Comparison of the Achievable Rates in OFDM and Single Carrier Modulation with I.I.D. Inputs

Yair Carmon*, Shlomo Shamai* and Tsachy Weissman†

Abstract

We compare the maximum achievable rates in single-carrier and OFDM modulation schemes, under the practical assumptions of i.i.d. finite alphabet inputs and linear ISI with additive Gaussian noise. We show that the Shamai-Laroia approximation serves as a bridge between the two rates: while it is well known that this approximation is often a *lower bound* on the single-carrier achievable rate, it is revealed to also essentially *upper bound* the OFDM achievable rate. We apply Information-Estimation relations in order to rigorously establish this result for both general input distributions and to sharpen it for commonly used PAM and QAM constellations. To this end, novel bounds on MMSE estimation of PAM inputs to a scalar Gaussian channel are derived, which may be of general interest. Our results show that, under reasonable assumptions, optimal single-carrier schemes may offer spectral efficiency significantly superior to that of OFDM, motivating further research of such systems.

I. INTRODUCTION

Intersymbol interference (ISI) is a ubiquitous impairment in communication and data storage media [1]. Techniques of information transmission over ISI channels can be roughly divided into two types: single-carrier (SC) modulation and multi-carrier modulation.

In SC modulation, symbols are transmitted at a rate approximately equal to the available bandwidth, and are distorted by ISI. The distortion can be compensated for at the receiver using different equalization techniques such as maximum-likelihood sequence estimation [2] and linear MMSE estimation, possibly with decision-feedback [3]. Equalization can also be combined with decoding, resulting in various iterative schemes [4].

In multi-carrier modulation, the available bandwidth is divided among several lower symbol-rate signals, each with a different carrier frequency, often referred to as subcarriers. Orthogonal frequency-division multiplexing (OFDM) [5], [6] is the most important multi-carrier modulation technique. In OFDM, the frequency spacing between subcarriers is chosen so that they are orthogonal on every signaling interval. Additionally, a portion of the end of each interval is copied to its beginning and is commonly referred to as the cyclic prefix. These two modifications allow for low-complexity optimal equalization using the FFT algorithm, which is the primary advantage of OFDM.

Today, OFDM is the predominant modulation technique in high-bandwidth communications over channels with significant ISI, and is featured in a large number of standards, including DSL [7], WiFi [8], WiMAX [9], DVB-T [10] and the LTE downlink [11]. However, OFDM waveforms suffer from a higher peak to average power ratio (PAPR) than SC waveforms. Due to their lower PAPR, and due to the introduction of efficient frequency-domain decision feedback equalization techniques [12], SC schemes have become a viable alternative to OFDM in certain settings. In particular, small, cheap and power-efficient amplifiers require low PAPR input due to their limited dynamic range, making SC desirable when they are used. Two such examples are the uplink air-interface of LTE, where a scheme called SC-FDMA is used [13], and the new 802.11ad specification, which includes a SC option [14].

The purpose of this paper is to compare the maximum achievable rates of reliable communication in OFDM and SC modulations, when optimal equalization and channel coding is assumed. Since optimal OFDM equalization is trivial, modern OFDM systems are able to approach this maximum theoretical rate using advanced coding schemes such as turbo codes or LDPC [15]. In contrast, optimal equalization and decoding cannot be decoupled in SC schemes, and instead must be approximated by iterative turbo equalization techniques [4]. Such techniques incur a high computational cost which currently renders them infeasible in practice. Nonetheless, with ever-growing computation resources and steadily improving iterative receivers (c.f. [16]), SC schemes able to approach the maximum achievable rate might soon become feasible. Therefore, comparison of the two maximum achievable rates is of practical as well as theoretical interest.

When only average power constraints apply on the channel input, it is well known that Gaussian signaling achieves the same maximum rate in both OFDM and SC schemes [17]. However, we impose two practical restrictions, which rule out the classical Gaussian solution. The first restriction is that the channel inputs take values in a certain fixed finite alphabet (constellation), as is always the case in practice. The second restriction is that the inputs are i.i.d., or more precisely that an i.i.d. random coding distribution is used. This restriction is justified as long as we limit ourselves to channel coding schemes that were designed for memoryless channels, as is common in practice. For OFDM the latter restriction is taken to mean that all subcarriers use the same distribution — this will simplify our calculations and it also accurately models wireless ISI channels, which most often change too rapidly for constellation loading to become practical.

Under these constraints, there is no closed-form expression for the achievable rate of SC modulation. However, an expression introduced by Shamai and Laroia [18] is known to tightly approximate the SC achievable rate in virtually all scenarios. The Shamai-Laroia approximation is intimately connected to the performance of decision-feedback equalization of SC signals.

In this paper, we show that in the sense of achievable rate and under the above-mentioned restrictions, SC modulation will often offer performance superior or equal to that of OFDM. In particular, we prove that for BPSK and QPSK inputs, the OFDM achievable rate is lower than the Shamai-Laroia approximation regardless of the ISI channel. A numerical study indicates this holds also for 4-PAM, 8-PSK, 16-QAM and 32-QAM inputs. For general finite-alphabet input distributions, we prove the same result in the low- and high-SNR regimes. For PAM and square QAM inputs, we find an SNR threshold above which the high-SNR regime is in effect. For square QAM inputs of order 256 and above, this threshold corresponds to a symbol error rate of more than 50% and is therefore reasonably

low. We provide an exact characterization of the maximum advantage OFDM may offer over the Shamai-Laroia approximation. Numerical evaluation shows that this advantage is very small for QAM constellations of orders up to 4096 (less than 0.1 bit). Complementing this result, we show that the advantage of the Shamai-Laroia approximation over the OFDM rate becomes arbitrarily large for some of ISI channels. Thus, SC schemes may offer a significant performance gain over OFDM in certain cases, but not vice versa. Finally, we study continuous uniform input, which is limiting case of increasingly high-order QAM inputs. We show that unlike the finite-alphabet cases, under uniform input the Shamai-Laroia approximation cannot significantly exceed the OFDM achievable rate. This indicates that the advantage of SC over OFDM will become small if a sufficiently dense input constellation is chosen. However, such inputs are not necessarily feasible. We provide a detailed discussion on practical scenarios in which SC is expected to offer a considerable performance gain over OFDM.

Our results stem from the concavity properties of the input-output mutual information in a scalar Gaussian channel, as a function of a rescaled SNR variable that we call the "log-SNR". In order to investigate these properties, we make extensive use of Information-Estimation results that link derivatives of the mutual information function and estimation-theoretic quantities [19]–[21]. We derive new bounds on MMSE estimation of PAM inputs to an additive Gaussian channel in the high-SNR regime. Besides their use in proving some of our main results, we believe them to be of general interest.

There is a considerable amount of literature that deals with comparison between SC and OFDM modulations. It is mostly concerned with comparison of specific schemes and quantities such as PAPR and bit or frame error rates (c.f. [22]–[27]). Some works also compare fundamental limits. In [28]–[30] achievable rates are considered, but Gaussian inputs are assumed in order to model adaptive constellation loading, leading to the expected conclusion that both methods offer the same rates. In [31] achievable rates are compared via simulation for binary input and two-taps ISI channels. The authors report that in these settings SC is superior, but make no general or theoretically-backed claims. In [32] it is conjectured that the SC achievable rate is always higher than the OFDM rate. This conjecture is supported by numerical evidence, but no theoretical analysis is performed. In [33] and [34] the cut-off rate is compared analytically, and the SC rate is shown to exceed the OFDM rate in several scenarios.

In a recent report [35] the authors independently found that concavity with respect to log-SNR yields an inequality between the OFDM achievable rate and the Shamai-Laroia approximation. Based on numerical evidence, they argue that concavity holds for QPSK and 16-QAM inputs and that while it does not always hold for higher order constellations, the maximum difference in favor of OFDM is small. However, the majority of our results remain exclusive to this work, including the analytic proofs of concavity, the study of the concave envelope and the application of Information-Estimation tools.

The rest of this paper is organized as follows. Section II formulates the problem and presents the main results, briefly discussing their extension when linear precoding is allowed. Section III introduces the log-SNR scale and studies the concavity of the mutual information function with respect to it, culminating in a proof to our main result for general finite-alphabet inputs. Section IV focuses on the nonlinear MMSE estimation of PAM random variables corrupted by additive Gaussian noise. In this section, a novel "pointwise" result is presented, bounding the conditional variance of the channel input given an observation of the channel output. This bound is then applied to

derive a tight high-SNR characterization of the MMSE function and its derivative. Section V uses insights from the two previous sections in order to prove our results pertaining PAM and square QAM inputs, as well as continuous uniform input. Section VI contains an in-depth discussion on the differences between the SC and OFDM achievable rates likely to occur in practice, and the implications of increasing the constellation order. Section VII concludes this paper.

II. PRELIMINARIES AND MAIN RESULTS

We consider a complex-valued, discrete-time ISI channel model,

$$y_k = \sum_{i=0}^{L-1} h_i x_{k-i} + n_k, \tag{1}$$

where $x_{-\infty}^{\infty}$ is the channel input sequence[1], h_0^{L-1} are arbitrary complex-valued ISI taps and $n_{-\infty}^{\infty}$ is an i.i.d. standard complex Gaussian[2] sequence independent of the input. Here L denotes the length of the channel impulse response. Let $H(\theta) = \sum_{k=0}^{L-1} h_k e^{-jk\theta}$ be the ISI channel transfer function. We assume throughout the paper that the input sequence has zero mean and unit average power, *i.e.* $E|x_i|^2 = 1$. Since the input and noise are both normalized to unit variance, the quantity $\sum_{k=0}^{L-1} |h_k|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |H(\theta)|^2 d\theta$, which is not normalized to unity, expresses the input signal-to-noise ratio (SNR).

A. Single carrier modulation model

In our model for single-carrier modulation, the channel input sequence is assumed to be i.i.d., zero-mean and unit power, with every input symbol drawn from a finite complex-valued alphabet also known as a signal constellation. Conventional constellations are composed of 2^m uniformly spaced symbols, each representing m data bits. Commonly used constellations include BPSK, QPSK, 8-PSK, 16-QAM and 64-QAM [1]. Unless specifically mentioned otherwise, our results apply for any (finite alphabet) input distribution. However, when we refer to a certain input distribution by its constellation name (*e.g.* "BPSK input" or "256-QAM" input), a uniform distribution over the constellation points will be assumed. The maximum achievable rate for reliable communication under the assumptions of this model is given by the input-output Average Mutual Information [36]:

$$\mathcal{I}_{\text{SC}} \triangleq \lim_{K \to \infty} \frac{1}{2K+1} I\left(x_{-K}^K \,;\, y_{-K}^K\right) = I\left(x_0 \,;\, y_{-\infty}^{\infty} \mid x_{-\infty}^{-1}\right) \tag{2}$$

where $I(A \,;\, B)$ denotes the mutual information between A and B, and $I(A \,;\, B \mid C)$ denotes the mutual information between A and B, conditioned on C (c.f. [37]).

When the input distribution is symmetric Gaussian, a closed-form expression for $\mathcal{I}_{\text{SC}}$ can be easily derived (cf. [17], [18]),

$$\mathcal{I}_{\text{SC,Gaussian}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log\left(1 + |H(\theta)|^2\right) d\theta \tag{3}$$

---

[1]We use the standard notation $a_{N_1}^{N_2}$ for the sequence $[a_{N_1}, a_{N_1+1}, ..., a_{N_2}]$ with the natural interpretation when $N_1 = -\infty$ and/or $N_2 = \infty$.
[2]A standard complex Gaussian variable is of the form $n^I + jn^Q$, where $n^I, n^Q \sim \mathcal{N}(0, 1/2)$ and $n^I \perp n^Q$.

Let

$$I_x\left(\gamma\right) \triangleq I\left(x \,;\, \sqrt{\gamma}x + n\right) \tag{4}$$

stand for the input-output mutual information in a scalar complex-valued Gaussian channel with unit-power input[3] x, SNR γ and standard complex Gaussian noise n, independent of x. The Gaussian achievable rate can also be expressed as

$$\mathcal{I}_{\text{SC,Gaussian}} = I_{\text{Gaussian}}\left(\text{SNR}_{\text{MMSE-DFE-U}}\right) \tag{5}$$

where

$$\text{SNR}_{\text{MMSE-DFE-U}} = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left(1 + |H(\theta)|^2\right)d\theta\right\} - 1 \tag{6}$$

is the output SNR of the unbiased MMSE linear estimator of x_0 given $x_{-\infty}^{-1}$ and $y_{-\infty}^{\infty}$, known as the MMSE decision-feedback equalizer (DFE), and $I_{\text{Gaussian}}\left(\gamma\right) = \log\left(1 + \gamma\right)$ is the input-ouput mutual information for a complex-valued Gaussian channel with standard complex Gaussian input. A concise presentation of the MMSE descision-feedback equalizer, as well as a derivation of its output SNR can be found in [3].

When the input distribution in not Gaussian, no closed-form expression for $\mathcal{I}_{\text{SC}}$ is known and it must be approximated either analytically [18], [38], [39] or by Monte-Carlo simulations [40]–[42]. A simple and often-used approximation for $\mathcal{I}_{\text{SC}}$ was first proposed by Shamai and Laroia [18],

$$\mathcal{I}_{\text{SC}} \approx I_{\text{SL}} \triangleq I_x\left(\text{SNR}_{\text{MMSE-DFE-U}}\right) \tag{7}$$

where x is a random variable distributed as one of the input symbols to the ISI channel. This approximation can be derived by applying the MMSE DFE on the channel output sequence and replacing the residual ISI by independent Gaussian variables with equal power — however, as explained in [18], the central limit theorem cannot be used to rigorously justify this approximation, as the residual ISI coefficients do not meet its conditions.

