding over sufficiently long blocks.

Throughout this paper, we restrict the channel inputs to complex proper PSK, which is not optimal in general for Rayleigh fading channels without CSI. There are two main benefits of this choice. First, compared to other channel inputs such as circular complex Gaussian, PSK leads to a significant simplification of the analytical developments. As we saw, recursive training with PSK converts the original fading channel without CSI into a series of parallel

sub-channels, each with estimated receive CSI but additional noise that remains circular complex white Gaussian. In this paper we mainly investigate the steady-state limiting channel behavior; however, it may worth mentioning that, using the induced channel model presented in Section II, exact evaluation of the transient channel behavior is straightforward, with the aid of numerical methods.

Second, PSK inputs perform reasonably well in the moderate to low SNR regime. This is due to the fact that, for fading channels with perfect receive CSI, as SNR vanishes, channel capacity can be asymptotically achieved by rather general complex proper inputs besides circular complex Gaussian [10]. The main contribution of our work is that it clearly separates the effect of an input peak-power constraint and the effect of replacing optimal peak-limited inputs with PSK, which is bursty neither in time nor in frequency. It is shown that, for slowly time-varying fading processes, the rate loss from PSK inputs is essentially negligible. Furthermore, as revealed by the non-asymptotic analysis for discrete-time Gauss-Markov fading processes, there appear to be non-vanishing SNRs at which near-coherent performance is attainable with recursive training and PSK.

## REFERENCES

- [1] V. Sethuraman and B. Hajek, "Capacity per Unit Energy of Fading Channels with a Peak Constraint," *IEEE Trans. Inform. Theory*, vol. 51, no. 9, pp. 3102–3120, Sept. 2005.
- [2] W. Zhang and J. N. Laneman, "How Good is Phase-Shift Keying for Peak-Limited Rayleigh Fading Channels in the Low-SNR Regime?," *IEEE Trans. Inform. Theory*, Aug. 2005, Submitted for publication.
- [3] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading Channels: Information-Theoretic and Communications Aspects," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.
- [4] J. G. Proakis, *Digital Communications*, McGraw-Hill, Inc., New York, third edition, 1995.
- [5] T. Kailath, "Correlation Detection of Signals Perturbed by a Random Channel," *IEEE Trans. Inform. Theory*, vol. 6, no. 3, pp. 361–366, June 1960.
- [6] F. Neeser and J. L. Massey, "Proper Complex Random Processes with Applications to Information Theory," *IEEE Trans. Inform. Theory*, vol. 39, no. 4, pp. 1293–1302, July 1993.
- [7] A. J. Goldsmith and P. P. Varaiya, "Capacity, Mutual Information, and Coding for Finite-State Markov Channels," *IEEE Trans. Inform. Theory*, vol. 42, no. 3, pp. 868–886, May 1996.
- [8] T. Kailath, *Lectures on Wiener and Kalman Filtering*, Springer-Verlag, New York, 1981.
- [9] J. L. Doob, *Stochastic Processes*, John Wiley & Sons, Inc., New York, 1953.
- [10] V. V. Prelov and S. Verdú, "Second-Order Asymptotics of Mutual Information," *IEEE Trans. Inform. Theory*, vol. 50, no. 8, pp. 1567–1580, Aug. 2004.
- [11] M. Medard and R. G. Gallager, "Bandwidth Scaling for Fading Multipath Channels," *IEEE Trans. Inform. Theory*, vol. 48, no. 4, pp. 840–852, Apr. 2002.
- [12] B. Hajek and V. G. Subramanian, "Capacity and Reliability Function for Small Peak Signal Constraints," *IEEE Trans. Inform. Theory*, vol. 48, no. 4, pp. 828–839, Apr. 2002.
- [13] A. Lapidoth, "On the Asymptotic Capacity of Stationary Gaussian Fading Channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 437–446, Feb. 2005.
- [14] A. Lapidoth and S. M. Moser, "Capacity Bounds via Duality with Applications to Multiple-Antenna Systems on Flat Fading Channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2426–2567, Oct. 2003.
- [15] R. Etkin and D. N. C. Tse, "Degrees of Freedom in Underspread MIMO Fading Channels," *IEEE Trans. Inform. Theory*, May 2003, submitted; available at [http://www.eecs.berkeley.edu/~dtse/raul\\_hisnr.pdf](http://www.eecs.berkeley.edu/~dtse/raul_hisnr.pdf).
- [16] A. D. Wyner, "Bounds on Communication with Polyphase Coding," *Bell Sys. Tech. J.*, vol. 45, pp. 523–559, Apr. 1966.
- [17] A. J. Viterbi, "Performance of an M-ary Orthogonal Communication System Using Stationary Stochastic Signals," *IEEE Trans. Inform. Theory*, vol. 13, no. 3, pp. 414–422, July 1967.
- [18] W. C. Jakes and D. C. Cox, Eds., *Microwave Mobile Communications*, IEEE Press, New York, 1994.