The Shamai-Laroia approximation was originally conjectured to be a lower bound on $\mathcal{I}_{\text{SC}}$. However, in [39] a counterexample based on highly skewed binary input is constructed, showing that this conjecture does not hold for all input distributions. Nonetheless, extensive experimentation has shown that when conventional input distributions are used, I_{SL} is an extremely tight lower bound for $\mathcal{I}_{\text{SC}}$ for any ISI channel and SNR [18], [38], [40]. In particular, there is no known counterexample to $\mathcal{I}_{\text{SC}} \geq I_{\text{SL}}$ that involves symmetrically distributed inputs (as all conventional inputs are). Moreover, in [39] the lower bound $\mathcal{I}_{\text{SC}} \geq I_{\text{SL}}$ is proven to hold for sufficiently high SNR, further establishing its validity. Whether $\mathcal{I}_{\text{SC}} \geq I_{\text{SL}}$ can be proven to always hold for specific input distributions is a question open to future research.

B. OFDM modulation model

In OFDM, information is transmitted in blocks of $N + N_{CP}$ channel inputs, where the first N_{CP} elements of each block are identical to its last N_{CP} elements and thus constitute a cyclic prefix (CP). With the CP discarded at

---

[3]By "Gaussian channel" we mean a channel with additive Gaussian noise independent of the input, which is not necessarily Gaussian. The subscript x will commonly be used to indicate that a general input distribution x is discussed, and subscripts with indicative names will be employed when referring to specific input distributions, *e.g.* $I_{\text{BPSK}}\left(\gamma\right)$ and $I_{\text{Gaussian}}\left(\gamma\right)$.

the receiver, the ISI channel is transformed into a vector channel,

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{8}$$

The vectors $\mathbf{y}$ and $\mathbf{x}$ represent blocks of N channel outputs and inputs, respectively, $\mathbf{n}$ is a standard complex Gaussian noise vector, and $\mathbf{H}$ is a matrix representing the ISI. Practical OFDM schemes are designed so that the cyclic prefix is longer than the channel memory ($N_{CP} > L$). Moreover, we consider a sequence of schemes in which the block size N and the cyclic prefix size N_{CP} grow to infinity, so that N_{CP} is guaranteed to exceed L eventually. Assuming $N_{CP} > L$, $\mathbf{H}$ is a circulant matrix, with first row equal to $[h_0, 0 \cdots 0, h_{L-1}, \cdots, h_1]$. Therefore, $\mathbf{H}$ is diagonalized by the the DFT matrix of order N,

$$\mathbf{H^d} = \mathbf{WHW}^{-1} \tag{9}$$

with $\mathbf{H^d}$ a diagonal matrix and $W_{m,k} = \frac{1}{\sqrt{N}} e^{-2\pi jmk/N}$ the DFT matrix. Applying the input precoding $\mathbf{x} = \mathbf{W}^{-1}\tilde{\mathbf{x}}$ and output transformation $\tilde{\mathbf{y}} = \mathbf{W}\mathbf{y}$ thus yields an equivalent diagonal vector channel,

$$\tilde{\mathbf{y}} = \mathbf{H^d}\tilde{\mathbf{x}} + \tilde{\mathbf{n}} \tag{10}$$

We assume that the elements of $\tilde{\mathbf{x}}$ are i.i.d.[4], zero mean and have unit average power. Since the channel model in (10) is simply N parallel channels, the maximum achievable rate per channel input is given by

$$\mathcal{I}_{\text{OFDM}}^{(N)} \triangleq \frac{1}{N + N_{CP}} I\left(\tilde{\mathbf{x}}; \tilde{\mathbf{y}}\right) = \frac{1}{N + N_{CP}} \sum_{i=1}^{N} I_x\left(|H_{i,i}^d|^2\right) \tag{11}$$

with $I_x\left(\cdot\right)$ as defined in (4) and x distributed as one of the elements of $\tilde{\mathbf{x}}$. The Toeplitz Distribution Theorem [43] allows us to take the limit of the large block size,

$$\mathcal{I}_{\text{OFDM}} \triangleq \lim_{N \to \infty} \mathcal{I}_{\text{OFDM}}^{(N)} = \frac{1}{2\pi} \int_{-\pi}^{\pi} I_x\left(|H(\theta)|^2\right) d\theta \tag{12}$$

where it is assumed that N_{CP} grows as $o(N)$, so that the rate overhead of the cyclic prefix vanishes.

C. MMSE estimation in a scalar Gaussian channel

Consider once more the scalar complex-valued Gaussian channel $y = \sqrt{\gamma}x + n$ with input x and standard complex Gaussian noise n independent of x. The minimum mean square error (MMSE) in estimating x from y is given by

$$\text{mmse}_x\left(\gamma\right) = \text{E}\left|x - \text{E}\left[x|y\right]\right|^2 \tag{13}$$

For a standard complex Gaussian input we have $\text{mmse}_{\text{Gaussian}}\left(\gamma\right) = 1/\left(1 + \gamma\right) \geq \text{mmse}_x\left(\gamma\right)$, for any other unit-power input x.

---

[4]This is usually the case in wireless links, where the communication overhead of coordinating different powers and constellations for different subcarriers often makes doing so undesirable.

Assuming x has zero mean and unit variance (*i.e.* $\mathrm{E}\,|x|^2 = 1$), the connection between the mutual information (4) and the MMSE (13) is given by:

$$I'_x(\gamma) = \mathrm{mmse}_x(\gamma) \tag{14}$$

Note that the above equation differs from the familiar Guo-Shamai-Verdú formula [19] by a factor of 2 on the right-hand side. This is due to the fact that our channel model is complex-valued. The relation (14) can be easily derived from the vector version of the GSV theorem (eq. (22) in [19]), by considering a two-dimensional scalar channel matrix.

D. Statement of results

Our main result provides a connection between $\mathcal{I}_{\mathrm{OFDM}}$ and I_{SL} for finite-alphabet inputs. First, we show that an inequality of the form $\mathcal{I}_{\mathrm{OFDM}} \leq I_{\mathrm{SL}} + \Delta_x$ always holds, where $\Delta_x \geq 0$ depends only on the input distribution (and not on the ISI channel). Second, we characterize low and high SNR regions in which the strengthened inequality $\mathcal{I}_{\mathrm{OFDM}} \leq I_{\mathrm{SL}}$ holds, even when $\Delta_x \neq 0$. To this end, we introduce two pairs of SNR thresholds. The first pair, denoted $\underline{\gamma}_1$ and $\bar{\gamma}_2$, constitutes low and high SNR thresholds with respect to $\mathsf{SNR}_{\mathrm{MMSE\text{-}DFE\text{-}U}}$, defined in (6) — when $\mathsf{SNR}_{\mathrm{MMSE\text{-}DFE\text{-}U}}$ is below $\underline{\gamma}_1$ or above $\bar{\gamma}_2$, we have $\mathcal{I}_{\mathrm{OFDM}} \leq I_{\mathrm{SL}}$. The second threshold pair is denoted $\underline{\gamma}_0$ and $\bar{\gamma}_0$, and relates to the channel's frequency response — when $|H(\theta)|^2$ is bounded by $\underline{\gamma}_0$ from above or by $\bar{\gamma}_0$ from below, we are guaranteed once more to have $\mathcal{I}_{\mathrm{OFDM}} \leq I_{\mathrm{SL}}$. Like Δ_x, these thresholds depend only on the input distribution.

The explicit construction of $\underline{\gamma}_0$, $\bar{\gamma}_0$, $\underline{\gamma}_1$, $\bar{\gamma}_2$ and Δ_x is deferred to Section III and Definitions 1, 2 and 3 therein, as it relies on concepts and results developed there. These quantities are defined in terms of concavity properties of a certain function, and are straightforward to evaluate numerically. In particular, we provide an expression for Δ_x in terms of a three-variable optimization problem which is easily solved numerically — see (16) below.

Formally stated, our main result is as follows,

Theorem 1. *For any ISI channel and any finite alphabet distribution x,*

$$\mathcal{I}_{\mathrm{OFDM}} \leq I_{\mathrm{SL}} + \Delta_x \tag{15}$$

Where $\Delta_x \geq 0$ is given by

$$\Delta_x \triangleq \sup_{\substack{\gamma_1, \gamma_2, \gamma \text{ s.t.} \\ \gamma_1 \leq \gamma \leq \gamma_2}} \frac{\log\left(\frac{1+\gamma}{1+\gamma_1}\right)[I_x(\gamma_2) - I_x(\gamma)] - \log\left(\frac{1+\gamma_2}{1+\gamma}\right)[I_x(\gamma) - I_x(\gamma_1)]}{\log\left([1+\gamma_2]/[1+\gamma_1]\right)} \tag{16}$$

Additionally, if $\Delta_x > 0$, there exit $0 < \underline{\gamma}_1 \leq \underline{\gamma}_0 \leq \bar{\gamma}_0 \leq \bar{\gamma}_2 < \infty$ that depend only on the input distribution, such that $\mathcal{I}_{\mathrm{OFDM}} \leq I_{\mathrm{SL}}$ holds whenever the channel transfer function $H(\theta)$ satisfies at least one of the following conditions:

1) $\mathsf{SNR}_{\mathrm{MMSE\text{-}DFE\text{-}U}} \in [0, \underline{\gamma}_1] \cup [\bar{\gamma}_2, \infty)$
2) $|H(\theta)|^2 \leq \underline{\gamma}_0$ *for every* $\theta \in (-\pi, \pi)$
3) $|H(\theta)|^2 \geq \bar{\gamma}_0$ *for every* $\theta \in (-\pi, \pi)$

Next, we provide the following results, which sharpen Theorem 1 for specific input distributions,

Theorem 2. *For BPSK and QPSK inputs, $\Delta_x = 0$ and so $\mathcal{I}_{\text{OFDM}} \leq I_{\text{SL}}$ for every ISI channel.*

Theorem 3. *For M-PAM and square M^2-QAM inputs, $(d_{\min}/2)^2 \bar{\gamma}_0 \leq 1$, where $d_{\min}$ is the minimum distance between input symbols, assuming unit input power.*

Combined with Theorem 1, Theorem 3 implies that for M-PAM and square M^2-QAM inputs, $\mathcal{I}_{\text{OFDM}} \leq I_{\text{SL}}$ whenever the ISI channel is such that $(d_{\min}/2)^2 |H(\theta)|^2 \geq 1$ for every $\theta \in (-\pi, \pi)$. Since the uncoded symbol error rate in OFDM subcarrier frequency θ_0 is a function of $(d_{\min}/2)^2 |H(\theta_0)|^2$, the following corollary is immediate,

Corollary 1. *For a given ISI channel and square M^2-QAM inputs with $M \geq 16$, if the uncoded symbol error rate is below 50% in all OFDM subcarriers, $\mathcal{I}_{\text{OFDM}} \leq I_{\text{SL}}$.*

Our last result deals with uniformly distributed input. This input distribution represents the limit of infinitely high-order QAM, and is therefore referred to also as ∞-QAM. Since this input has an infinite alphabet, Theorem 1 does not apply to it. Instead, we have

Theorem 4. *For uniformly distributed complex input and any ISI channel,*

$$- \tilde{\Delta}_{\infty\text{-QAM}} \leq \mathcal{I}_{\text{OFDM}} - I_{\text{SL}} \leq \bar{\Delta}_{\infty\text{-QAM}} \left(\max_{\theta \in (-\pi, \pi)} |H(\theta)|^2 \right) \tag{17}$$

where $\tilde{\Delta}_{\infty\text{-QAM}} \approx 0.0608$ [bit], and $\bar{\Delta}_{\infty\text{-QAM}} (\gamma)$, is a non-decreasing function that satisfies $\bar{\Delta}_{\infty\text{-QAM}} (\gamma) = 0$ for every $\gamma \leq \underline{\gamma}_0^{(\infty\text{-QAM})} \approx 8.76$ [dB], and

$$\lim_{\gamma \to \infty} \bar{\Delta}_{\infty\text{-QAM}} (\gamma) \triangleq \Delta_{\infty\text{-QAM}} = \log (\pi e / 6) \approx 0.509 \text{ [bit]} \tag{18}$$

is the uniform input shaping loss with respect to Gaussian input.

Moreover, $\mathcal{I}_{\text{OFDM}} \geq I_{\text{SL}}$ when $H(\theta)$ satisfies at least one of the following conditions:
1) *$\text{SNR}_{\text{MMSE-DFE-U}} \geq \tilde{\gamma}_2^{(\infty\text{-QAM})} \approx 16.5$ [dB]*
2) *$|H(\theta)|^2 \geq \underline{\gamma}_0^{(\infty\text{-QAM})} \approx 8.76$ [dB] for every $\theta \in (-\pi, \pi)$*

Note that Theorem 4 provides some of the guarantees of Theorem 1. In particular, we have $\mathcal{I}_{\text{OFDM}} \leq I_{\text{SL}} + \Delta_{\infty\text{-QAM}}$ for any ISI channel, as well as $\mathcal{I}_{\text{OFDM}} \leq I_{\text{SL}}$ for every channel that satisfies $|H(\theta)|^2 \leq \underline{\gamma}_0^{(\infty\text{-QAM})}$. Figure 1 graphs $\bar{\Delta}_{\infty\text{-QAM}} (\gamma)$, which is evaluated numerically based on the analysis carried out in subsection V-D. As seen in the figure, the convergence of $\bar{\Delta}_{\infty\text{-QAM}} (\gamma)$ to $\Delta_{\infty\text{-QAM}}$ is extremely slow. Thus, for any practical purpose we may select an SNR level $\bar{\gamma}$ that is much higher than any plausible value of $|H(\theta)|^2$, and use $\bar{\Delta}_{\infty\text{-QAM}}(\bar{\gamma})$ instead of $\Delta_{\infty\text{-QAM}}$. Depending on the application, appropriate choices of $\bar{\gamma}$ are likely to yield values between $\bar{\Delta}_{\infty\text{-QAM}} (30 \text{ dB}) \approx 0.0841$ and $\bar{\Delta}_{\infty\text{-QAM}} (60 \text{ dB}) \approx 0.228$.

Table I summarizes a numerical study of the quantities that appear in Theorem 1. The table is consistent with Theorem 2 and indicates it also extends to 4-PAM, 8-PSK, 16-QAM and 32-QAM inputs. It also reveals that while nonzero, $\Delta_{64\text{-QAM}}$ is negligible, being of the order of a millionth of a bit. For higher order constellations Δ_x is more significant, but remains quite small even in very high-order constellations such as 4096-QAM. The limiting

Table I
NUMERICAL EVALUATION OF THE QUANTITIES APPEARING IN THEOREM 1

Input	$\left(\frac{d_{\min}}{2}\right)^2$ [dB]	$\underline{\gamma}_1$ [dB]	$\underline{\gamma}_0$ [dB]	$\bar{\gamma}_0$ [dB]	$\bar{\gamma}_2$ [dB]	Δ_x [bits]
BPSK, 4-PAM, QPSK, 8-PSK, 16-QAM, 32-QAM	(varies)	-	-	-	-	0
64-QAM	-16.2	10.7	11.0	11.7	12.0	$1.86 \cdot 10^{-6}$
256-QAM	-22.3	5.29	9.01	19.2	21.0	0.0202
1024-QAM	-28.3	3.63	8.81	25.2	27.5	0.0585
4096-QAM	-34.4	2.60	8.77	31.0	33.6	0.0987
∞-QAM	$-\infty$	-	8.76	-	-	0.509

Note: All values are rounded to three significant digits.

case of ∞-QAM input is also included in the table. It is seen that 4096-QAM has a value of $\underline{\gamma}_0$ quite close to that of ∞-QAM, but that $\Delta_{4096\text{-QAM}}$ is still far from converging to $\Delta_{\infty\text{-QAM}}$. Examining Figure 1, it is seen that for the various QAM inputs considered, Δ_x is well-approximated by $\bar{\Delta}_{\infty\text{-QAM}}(\bar{\gamma}_2)$. Finally, the table shows that the general bound provided by Theorem 3 is slack by an approximate factor of 2 — *i.e.*, for higher-order QAM, $(d_{\min}/2)^2 \bar{\gamma}_0 \approx 1/2$.

Our results show that $\mathcal{I}_{\text{OFDM}}$ may only exceed I_{SL} by a small amount, but the opposite is not true. Indeed, in subsection VI-A we construct a family of channels for which $\mathcal{I}_{\text{OFDM}}$ tends to zero while I_{SL} tends to the input entropy and in subsection VI-B we discuss practical scenarios in which the SC achievable rate is significantly higher than the OFDM rate. However, Theorem 4 indicates that this difference can be made small by increasing the constellation order. Subsection VI-C further discusses this course of action.

E. A note on linear precoding

Our results extend straightforwardly to the following generalized problem setting. In the SC case, we add a linear precoding filter that is applied on the i.i.d. inputs prior to their transmission. In the OFDM case, we allow a different power allocation for each subcarrier. More concretely, the SC precoded transmitted symbols are

$$x_n^{\text{precoded}} = \sum_i c_i x_{n-i} \qquad (19)$$

where $x_{-\infty}^{\infty}$ is an i.i.d. input sequence and the precoder taps satisfy $\sum_i |c_i|^2 = 1$. The OFDM transmitted symbols are

$$\tilde{x}_i^{\text{precoded}} = \sqrt{P_i}\tilde{x}_i \qquad (20)$$

where $\tilde{x}_1^N$ are the i.i.d. OFDM block inputs as in (10), and the power allocations satisfy $\sum_i P_i = 1$.

Clearly, precoded SC modulation with taps $c_{-\infty}^{\infty}$ and ISI channel $|H(\theta)|^2$ is equivalent to normal SC with channel
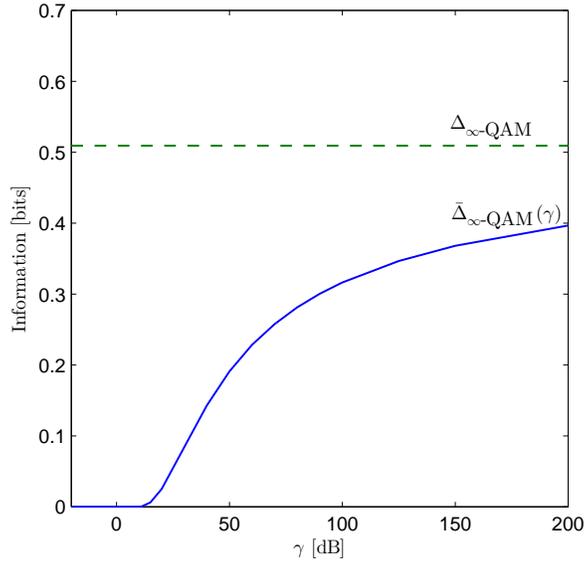
Figure 1. Numerical evaluation of $\bar{\Delta}_{\infty\text{-QAM}}(\gamma)$.

$|C(\theta)H(\theta)|^2$ where $C(\theta) = \sum_k c_k e^{-ik\theta}$. Moreover, OFDM with non-uniform power allocation is equivalent to normal OFDM with channel $P(\theta)|H(\theta)|^2$ where $P(2\pi i/N) = P_i$. Equating $|C(\theta)|^2$ with $P(\theta)$ we conclude that for any SC linear precoder there exist an OFDM power allocation such that both yield the same equivalent ISI channel. Hence, given a degree of freedom in choosing any SC linear precoder and any OFDM power allocation policy, our results are still applicable, revealing that SC has significant advantages over OFDM in this case as well.

The introduction of linear precoding lends additional viability to our assumption of i.i.d. input, as the capacity-achieving SC scheme can be viewed as i.i.d. Gaussian inputs linearly precoded with an optimal Waterfilling filter. Thus, it is reasonable to assume that when combined with a suitably chosen linear precoder, statistically independent symbols will be close to optimal even when input alphabet constraints prohibit Gaussian signaling. For OFDM with independent non-Gaussian inputs, an optimal power allocation policy called Mercury/Waterfilling was proposed in [44]. However, the Mercury/Waterfilling spectrum does not necessarily describe the optimal linear precoder for the i.i.d. non-Gaussian single carrier case. Using the methods described in [45], it should be possible to find the taps of this optimal precoder.

III. CONCAVITY OF MUTUAL INFORMATION WITH RESPECT TO LOG-SNR

A. Log-SNR scale

For a given SNR γ, define the log-SNR as $\zeta = \log(1 + \gamma)$, and let

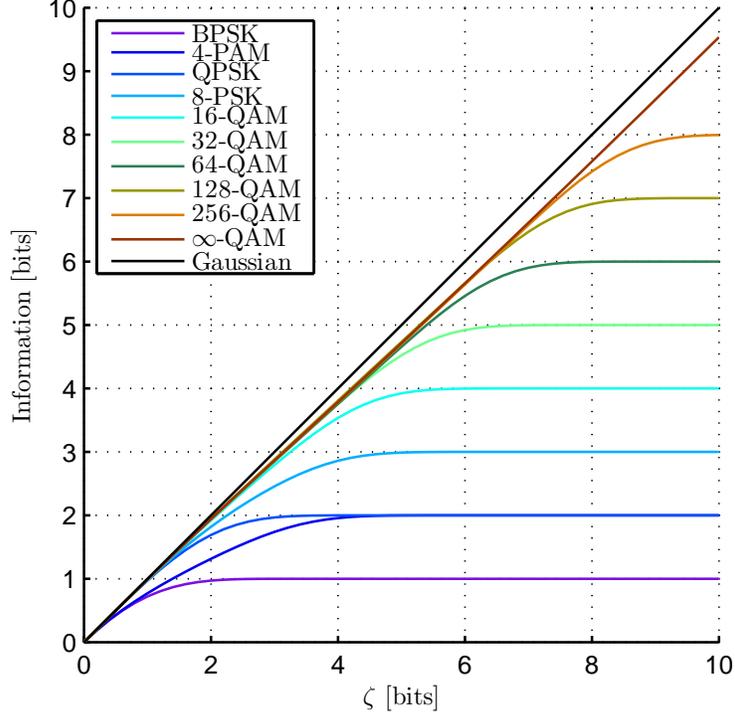$$I_x^{\log}(\zeta) \triangleq I_x(e^\zeta - 1) \tag{21}$$

Figure 2. $I_x^{\log}(\zeta)$ for some common input distributions.

be the input output mutual information, as a function of log-SNR, for a scalar complex Gaussian channel with zero-mean, unit-variance input x. Since ζ is identical to $I_{\mathrm{Gaussian}}^{\log}(\zeta)$, it is naturally measured in units of information. Moreover, we have $I_x^{\log}(\zeta) \leq I_{\mathrm{Gaussian}}^{\log}(\zeta) = \zeta$ for all inputs. Figure 2 shows $I_x^{\log}(\zeta)$ for some common input distributions. As can be seen in the figure, $I_x^{\log}(\zeta)$ is nearly linear for low ζ and, for finite alphabet inputs, it is nearly constant for high ζ, with the shoulder occurring at around the input entropy.

The main results of this paper hinge on the concavity properties of $I_x^{\log}(\zeta)$. In this section we study these properties for a general input distribution. We begin by showing that $I_x^{\log}(\zeta)$ is concave for sufficiently low and sufficiently high ζ. Next, we consider the concave envelope of $I_x^{\log}(\zeta)$ and show that it must equal $I_x^{\log}(\zeta)$ for sufficiently low and sufficiently high ζ. Finally, we apply these conclusions to prove Theorem 1.

B. Asymptotic concavity results

Proposition 1. *For every input distribution x, there exists $0 < \zeta_0$ such that $I_x^{\log}(\zeta)$ is concave for every $\zeta \in [0, \zeta_0]$.*

Proof: Setting $\gamma = e^\zeta - 1$ and differentiating $I_x^{\log}$ twice, we find that

$$
\begin{aligned}
I_x^{\log}{}''(\zeta) &= e^{2\zeta} I_x''(e^\zeta - 1) + e^\zeta I_x'(e^\zeta - 1) \\
&= (1+\gamma)^2 I_x''(\gamma) + (1+\gamma) I_x'(\gamma) \\
&= (1+\gamma) [\mathrm{mmse}_x(\gamma) + (1+\gamma) \mathrm{mmse}_x'(\gamma)] \tag{22} \\
&= (1+\gamma) \frac{d}{d\gamma} [(1+\gamma) \mathrm{mmse}_x(\gamma)] = (1+\gamma) r_x'(\gamma) \tag{23}
\end{aligned}
$$

where the transition to (22) is due to the I-MMSE relation (14). The function $r_x(\gamma) \triangleq (1+\gamma) \mathrm{mmse}_x(\gamma)$ denotes the ratio between the MMSE's of the non-linear and linear optimal estimators of x in the scalar complex Gaussian channel with SNR γ. Clearly, $r_x(\gamma) \le 1$, and $r_x(0) = 1$. Therefore, by continuity there must be a neighborhood of 0, denoted by $[0, \gamma_0]$, in which r_x is decreasing. Hence, by (23) we find that $I_x^{\log}(\zeta)$ is concave in $[0, \zeta_0]$, with $\zeta_0 = \log(1 + \gamma_0)$. ∎

Proposition 2. *For every input distribution x over a finite alphabet, there exists $\zeta_0 < \infty$ such that $I_x^{\log}(\zeta)$ is concave for every $\zeta \in [\zeta_0, \infty]$.*

Proof: Let $d_{\min}$ denote the minimum distance between any two symbols in the input alphabet. By the standard probability of error upper bound (c.f. Appendix C in [44]), we have

$$
\mathrm{mmse}_x(\gamma) \le D^2 e^{-(d_{\min}/2)^2 \gamma} \tag{24}
$$

for some $D > 0$. Moreover in Appendix A it is shown that

$$
\mathrm{mmse}_x'(\gamma) \le -C \frac{e^{-(d_{\min}/2)^2 \gamma}}{\sqrt{\gamma}} \tag{25}
$$

for sufficiently large γ and some $C > 0$. Therefore, denoting again $\gamma = e^\zeta - 1$ and substituting the above bounds in (22), we find that

$$
I_x^{\log}{}''(\zeta) \le (1+\gamma) \left(D^2 - \frac{1+\gamma}{\sqrt{\gamma}} C \right) e^{-(d_{\min}/2)^2 \gamma} \tag{26}
$$

for some $C > 0$ and sufficiently large γ. Clearly, this implies $I_x^{\log}{}''(\zeta) < 0$ for sufficiently large ζ. ∎

Definition 1. Let $0 < \underline{\zeta}_0 \le \infty$ be the *maximal* ζ_0 for which $I_x^{\log}(\zeta)$ is concave for every $\zeta \in [0, \zeta_0]$ and similarly let $0 \le \bar{\zeta}_0 < \infty$ be the *minimal* ζ_0 for which $I_x^{\log}(\zeta)$ is concave for every $\zeta \in [\zeta_0, \infty)$. Let $\underline{\gamma}_0 = e^{\underline{\zeta}_0} - 1$ and $\bar{\gamma}_0 = e^{\bar{\zeta}_0} - 1$ denote the SNR's corresponding to $\underline{\zeta}_0$ and $\bar{\zeta}_0$, respectively.

Notice that $\underline{\zeta}_0 < \bar{\zeta}_0$ if and only if $I_x^{\log}(\zeta)$ is not a concave function of ζ, in which case

$$
I_x^{\log}{}''(\underline{\zeta}_0) = I_x^{\log}{}''(\bar{\zeta}_0) = 0 \tag{27}
$$

C. Concave envelope

Let $\hat{I}_x^{\log}(\zeta)$ denote the concave envelope [46] of $I_x^{\log}$, *i.e.* the smallest concave function that upper bounds $I_x^{\log}$, also given by,

$$\hat{I}_x^{\log}(\zeta) = \sup_{\substack{\zeta_1,\zeta_2 \text{ s.t.}\\ \zeta_1 \leq \zeta \leq \zeta_2}} \left[\frac{(\zeta - \zeta_1) I_x^{\log}(\zeta_2) + (\zeta_2 - \zeta) I_x^{\log}(\zeta_1)}{\zeta_2 - \zeta_1} \right] \tag{28}$$

Clearly, $\hat{I}_x^{\log}$ exists and is upper bounded by ζ, which is a concave functions that upper bounds $I_x^{\log}(\zeta)$. Since $I_x^{\log}$ is real-analytic, $\hat{I}_x^{\log}(\zeta)$ is continuous and has a continuous derivative. Moreover, our previous results allow the following,

Proposition 3. *For every input distribution with finite alphabet, There exist $\zeta_1 > 0$ and $\zeta_2 < \infty$ such that $\hat{I}_x^{\log}(\zeta) = I_x^{\log}(\zeta)$ for every $\zeta \in [0, \zeta_1] \cup [\zeta_2, \infty)$.*

Proof: At any point ζ, the concave envelope $\hat{I}_x^{\log}(\zeta)$ is either equal to $I_x^{\log}(\zeta)$ or linear in an interval containing ζ, such that the concave envelope is equal to $I_x^{\log}$ at the edges of the intervals. Put in other words, there exists a set of disjoint intervals $\{[\zeta_{1,i}, \zeta_{2,i}]\}_{i \in S}$ such that

$$\hat{I}_x^{\log}(\zeta) = \begin{cases} \frac{\zeta - \zeta_{1,i}}{\zeta_{2,i} - \zeta_{1,i}} I_x^{\log}(\zeta_{2,i}) + \frac{\zeta_{2,i} - \zeta}{\zeta_{2,i} - \zeta_{1,i}} I_x^{\log}(\zeta_{1,i}) & \zeta \in [\zeta_{1,i}, \zeta_{2,i}] \\ I_x^{\log}(\zeta) & \text{otherwise} \end{cases} \tag{29}$$

Moreover, since $\hat{I}_x^{\log\prime}(\zeta)$ is also continuous, the above statement implies that $\hat{I}_x^{\log\prime}(\zeta) = I_x^{\log\prime}(\zeta_{1,i}) = I_x^{\log\prime}(\zeta_{2,i})$ for every $\zeta \in [\zeta_{1,i}, \zeta_{2,i}]$.

Suppose by contradiction that there exists i_0 such that $\zeta_{1,i_0} = 0$. Denoting $\gamma = e^\zeta - 1$, By the I-MMSE relationship we have $\hat{I}_x^{\log\prime}(\zeta) = r_x(\zeta) = (1 + \gamma) \operatorname{mmse}_x(\gamma)$. Since the input is normalized to unit power, we have $I_x^{\log\prime}(0) = r_x(0) = 1$, and so there must exist $\zeta_{2,i} > 0$ such that $I_x^{\log\prime}(\zeta_{2,i}) = I_x^{\log\prime}(\zeta_{1,i}) = 1$. However, since the input is not Gaussian (it has finite alphabet), and since $\operatorname{mmse}_x(0) = 1$, the single-crossing property [20] implies that $\operatorname{mmse}_x(\gamma) < 1/(1 + \gamma)$ for every $\gamma > 0$ and therefore $I_x^{\log\prime}(\zeta) < 1$ for every $\zeta > 0$, contradicting $I_x^{\log\prime}(\zeta_{2,i}) = 1$. Hence, setting $\zeta_1 = \min_{i \in S} \zeta_{i,1}$, we find that $I_x^{\log}(\zeta) = \hat{I}_x^{\log}(\zeta)$ for any $\zeta \in [0, \zeta_1]$.

Since the input has finite alphabet, $\hat{I}_x^{\log}(\zeta)$ converges to a finite value (the input entropy). By the data-processing inequality, mutual information is an increasing function of SNR, and since the mapping $\zeta = \log(1 + \gamma)$ is monotonic and increasing, it follows that $I_x^{\log}(\zeta)$ is also increasing in ζ. Therefore, $I_x^{\log\prime}(\zeta) > 0$ for any $0 \leq \zeta < \infty$ and $\lim_{\zeta \to \infty} I_x^{\log\prime}(\zeta) = 0$. Moreover, by Proposition 2 we know that there exist $\zeta_0 < \infty$ such that $I_x^{\log\prime}(\zeta)$ is monotonically decreasing for every $\zeta > \zeta_0$. By the above observations, there must exist $\zeta_0 \leq \zeta_2 < \infty$ such that $\zeta_a \leq \zeta_2 \leq \zeta_b$ implies $I_x^{\log\prime}(\zeta_a) > I_x^{\log\prime}(\zeta_2) \geq I_x^{\log\prime}(\zeta_b)$. Clearly $\zeta_{2,i} \leq \zeta_2$ for any $i \in S$, as otherwise the equality $I_x^{\log\prime}(\zeta_{2,i}) = I_x^{\log\prime}(\zeta_{1,i})$ contradicts $I_x^{\log\prime}(\zeta_{1,i}) > I_x^{\log\prime}(\zeta_2) \geq I_x^{\log\prime}(\zeta_{2,i})$. Therefore, we have $I_x^{\log}(\zeta) = \hat{I}_x^{\log}(\zeta)$ for any $\zeta \in [\zeta_2, \infty)$, concluding our proof. ∎

Definition 2. Let $0 < \underline{\zeta}_1 \leq \infty$ be the *maximal* ζ_1 for which $\hat{I}_x^{\log}(\zeta) = I_x^{\log}(\zeta)$ for every $\zeta \in [0, \zeta_1]$ and similarly let $0 \leq \bar{\zeta}_2 < \infty$ be the *minimal* ζ_2 for which $\hat{I}_x^{\log}(\zeta) = I_x^{\log}(\zeta)$ for every $\zeta \in [\zeta_2, \infty)$. Let $\underline{\gamma}_1 = e^{\underline{\zeta}_1} - 1$ and $\bar{\gamma}_2 = e^{\bar{\zeta}_2} - 1$ denote the SNR's corresponding to $\underline{\zeta}_1$ and $\bar{\zeta}_2$, respectively.
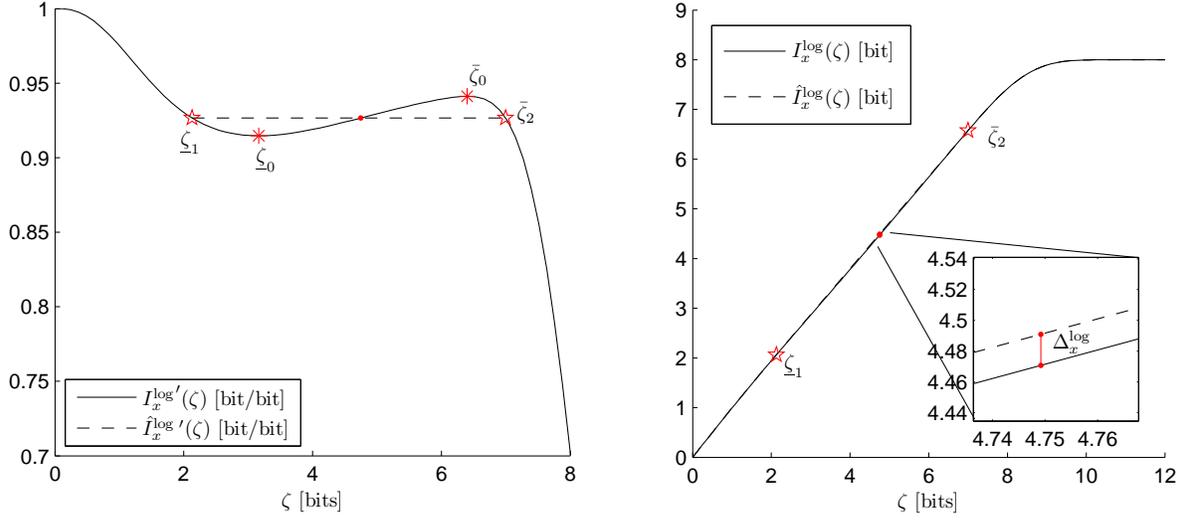
Figure 3. $I_x^{\log}$ and $\hat{I}_x^{\log}$ (right), and their derivatives with respect to ζ (left), with $\underline{\zeta}_0, \bar{\zeta}_0, \underline{\zeta}_1, \bar{\zeta}_2$ and Δ_x highlighted, for 256-QAM input.

Notice that in light of the above definition, the optimization in the definition of the concave envelope (28) can be limited to ζ_1 and ζ_2 in the interval $[\underline{\zeta}_1, \bar{\zeta}_2]$.

Definition 3. Let Δ_x denote that maximum difference between $I_x^{\log}(\zeta)$ and its concave envelope.

By the above definition and (28) we have,

$$
\Delta_x = \max_\zeta \left[\hat{I}_x^{\log}(\zeta) - I_x^{\log}(\zeta) \right] \tag{30}
$$

$$
= \sup_{\substack{\gamma_1, \gamma_2, \gamma \text{ s.t.} \\ \gamma_1 \le \gamma \le \gamma_2}} \frac{\log\left(\frac{1+\gamma}{1+\gamma_1}\right) \left[I_x(\gamma_2) - I_x(\gamma) \right] - \log\left(\frac{1+\gamma_2}{1+\gamma}\right) \left[I_x(\gamma) - I_x(\gamma_1) \right]}{\log\left([1+\gamma_2]/[1+\gamma_1]\right)} \tag{31}
$$

Using (29), it is seen that

$$
\Delta_x = \hat{I}_x^{\log}(\zeta_m) - I_x^{\log}(\zeta_m) = (\zeta_m - \zeta_{1,i}) I_x^{\log\prime}(\zeta_m) - \left[I_x^{\log}(\zeta_m) - I_x^{\log}(\zeta_{1,i}) \right]
$$
$$
= \left[I_x^{\log}(\zeta_{2,i}) - I_x^{\log}(\zeta_m) \right] - (\zeta_{2,i} - \zeta_m) I_x^{\log\prime}(\zeta_m) \tag{32}
$$

for some $i \in S$ and $\zeta_m \in [\zeta_{1,i}, \zeta_{2,i}]$ that satisfies $I_x^{\log\prime}(\zeta_m) = I_x^{\log\prime}(\zeta_{1,i}) = I_x^{\log\prime}(\zeta_{2,i})$.

Clearly, $\Delta_x = 0$ if and only if $I_x^{\log}$ is concave in $\mathbb{R}^+$. As seen from Table I, Δ_x is quite small, even when $I_x^{\log}$ is not concave. Figure 3 illustrates $\hat{I}_x^{\log}, \underline{\zeta}_0, \bar{\zeta}_0, \underline{\zeta}_1, \bar{\zeta}_2$ and Δ_x as defined above, for an input uniformly distributed on a 256-QAM constellation (two 16-PAM constellations in quadrature).

D. Proof of Theorem 1

Proof: The inequality (15) is immediate from the definitions of I_{SL}, $\mathcal{I}_{OFDM}$, $I_x^{\log}$, $\hat{I}_x^{\log}$ and Δ_x:

$$\mathcal{I}_{OFDM} = \frac{1}{2\pi} \int_{-\pi}^{\pi} I_x \left(|H(\theta)|^2 \right) d\theta \tag{33}$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} I_x^{\log} \left(\log \left(1 + |H(\theta)|^2 \right) \right) d\theta \tag{34}$$

$$\leq \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{I}_x^{\log} \left(\log \left(1 + |H(\theta)|^2 \right) \right) d\theta \tag{35}$$

$$\leq \hat{I}_x^{\log} \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \log \left(1 + |H(\theta)|^2 \right) d\theta \right) \tag{36}$$

$$= \hat{I}_x^{\log} \left(\log \left(1 + \mathsf{SNR}_{\text{MMSE-DFE-U}} \right) \right) \tag{37}$$

$$\leq I_x \left(\mathsf{SNR}_{\text{MMSE-DFE-U}} \right) + \Delta_x = I_{SL} + \Delta_x \tag{38}$$

where in (36) the concavity of $\hat{I}_x^{\log}$ was used to invoke Jensen's inequality. Choosing $\underline{\gamma}_1$ and $\bar{\gamma}_2$ to be as defined in Definition 2, it is clear that if condition 1 holds, then

$$\hat{I}_x^{\log} \left(\log \left(1 + \mathsf{SNR}_{\text{MMSE-DFE-U}} \right) \right) = I_x^{\log} \left(\log \left(1 + \mathsf{SNR}_{\text{MMSE-DFE-U}} \right) \right) = I_{SL} \tag{39}$$

and $\mathcal{I}_{OFDM} \leq I_{SL}$ is therefore valid. Choosing $\underline{\gamma}_0$ and $\bar{\gamma}_0$ according to Definition 1, if either condition 2 or conditions 3 hold, then $I_x^{\log}$ is a concave function for every value of $|H(\theta)|^2$, and we may therefore exchange $\hat{I}_x^{\log}$ with $I_x^{\log}$ in (36), yielding $\mathcal{I}_{OFDM} \leq I_{SL}$ once more. ∎

IV. INTERLUDE — MMSE BOUNDS FOR PAM SIGNALING

In order to prove Theorems 2 and 3, we first need to derive a tight high-SNR characterization of the MMSE function and its derivative, for PAM inputs and Gaussian noise. In this section, we first present a novel "pointwise" bound on the MMSE of PAM signaling conditioned on the channel output. The bound is then applied to derive inequalities that characterize the MMSE and its derivative in the high-SNR regime, as required. Finally, some bounds on the MMSE and its derivative are presented for the special case of BPSK input. Besides their use in proving Theorems 2 and 3, the results presented in this section may be of general interest.

A. A pointwise MMSE inequality

Let X be a real-valued random variable and define[5] $Y_\gamma = X + \frac{1}{\sqrt{\gamma}} N$ with $N \sim \mathcal{N}(0, \frac{1}{2})$ and independent of X. Y_γ is the Gaussian-noise contaminated version of X, at SNR γ. Let

$$\phi_X(y; \gamma) = \mathrm{E}_X \left[|X - \mathrm{E}[X|Y_\gamma = y]|^2 \mid Y_\gamma = y \right] \tag{40}$$

---

[5] We have chosen here to let γ scale N and not X, as opposed to the convention in the I-MMSE literature. This definition ensures that Y_γ and X are on the same scale, which simplifies many of the derivations that follow. We have also set the noise variance to be $1/2$ when $\gamma = 1$, in the purpose of making these results easily applicable in the complex setting of the rest of the paper.

denote the "point-wise" conditional variance of X given a noisy channel observation. Clearly, $\mathrm{mmse}_X(\gamma) = \mathrm{E}_{Y_\gamma} \phi_X(Y_\gamma; \gamma)$. Moreover, exploration of Information-Estimation relations [20] has revealed that,

$$\mathrm{mmse}'_X(\gamma) = -2\mathrm{E}_{Y_\gamma} \phi_X(Y_\gamma; \gamma) \tag{41}$$

Hence, intimate understanding of $\phi_X(y; \gamma)$ is expected to provide insights on both the MMSE function and its derivative. For the case of PAM input distribution, this understanding is presented in the form of the following,

Theorem 5. *Let $\mathcal{X} = \{x_1, ..., x_M\}$ be the alphabet of an M-ary PAM constellation such that $x_{m+1} - x_m = d$ for every $1 \leq m < M$. Let X be uniformly distributed in $\mathcal{X}$. Fix $y \in \mathbb{R}$ and choose $1 \leq J < M$ such that x_J, x_{J+1} are the nearest values to y in $\mathcal{X}$. Let B_J be uniformly distributed on $\{x_J, x_{J+1}\}$. Then,*

$$0 \leq \phi_X(y; \gamma) - \phi_{B_J}(y; \gamma) \leq \left(\frac{d}{2}\right)^2 \bar{D}\left(\left(\frac{d}{2}\right)^2 \gamma\right) \tag{42}$$

with

$$\bar{D}(\gamma) = 4\sum_{k=1}^{\infty} (k+1)^2 e^{-4\gamma k^2} \leq \frac{16 e^{-4\gamma}}{(1 - e^{-4\gamma})^3} \tag{43}$$

Proof: Appendix B. ∎

Note that

$$\phi_{B_J}(y; \gamma) = \left(\frac{d}{2}\right)^2 \phi_{\mathrm{BPSK}}\left(\left(\frac{d}{2}\right)^{-1}\left[y - \frac{x_J + x_{J+1}}{2}\right]; \left(\frac{d}{2}\right)^2 \gamma\right) \tag{44}$$

where $\phi_{\mathrm{BPSK}}(y; \gamma) = 1 - \tanh^2(2\gamma y)$ is the conditional variance function for BPSK input.

Put in words, Theorem 5 means that for PAM input and given channel output, the expected value of the MMSE is lower-bounded by the expected MMSE given the same channel output and assuming an input equally distributed on the two PAM symbols nearest to it.

Figure 4 illustrates the bounds in (42) for 4-PAM input. As the figure indicates, the lower bound is reasonably tight for $(d/2)^2 \gamma$ as low as 0 dB, and both bounds are very tight for $(d/2)^2 \gamma = 3$ dB and above.

B. High-SNR characterization of the MMSE

Let $\mathrm{mmse}_{d,M\text{-PAM}}(\gamma)$ stand for the $\mathrm{mmse}_X(\gamma)$, with X uniformly distributed on an M-PAM constellation with distance d between adjacent points. We show that

$$\mathrm{mmse}_{d,M\text{-PAM}}(\gamma) \approx 2\frac{M-1}{M}\left(\frac{d}{2}\right)^2 \mathrm{mmse}_{\mathrm{BPSK}}\left(\left(\frac{d}{2}\right)^2 \gamma\right) \tag{45}$$

in the sense that the difference between the terms tends to zero with a faster exponential rate than $\mathrm{mmse}_{M-PAM}(\gamma)$, where $\mathrm{mmse}_{\mathrm{BPSK}}(\gamma) \equiv \mathrm{mmse}_{2,2-PAM}(\gamma)$. Moreover, we use similar techniques in order to show that this
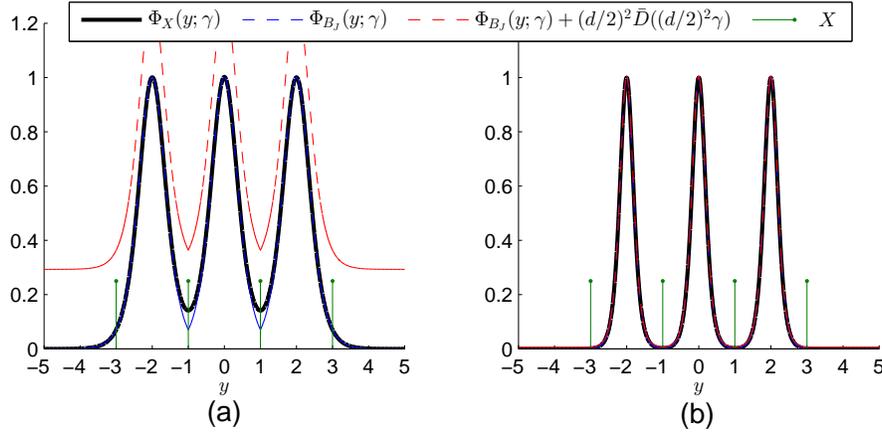
Figure 4. Illustration of Theorem 5 for 4-PAM input with $d = 2$ and two different SNR's: (a) $(d/2)^2 \gamma = 1$ and (b) $(d/2)^2 \gamma = 2$.

approximation also applies to the derivative of the MMSE with respect to γ, *i.e.,*

$$\text{mmse}'_{d,M\text{-PAM}}(\gamma) \approx 2\frac{M-1}{M}\left(\frac{d}{2}\right)^4 \text{mmse}'_{\text{BPSK}}\left(\left(\frac{d}{2}\right)^2 \gamma\right) \tag{46}$$

It should be noted that (45) can be seen as a special case of the high-SNR MMSE characterization for general discrete inputs that was recently presented in [47]. However, the following analysis provides two important advantages. First, it enables us to also establish (46) — a characterization of the derivative of the MMSE. Second, it allows for explicit and tight bounds on the difference between the exact M-PAM quantities and their BPSK approximations. Both of these features are crucial for establishing the bound in Theorem 3.

Letting

$$Q(x) \triangleq \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \tag{47}$$

denote the standard error function, the result (45) is stated formally as follows,

Theorem 6. *The following bounds hold for every $M \geq 2$, $d \geq 0$ and $\gamma \geq 0$:*

$$\text{mmse}_{d,M\text{-PAM}}(\gamma) \leq 2\frac{M-1}{M}\left(\frac{d}{2}\right)^2 \left[\text{mmse}_{\text{BPSK}}\left(\left(\frac{d}{2}\right)^2 \gamma\right) + \bar{B}\left(\left(\frac{d}{2}\right)^2 \gamma\right)\right] \tag{48}$$

with

$$\bar{B}(\gamma) = 16Q\left(\sqrt{8\gamma}\right) + 4\sum_{k=2}^\infty (2k+1) Q\left(k\sqrt{8\gamma}\right) \tag{49}$$

and

$$\text{mmse}_{d,M\text{-PAM}}(\gamma) \geq 2\frac{M-1}{M}\left(\frac{d}{2}\right)^2 \left[\text{mmse}_{\text{BPSK}}\left(\left(\frac{d}{2}\right)^2 \gamma\right) - \underline{B}\left(\left(\frac{d}{2}\right)^2 \gamma\right)\right] \tag{50}$$

with

$$\underline{B}(\gamma) = 4Q\left(\sqrt{8\gamma}\right) \leq \frac{1}{\sqrt{\pi\gamma}}e^{-4\gamma} \tag{51}$$

Proof: Appendices C-A and C-D. ∎

Similarly, (46) has the following formal form,

Theorem 7. *The following bounds hold for every $M \geq 2$, $d \geq 0$ and $\gamma \geq 0$:*

$$\mathrm{mmse}'_{d,M\text{-PAM}}(\gamma) \leq 2\frac{M-1}{M}\left(\frac{d}{2}\right)^4\left[\mathrm{mmse}'_{\mathrm{BPSK}}\left(\left(\frac{d}{2}\right)^2\gamma\right) + \bar{C}\left(\left(\frac{d}{2}\right)^2\gamma\right)\right] \tag{52}$$

with

$$\bar{C}(\gamma) = 32e^{8\gamma}Q\left(\sqrt{32\gamma}\right) \leq \frac{4}{\sqrt{\pi\gamma}}e^{-8\gamma} \tag{53}$$

and

$$\mathrm{mmse}'_{d,M\text{-PAM}}(\gamma) \geq 2\frac{M-1}{M}\left(\frac{d}{2}\right)^4\left[\mathrm{mmse}'_{\mathrm{BPSK}}\left(\left(\frac{d}{2}\right)^2\gamma\right) - \underline{C}\left(\left(\frac{d}{2}\right)^2\gamma\right)\right] \tag{54}$$

with

$$\underline{C}(\gamma) = 2\left[4\left(\sum_{k=1}^{\infty}(k+1)^2 e^{-4\gamma k^2}\right)\left(8\sum_{k=1}^{\infty}(k+1)^2 e^{-4\gamma k^2} + 1\right) + Q\left(\sqrt{8\gamma}\right)\right] \tag{55}$$

Proof: Appendices C-C and C-B. ∎

Figure 5 illustrates the high SNR behavior of $\mathrm{mmse}_{d,M\text{-PAM}}(\gamma)\big/\left[2\frac{M-1}{M}\left(\frac{d}{2}\right)^2\right]$ and $\mathrm{mmse}'_{d,M\text{-PAM}}(\gamma)\big/\left[2\frac{M-1}{M}\left(\frac{d}{2}\right)^4\right]$ for different values of M. It is seen that the above bounds become tight at $(d/2)^2\gamma$ values of around 3 dB.

C. Bounds for BPSK inputs

We present some upper and lower bound on the MMSE function and its derivative for the case of BPSK inputs. These bounds will be of use in proving analytically that $\Delta_x = 0$ for BPSK and QPSK inputs, as claimed in Theorem 2.

Theorem 8. *The following bounds on* $\mathrm{mmse}_{\mathrm{BPSK}}(\gamma)$ *hold*

$$\left(1 - \frac{1}{2\gamma}\frac{\pi^2}{8}\right)\frac{\sqrt{\pi}}{2}\frac{1}{\sqrt{\gamma}}e^{-\gamma} \leq \mathrm{mmse}_{\mathrm{BPSK}}(\gamma) \leq \frac{\sqrt{\pi}}{2}\frac{1}{\sqrt{\gamma}}e^{-\gamma} \tag{56}$$

$$\frac{e^{-\gamma}}{\sqrt{1+2\gamma}} \leq \mathrm{mmse}_{\mathrm{BPSK}}(\gamma) \leq e^{-\gamma} \tag{57}$$
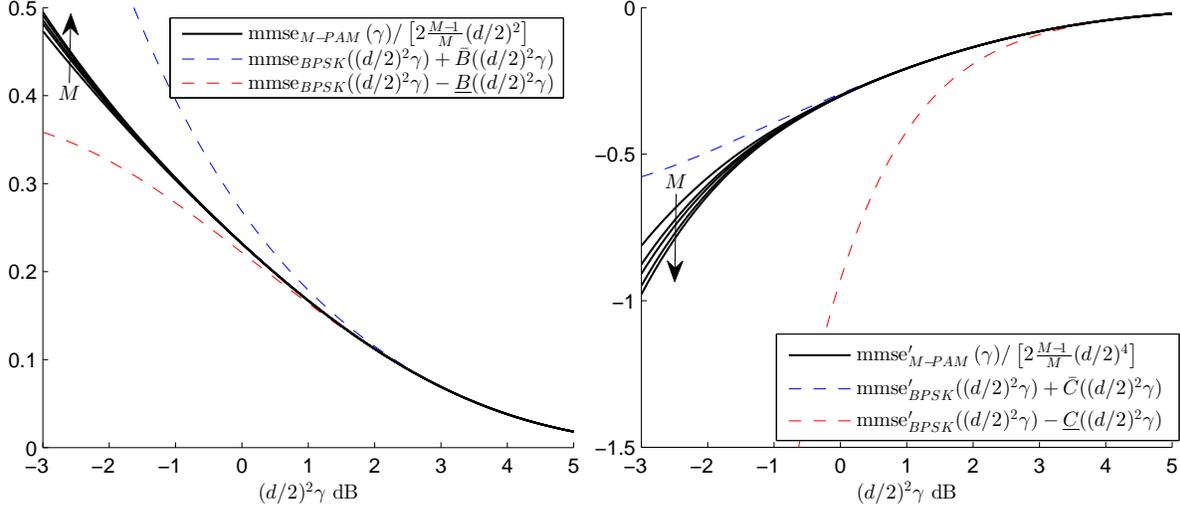
Figure 5. Illustration of Theorems 6 and 7 for M values of 3, 4, 5, 8 and 16.

similarly, the following bounds on $\mathrm{mmse}'_{\mathrm{BPSK}}(\gamma)$ *hold*

$$\left(1 - \frac{1}{2\gamma}\left(\frac{\pi^2}{8} - 1\right)\right)\frac{\sqrt{\pi}}{2}\frac{1}{\sqrt{\gamma}}e^{-\gamma} \leq -\mathrm{mmse}'_{\mathrm{BPSK}}(\gamma) \leq \frac{\sqrt{\pi}}{2}\frac{1}{\sqrt{\gamma}}e^{-\gamma} \tag{58}$$

$$\frac{2e^{-\gamma}}{\sqrt{1+6\gamma}} \leq -\mathrm{mmse}'_{\mathrm{BPSK}}(\gamma) \leq 2e^{-\gamma} \tag{59}$$

Proof: Appendix D. ∎

We note that the bounds in (56) are composed of the first two terms in the asymptotic high SNR expansion of $\mathrm{mmse}_{\mathrm{BPSK}}(\gamma)$ derived in [44]. Our contribution in this case is the proof that these approximations are upper and lower bounds. Our proof of the bounds also allows for a simpler derivation of the series expansion than the one in [44], as well as extension of these bounds to $\mathrm{mmse}'_{\mathrm{BPSK}}(\gamma)$. It is worth noting that while not asymptotically tight as $\gamma \to \infty$, the lower bound $e^{-\gamma}/\sqrt{1+2\gamma}$ is a good approximation for $\mathrm{mmse}_{\mathrm{BPSK}}(\gamma)$ for all values of γ, with a maximum slackness of less than 0.022.

V. RESULTS FOR PAM AND SQUARE QAM INPUTS

In this section we utilize the results from Sections III and IV in order to prove Theorems 3, 2 and 4.

A. Proof of Theorem 2

We recall the definition of $I_x^{\log}$ (21) and prove the following,

Lemma 1. $I_{\mathrm{BPSK}}^{\log}(\zeta)$ *and* $I_{\mathrm{QPSK}}^{\log}(\zeta)$ *are concave in* ζ.

Proof: Let $\gamma = e^\zeta - 1$. As shown in (22), $I_x^{\log ''}(\zeta) = (1 + \gamma) \left[\mathrm{mmse}_x(\gamma) + (1 + \gamma) \, \mathrm{mmse}'_x(\gamma) \right]$. By Theorem 8,

$$\mathrm{mmse}_{\mathrm{BPSK}}(\gamma) + (1 + \gamma) \, \mathrm{mmse}'_{\mathrm{BPSK}}(\gamma) \leq e^{-\gamma} \left[1 - \frac{2(1 + \gamma)}{\sqrt{1 + 6\gamma}} \right] < 0 \tag{60}$$

and so $I_{\mathrm{BPSK}}^{\log ''}(\zeta) < 0$ for every ζ, meaning $I_{\mathrm{BPSK}}^{\log}$ is concave in ζ. Since,

$$\mathrm{mmse}_{\mathrm{QPSK}}(\gamma) = \mathrm{mmse}_{\mathrm{BPSK}} \left(\frac{\gamma}{2} \right) \tag{61}$$

we have,

$$\mathrm{mmse}_{\mathrm{QPSK}}(\gamma) + (1 + \gamma) \, \mathrm{mmse}'_{\mathrm{QPSK}}(\gamma) \leq e^{-\gamma/2} \left[1 - \frac{1 + \gamma}{\sqrt{1 + 3\gamma}} \right] \tag{62}$$

and it is easily verified that $(1 + \gamma) / \sqrt{1 + 3\gamma} \geq 1$ for every $\gamma \geq 1$. For lower SNR's, we use the Gaussian upper bound $\mathrm{mmse}_{\mathrm{QPSK}}(\gamma) \leq (1 + \gamma)^{-1}$ as well as $e^{-\gamma/2} \geq 1 - \gamma/2$ to show that

$$\mathrm{mmse}_{\mathrm{QPSK}}(\gamma) + (1 + \gamma) \, \mathrm{mmse}'_{\mathrm{QPSK}}(\gamma) \leq \frac{1}{1 + \gamma} \left[1 - \frac{(1 + \gamma)^2 (1 - \gamma/2)}{\sqrt{1 + 3\gamma}} \right] \tag{63}$$

and once more it is easily verified that $(1 + \gamma)^2 (1 - \gamma/2) / \sqrt{1 + 3\gamma} \geq 1$ for every $\gamma \leq 1$. We thus conclude that $I_{\mathrm{QPSK}}^{\log ''}(\zeta) < 0$ for every ζ, and so $I_{\mathrm{QPSK}}^{\log}$ is concave in ζ. ∎

Clearly, if $I_x^{\log}$ is concave then $I_x^{\log}(\zeta) = \hat{I}_x^{\log}(\zeta)$ for every ζ and by its definition (31), $\Delta_x = 0$. Thus, Lemma 1 proves Theorem 2. While numeric investigation shows that $I_{\text{4-PAM}}^{\log}$, $I_{\text{8-PSK}}^{\log}$, $I_{\text{16-QAM}}^{\log}$ and $I_{\text{32-QAM}}^{\log}$ are also concave, no analytical proof of concavity has been found for these cases.

B. Proof of Theorem 3

Proof: Let $\gamma = e^\zeta - 1$ and let $\rho = \left(d_{\min}^{\mathrm{PAM}}/2 \right)^2 \gamma$ where $d_{\min}^{\mathrm{PAM}} = \sqrt{\frac{12}{M^2 - 1}}$ is the minimum distance between symbols of a unit-power M-PAM input. Using (22) and bounds provided in Theorems 6 and 7, we find that

$$\begin{aligned} I_{M\text{-PAM}}^{\log ''}(\zeta) &= (1 + \gamma) \left[\mathrm{mmse}_{M\text{-PAM}}(\gamma) + (1 + \gamma) \, \mathrm{mmse}'_{M\text{-PAM}}(\gamma) \right] \\ &\leq K \left[\rho \, \mathrm{mmse}'_{\mathrm{BPSK}}(\rho) + \mathrm{mmse}_{\mathrm{BPSK}}(\rho) + \bar{B}(\rho) + \rho \bar{C}(\rho) \right] \end{aligned} \tag{64}$$

with $\bar{B}(\rho)$ and $\bar{C}(\rho)$ given in (49) and (53), respectively, and $K = 2(1 + \gamma) \frac{M - 1}{M} \left(d_{\min}^{\mathrm{PAM}}/2 \right)^2$. Evaluating numerically the expression in square brackets, it is found that $I_{M\text{-PAM}}^{\log ''}(\zeta) < 0$ for $\rho \geq 1$, see Figure 6. Consequently, $\left(d_{\min}^{\mathrm{PAM}}/2 \right)^2 \bar{\gamma}_0 \leq 1$ for M-PAM inputs, as required.

An M^2-ary square QAM constellation is composed of two M-ary PAM constellations in quadrature. Letting $d_{\min}^{\mathrm{QAM}} = \sqrt{\frac{6}{M^2 - 1}}$ denote the minimum distance between symbols of a unit-power M^2-QAM input, we have

$$\mathrm{mmse}_{M^2\text{-QAM}}(\gamma) = 2 \mathrm{mmse}_{d_{\min}^{\mathrm{QAM}}, M\text{-PAM}}(\gamma) \tag{65}$$

and

$$\mathrm{mmse}'_{M^2\text{-QAM}}(\gamma) = 2 \mathrm{mmse}'_{d_{\min}^{\mathrm{QAM}}, M\text{-PAM}}(\gamma) \tag{66}$$
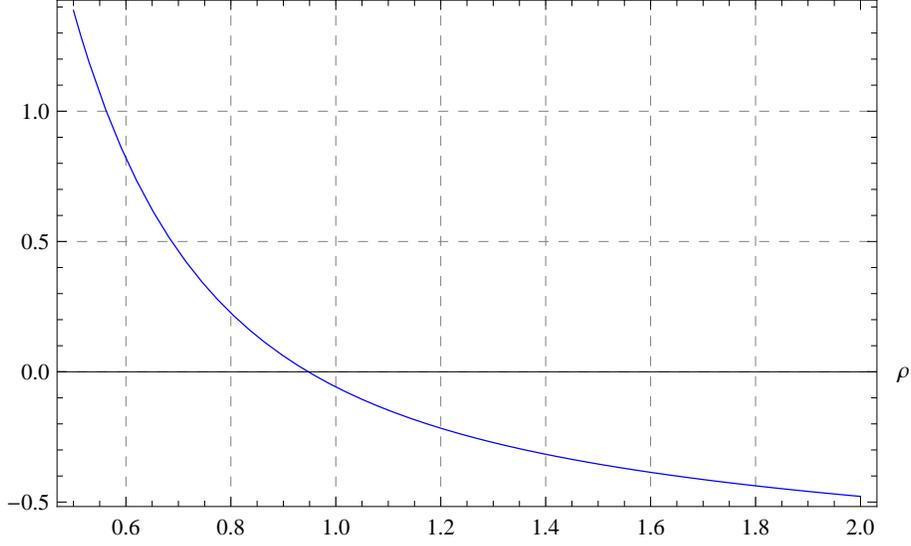
Figure 6. Evaluation of $\left(\sqrt{\rho}e^{-\rho}\right)^{-1}\left[\rho\,\mathrm{mmse}'_{\mathrm{BPSK}}\left(\rho\right) + \mathrm{mmse}_{\mathrm{BPSK}}\left(\rho\right) + \bar{B}\left(\rho\right) + \rho\bar{C}\left(\rho\right)\right]$.

where $\mathrm{mmse}_{d,M\text{-PAM}}\left(\gamma\right)$ denotes the MMSE for an M-PAM input with distance d between adjacent symbols and complex-valued additive Gaussian noise with power $1/\gamma$, as defined in subsection IV-B. Applying Theorems 6 and 7 to the above equations, we find that $I^{\log}_{M^2\text{-QAM}}{}''\left(\zeta\right)$ is also bounded from above by (64), with $\rho = \left(d^{\mathrm{QAM}}_{\min}/2\right)^2\gamma$. Therefore, we have $\left(d^{\mathrm{QAM}}_{\min}/2\right)^2\bar{\gamma}_0 \leq 1$ for M^2-QAM inputs, and the proof is complete. ∎

Examining Figure 6 more closely, we find that the term (64) becomes negative for ρ values around 0.95, and the bound on $\left(d_{\min}/2\right)^2\bar{\gamma}_0$ might be slightly tightened accordingly. However, as remarked on Table I, numerical evaluation of $\bar{\gamma}_0$ for square QAM inputs indicate that $\left(d_{\min}/2\right)^2\bar{\gamma}_0$ is closer to 0.5. Therefore, reducing the bound by 0.05 does not significantly improve its tightness.

C. Proof of Corollary 1

Proof: In the large block size limit, the input SNR at the k'th OFDM subcarrier is given by $\gamma_k = |H\left(\theta_k\right)|^2$ where $\theta_k = 2\pi k/N$ is the subcarrier frequency and k is its index spanning from 0 to $N-1$. Consider a unit-power square M^2-QAM input and Gaussian noise at SNR γ, and set $q = Q\left(\sqrt{\left(d_{\min}/2\right)^2\gamma}\right)$, with the error function $Q\left(\cdot\right)$ as given in (47). For M-PAM input with spacing $d_{\min}$, the probability of a symbol error is $2q$ for the $M-2$ inner constellation points, and q for the 2 outer points, *i.e.*

$$P^{M\text{-PAM}}_{err} = \frac{M-2}{M}2q + \frac{2}{M}q = 2\frac{M-1}{M}q \tag{67}$$

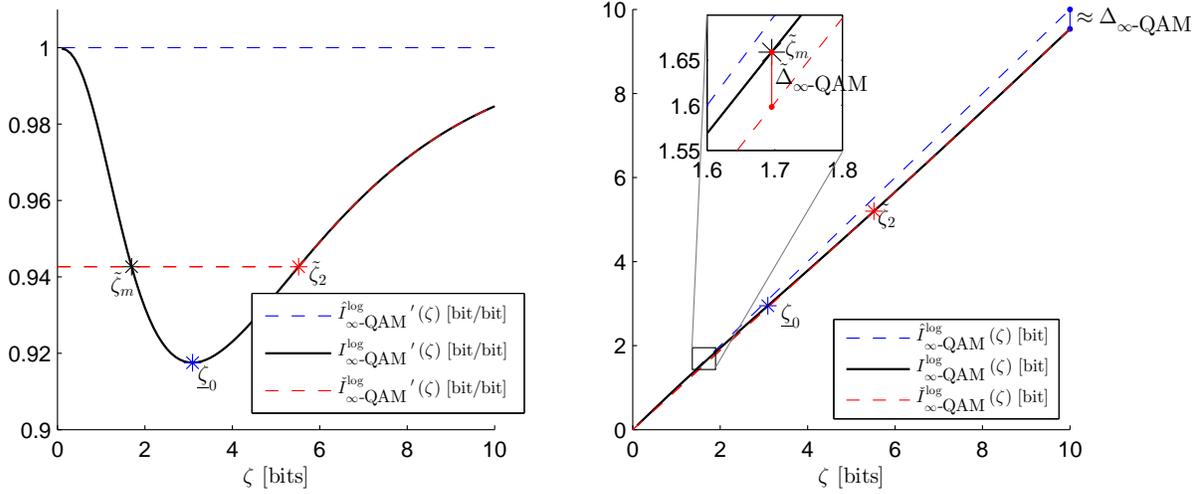For M^2-QAM input, a symbol error event is the union of two independent error events along the in-phase and

Figure 7. $I_x^{\log}$ with its concave and convex envelopes (right), and their derivatives with respect to ζ (left), with $\underline{\zeta}_0, \tilde{\zeta}_2, \Delta_x$ and $\tilde{\Delta}_x$ highlighted, for ∞-QAM input.

quadrature directions, each being M-ary PAM error events. Therefore,

$$P_{err}^{M^2\text{-QAM}} = 2P_{err}^{M\text{-PAM}} - \left(P_{err}^{M\text{-PAM}}\right)^2 = 4\frac{M-1}{M}q\left(1 - \frac{M-1}{M}q\right) \tag{68}$$

It can thus be seen that for $M \geq 16$, $P_{err}^{M^2\text{-QAM}} < 50\%$ implies $(d_{\min}/2)^2 \gamma > 1$ and hence $\gamma > \bar{\gamma}_0$. Assuming this holds for all subcarriers and assuming large enough OFDM block size, we find that $|H(\theta)|^2 > \bar{\gamma}_0$ for all values of θ, thus satisfying condition 3 of Theorem 1 and proving the corollary. ∎

D. Analysis of uniform input

We consider a unit-power input distributed uniformly on the square $\left[-\sqrt{\frac{3}{2}}, \sqrt{\frac{3}{2}}\right] \times \left[-\sqrt{\frac{3}{2}}, \sqrt{\frac{3}{2}}\right]$, also referred to as ∞-QAM input. As usual we let $I_{\infty\text{-QAM}}(\gamma)$ denote the mutual information between such input and its complex Gaussian noise corrupted version at SNR γ, and we let $I_{\infty\text{-QAM}}^{\log}(\zeta) = I_{\infty\text{-QAM}}(e^\zeta - 1)$ be the mutual information with respect to log SNR. Figure 7 illustrates $I_{\infty\text{-QAM}}^{\log}$ and its derivative, as well as the quantities to be defined in the following paragraphs.

For high SNR, it well known [48] that

$$I_{\infty\text{-QAM}}(\gamma) \approx \log\left(\frac{6}{\pi e}\gamma\right) \tag{69}$$

and therefore

$$\lim_{\zeta \to \infty} \left(\zeta - I_{\infty\text{-QAM}}^{\log}(\zeta)\right) = \log\left(\frac{\pi e}{6}\right) \approx 0.509 \text{ [bit]} \tag{70}$$

as ζ is also the mutual information for Gaussian input at log-SNR ζ, the above limit represent the loss of using

uniform input rather than Gaussian input at high SNR, and is commonly referred to as the shaping gain.

Proposition 1 applies to the case of uniform input, and so we know there must a constant $\underline{\zeta}_0$ such $I^{\log}_{\infty\text{-QAM}}(\zeta)$ is concave for every $\zeta \leq \underline{\zeta}_0$. However, the high-SNR behavior of $I^{\log}_{\infty\text{-QAM}}(\zeta)$ is quite different than the finite-alphabet case, and is characterized as follows,

Proposition 4. $I^{\log}_{\infty\text{-QAM}}(\zeta)$ *is concave for every $\zeta \leq \zeta_0$ and convex for every $\zeta \geq \underline{\zeta}_0$, where $\underline{\zeta}_0 \approx 3.09$ bits.*

Proof: Appendix E. ∎

Even though it becomes convex in high SNR's, $I^{\log}_{\infty\text{-QAM}}$ still has a concave envelope. Additionally, it is of interest to study the *convex envelope* of $I^{\log}_{\infty\text{-QAM}}$, *i.e.* the maximum convex function that lower bounds $I^{\log}_{\infty\text{-QAM}}$, which we denote by $\check{I}^{\log}_{\infty\text{-QAM}}$. Moreover, we are interested in the concave envelope of $I^{\log}_{\infty\text{-QAM}}$ when limited to the interval $[0, \bar{\zeta}]$, *i.e.* the minimum concave function that upper bounds $I^{\log}_{\infty\text{-QAM}}$ for every $\zeta \in [0, \bar{\zeta}]$. We will denote this function by $\hat{I}^{\log;[0,\bar{\zeta}]}_{\infty\text{-QAM}}(\zeta)$. These concave and convex envelopes are characterized as follows,

Proposition 5. *The concave envelope of $I^{\log}_{\infty\text{-QAM}}$ is given by*

$$\hat{I}^{\log}_{\infty\text{-QAM}}(\zeta) = \lim_{\bar{\zeta}\to\infty} \hat{I}^{\log;[0,\bar{\zeta}]}_{\infty\text{-QAM}}(\zeta) = \zeta \tag{71}$$

and satisfies

$$\Delta_{\infty\text{-QAM}} = \sup_\zeta \left(\hat{I}^{\log}_{\infty\text{-QAM}}(\zeta) - I^{\log}_{\infty\text{-QAM}}(\zeta) \right) = \log\left(\frac{\pi e}{6}\right) \tag{72}$$

Limited to the interval $[0, \bar{\zeta}]$, where $\bar{\zeta} > \underline{\zeta}_0$, the concave envelope of $I^{\log}_{\infty\text{-QAM}}$ is given by

$$\hat{I}^{\log;[0,\bar{\zeta}]}_{\infty\text{-QAM}}(\zeta) = \begin{cases} I^{\log}_{\infty\text{-QAM}}(\underline{\zeta}_1) + (\zeta - \underline{\zeta}_1)I^{\log}_{\infty\text{-QAM}}{}'(\underline{\zeta}_1) & \zeta \geq \underline{\zeta}_1 \\ I^{\log}_{\infty\text{-QAM}}(\zeta) & \text{otherwise} \end{cases} \tag{73}$$

where $\underline{\zeta}_1 < \underline{\zeta}_0$ depends on $\bar{\zeta}$ and is determined by the condition $\hat{I}^{\log;[0,\bar{\zeta}]}_{\infty\text{-QAM}}(\bar{\zeta}) = I^{\log}_{\infty\text{-QAM}}(\bar{\zeta})$. The function $\bar{\Delta}_{\infty\text{-QAM}}(\bar{\gamma})$ is given by,

$$\bar{\Delta}_{\infty\text{-QAM}}\left(e^{\bar{\zeta}} - 1\right) \triangleq \sup_{\zeta \leq \bar{\zeta}} \left(\hat{I}^{\log;[0,\bar{\zeta}]}_{\infty\text{-QAM}}(\zeta) - I^{\log}_{\infty\text{-QAM}}(\zeta) \right) = \hat{I}^{\log;[0,\bar{\zeta}]}_{\infty\text{-QAM}}(\zeta_m) - I^{\log}_{\infty\text{-QAM}}(\zeta_m) \tag{74}$$

where $\zeta_m > \underline{\zeta}_1$ depends on $\bar{\zeta}$ and satisfies $I^{\log}_{\infty\text{-QAM}}{}'(\zeta_m) = I^{\log}_{\infty\text{-QAM}}{}'(\underline{\zeta}_1)$. For $\bar{\zeta} \leq \underline{\zeta}_0$, $I^{\log}_{\infty\text{-QAM}}$ is concave on the interval $[0, \bar{\zeta}]$, so that $\hat{I}^{\log;[0,\bar{\zeta}]}_{\infty\text{-QAM}}(\zeta) = I^{\log}_{\infty\text{-QAM}}(\zeta)$ and $\bar{\Delta}_{\infty\text{-QAM}}\left(e^{\bar{\zeta}} - 1\right) = 0$.

The convex envelope of $I^{\log}_{\infty\text{-QAM}}$ is given by

$$\check{I}^{\log}_{\infty\text{-QAM}}(\zeta) = \begin{cases} \zeta I^{\log}_{\infty\text{-QAM}}{}'(\tilde{\zeta}_2) & \zeta \leq \tilde{\zeta}_2 \\ I^{\log}_{\infty\text{-QAM}}(\zeta) & \text{otherwise} \end{cases} \tag{75}$$

with $\tilde{\zeta}_2 \approx 5.52$ [bits] determined by the continuity condition $\check{I}^{\log}_{\infty\text{-QAM}}(\tilde{\zeta}_2) = I^{\log}_{\infty\text{-QAM}}(\tilde{\zeta}_2)$. The constant $\tilde{\Delta}_{\infty\text{-QAM}}$ is

given by

$$\tilde{\Delta}_{\infty\text{-QAM}} \triangleq \sup_{\zeta} \left[I^{\log}_{\infty\text{-QAM}}(\zeta) - \check{I}^{\log}_{\infty\text{-QAM}}(\zeta) \right] \approx 0.0608 \text{ [bit]} \tag{76}$$

Proof: Appendix F. ∎

We let $\underline{\gamma}_0 = e^{\underline{\zeta}_0} - 1 \approx 8.76$ dB and $\tilde{\gamma}_2 = e^{\tilde{\zeta}_2} - 1 \approx 16.5$ dB. Armed with the above results, the proof of Theorem 4 is straightforward,

Proof of Theorem 4: Letting $\bar{\gamma} = \max_{\theta \in (-\pi, \pi)} |H(\theta)|^2$ and $\bar{\zeta} = \log(1 + \bar{\gamma})$, the proof that $\mathcal{I}_{\text{OFDM}} \leq I_{\text{SL}} + \bar{\Delta}_{\infty\text{-QAM}}(\bar{\gamma})$ is identical to the proof of Theorem 1, with $\hat{I}^{\log;[0,\bar{\zeta}]}_{\infty\text{-QAM}}(\zeta)$ replacing $\hat{I}^{\log}_x(\zeta)$ and $\bar{\Delta}_{\infty\text{-QAM}}(\bar{\gamma})$ replacing Δ_x. To show that $I_{\text{SL}} \leq \mathcal{I}_{\text{OFDM}} + \tilde{\Delta}_{\infty\text{-QAM}}$ we reverse the direction of the derivation:

$$\mathcal{I}_{\text{OFDM}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} I_{\infty\text{-QAM}}\left(|H(\theta)|^2\right) d\theta \tag{77}$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} I^{\log}_{\infty\text{-QAM}}\left(\log\left(1 + |H(\theta)|^2\right)\right) d\theta \tag{78}$$

$$\geq \frac{1}{2\pi} \int_{-\pi}^{\pi} \check{I}^{\log}_{\infty\text{-QAM}}\left(\log\left(1 + |H(\theta)|^2\right)\right) d\theta \tag{79}$$

$$\geq \check{I}^{\log}_{\infty\text{-QAM}}\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \log\left(1 + |H(\theta)|^2\right) d\theta\right) \tag{80}$$

$$= \check{I}^{\log}_{\infty\text{-QAM}}\left(\log\left(1 + \text{SNR}_{\text{MMSE-DFE-U}}\right)\right) \tag{81}$$

$$\geq I_{\infty\text{-QAM}}\left(\text{SNR}_{\text{MMSE-DFE-U}}\right) - \tilde{\Delta}_{\infty\text{-QAM}} = I_{\text{SL}} - \tilde{\Delta}_{\infty\text{-QAM}} \tag{82}$$

where in (80) the convexity of $\check{I}^{\log}_x$ was used to invoke Jensen's inequality. When $|H(\theta)|^2 \geq \underline{\gamma}_0$ for all $\theta \in (-\pi, \pi)$, the function $I^{\log}_x$ is convex for all values of $\log\left(1 + |H(\theta)|^2\right)$, and we may therefore replace $\check{I}^{\log}_{\infty\text{-QAM}}$ with $I^{\log}_{\infty\text{-QAM}}$ and obtain $\mathcal{I}_{\text{OFDM}} \geq I_{\text{SL}}$. When $\text{SNR}_{\text{MMSE-DFE-U}} \geq \tilde{\gamma}_2$,

$$\check{I}^{\log}_{\infty\text{-QAM}}\left(\log\left(1 + \text{SNR}_{\text{MMSE-DFE-U}}\right)\right) = I_{\infty\text{-QAM}}\left(\text{SNR}_{\text{MMSE-DFE-U}}\right) \tag{83}$$

and hence the introduction of $\tilde{\Delta}_{\infty\text{-QAM}}$ is unnecessary, resulting once more in $\mathcal{I}_{\text{OFDM}} \geq I_{\text{SL}}$. ∎

VI. DISCUSSION

In this section we use the results obtained in the paper to draw insight on the differences between the maximum achievable rates of SC and OFDM that expected in practical scenarios. Additionally, we consider how increasing the constellation order affects these difference and discuss the implications of doing so.

A. Maximum and minimum difference $\mathcal{I}_{OFDM}$ and I_{SL}

Our analysis enables us to characterize the ISI channels for which $I_{\text{SL}} - \mathcal{I}_{\text{OFDM}}$ will be smallest, and the channels for which it will be the largest. If the input distribution is such that $\Delta_x = 0$ then clearly every memoryless (flat fading) channel achieves the minimum difference $I_{\text{SL}} - \mathcal{I}_{\text{OFDM}} = 0$. If $\Delta_x > 0$, the minimum difference is obtained

by a channel with two-level transfer function,

$$|H(\theta)|^2 = \begin{cases} \gamma_1 & |\frac{\theta}{2\pi}| \leq \log\left(\frac{1+\gamma_2}{1+\gamma_m}\right) / \log\left(\frac{1+\gamma_2}{1+\gamma_1}\right) \\ \bar{\gamma}_2 & \text{otherwise} \end{cases}$$ (84)

with $\gamma_m = e^{\zeta_m} - 1$ and ζ_m as defined in (32). Clearly for this channel $I_{\text{SL}} - \mathcal{I}_{\text{OFDM}} = -\Delta_x$, which is the minimum possible difference according to Theorem 1.

Consider the channel,

$$|H(\theta)|^2 = \begin{cases} e^{\Gamma^2} - 1 & |\theta| \leq \pi/\Gamma \\ 0 & \text{otherwise} \end{cases}$$ (85)

where $\Gamma > 0$ is an arbitrary sharpness parameter. As Γ increases, the channel's frequency response becomes narrower and steeper. For this channel, $\mathcal{I}_{\text{OFDM}} = I_x\left(e^{\Gamma^2} - 1\right)/\Gamma$ and $I_{\text{SL}} = I_x\left(e^{\Gamma} - 1\right)$. For any finite-alphabet input, we therefore have $I_{\text{SL}} - \mathcal{I}_{\text{OFDM}} \to H(x_0)$ as $\Gamma \to \infty$, where $H(x_0)$ denotes the input entropy. Clearly, $H(x_0)$ is the maximum possible difference between I_{SL} and $\mathcal{I}_{\text{OFDM}}$, as it upper bounds both quantities.

Now consider uniform input, characterized in Theorem 4 and analyzed in subsection V-D. Since $I_{\infty\text{-QAM}}(\gamma) \approx \log(\gamma) - \log(\pi e/6)$ at high SNR, for the above channel we will have $\mathcal{I}_{\text{OFDM}} - I_{\text{SL}} \to \log(\pi e/6) = \Delta_{\infty\text{-QAM}}$ as $\Gamma \to \infty$. Thus, the extreme case that maximizes $I_{\text{SL}} - \mathcal{I}_{\text{OFDM}}$ for finite-alphabet inputs also minimizes it for uniform inputs. However, $\Delta_{\infty\text{-QAM}}$ is approached only for highly impractical channels. For example, in order for $\mathcal{I}_{\text{OFDM}} - I_{\text{SL}}$ to reach 90% of $\Delta_{\infty\text{-QAM}}$, we need $\Gamma \approx 10$, which yields $|H(\theta)|^2 \approx 430$ dB! Indeed, $\Delta_{\infty\text{-QAM}}$ can only be approached as $|H(\theta)|^2$ becomes exceedingly large — this is proven by the very slow rate of convergence of $\bar{\Delta}_{\infty\text{-QAM}}(\gamma)$ (Figure 1).

B. Difference between $\mathcal{I}_{OFDM}$ and I_{SL} in practical settings

In OFDM wireless communication systems, the constellation and error correcting code will usually be chosen so that the code rate is between $1/2$ and $5/6$ [8]–[11]. Assuming the system is efficient enough to have performance close to the maximum achievable rate, this means we would have

$$\frac{1}{2}H(x_0) \leq \mathcal{I}_{\text{OFDM}} = \frac{1}{2\pi}\int_{-\pi}^{\pi} I_x^{\log}(\zeta_\theta)\, d\theta \leq \frac{5}{6}H(x_0)$$ (86)

where $H(x_0)$ is the input entropy which equals the number of uncoded bits per input symbol for equiprobably inputs, and $\zeta_\theta = \log\left(1 + |H(\theta)|^2\right)$ is the log-SNR at subcarrier frequency θ. Let ζ_{OFDM} be the log-SNR in a memoryless channel with achievable rate of $\mathcal{I}_{\text{OFDM}}$, i.e. $I_x^{\log}(\zeta_{\text{OFDM}}) = \mathcal{I}_{\text{OFDM}}$. Under this notation, the single carrier achievable rate satisfies,

$$\mathcal{I}_{\text{SC}} \approx I_{\text{SL}} = I_x(\text{SNR}_{\text{MMSE-DFE-U}}) = I_x^{\log}\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\zeta_\theta\, d\theta\right)$$ (87)

Examining Figures 2 and 3 while keeping (86) in mind, we are able to estimate the performance gain of SC over OFDM for different ISI channels. Clearly, for channels with little ISI (nearly constant $|H(\theta)|^2$), there will
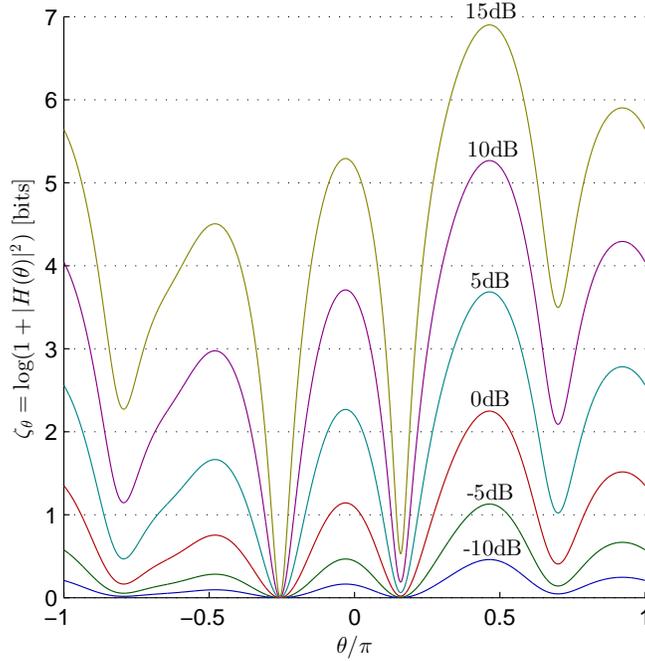
Figure 8. The channel transfer function expressed in terms of log-SNR $\zeta_\theta = \log\left(1 + |H(\theta)|^2\right)$, different values of input SNR.

be little difference between $\mathcal{I}_{\text{OFDM}}$ and $\mathcal{I}_{\text{SC}}$, as we will have $\zeta_\theta \approx \zeta_{\text{OFDM}}$ and $I_x^{\log}$ will typically be nearly linear around ζ_{OFDM}. For higher-order constellations, $\mathcal{I}_{\text{OFDM}}$ might even be slightly larger than I_{SL} as ζ_θ will take values in the interval where $I_x^{\log}$ is convex. Little performance gain is also expected when the overall code rate is low so that $\max_\theta I_x^{\log}(\zeta_\theta) \leq \frac{5}{6} H(x_0)$, since in such scenario $I_x^{\log}(\zeta_\theta)$ will be nearly linear. Conversely, a large difference between the SC and OFDM achievable rates is expected whenever there exists a significant bandwidth of sub-carriers for which $I_x^{\log}(\zeta_\theta)$ is close to the input entropy. This event is likely for channels with significant ISI, and the difference will become more pronounced as the code rate grows.

To demonstrate our conclusions we a present a numerical experiment using a 9-tap ISI channel randomly drawn from a distribution defined by the 802.11n NLOS channel model B [49]. The input SNR is given by $\frac{1}{2\pi}\int_{-\pi}^{\pi}|H(\theta)|^2\,d\theta$, and is varied by scaling $|H(\theta)|^2$. At unit input SNR and rounded to 2 significant digits, the taps of the specific channel used are given by,

$$h = \left[0.62e^{1.3j}, 0.42e^{2.8j}, 0.33e^{-1.3j}, 0.091e^{2.5j}, 0.51e^{0.66j}, 0.25e^{2j}, 0.039e^{-0.087j}, 0.028e^{-0.28j}, 0.039e^{1.7j}\right]$$

Figure 8 shows ζ_θ as a function of θ for different input SNR's. Clearly, this channel shows considerable variation in ζ_θ, and so observable differences between the achievable rates are expected. Figure 9 shows I_{SL}, $\mathcal{I}_{\text{OFDM}}$ and the difference between them for different input distributions. As expected, differences between the achievable rates are
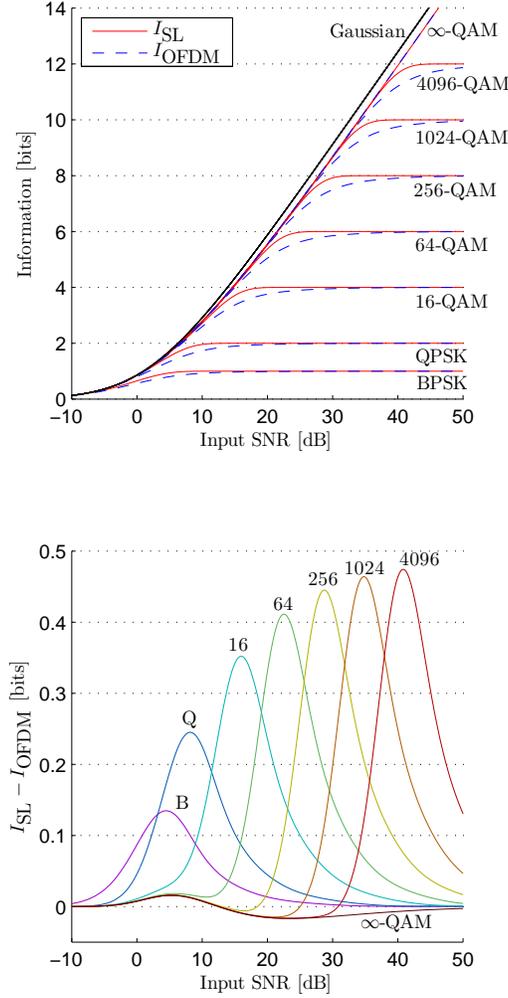
Figure 9. I_{SL} and $\mathcal{I}_{\text{OFDM}}$ (up) and $I_{\text{SL}} - \mathcal{I}_{\text{OFDM}}$ (down) as a function of input SNR, for different input distributions.

very small for low code rates, but become significant as the code rates grow. In some cases, I_{SL} is seen to exceed $\mathcal{I}_{\text{OFDM}}$ by over 15%, and for code rate 5/6 the differences between SC and OFDM in terms of required SNR reach up to 3dB. The very low rate in which $\mathcal{I}_{\text{OFDM}}$ converges to the input entropy as the SNR grows can be explained by the strong notches in the ISI transfer functions, where ζ_θ approaches the input entropy only for very high SNR's. For uniform input, $I_{\text{SL}} - \mathcal{I}_{\text{OFDM}}$ is positive for low SNR's and negative for high SNR's, as Theorem 4 predicts. Moreover, the difference is always very small, never exceeding 0.02 bits in magnitude. In particular, the maximum theoretical difference of about $\Delta_{\infty\text{-QAM}} \approx 0.509$ bits in favor of OFDM is never attained, and ever the tighter bound $\bar{\Delta}_{\infty\text{-QAM}}(60 \text{ dB}) \approx 0.228$ is quite loose for this channel.

Qualitatively, when the log-SNR at a given frequency grows, the contribution of that frequency to the overall OFDM achievable rate saturates, while its contribution to the single-carrier achievable rate continues to grow, resulting in a growing difference between the two rates. We note that this saturation effect is due to the finite-entropy nature of the input, and does not occur in the Gaussian and uniform cases. This behavior echoes the Mercury/Water-filling results of [44], where it is seen that the optimal OFDM power allocation policy for finite-entropy inputs differs significantly from classical Waterfilling in the high-SNR regime.

C. The implications of increasing the constellation order

From our analysis of uniform QAM input in Theorem 4 and from the discussion above, it is clear that for a given ISI channel and SNR, the performance of OFDM can be made close to that of SC, by sufficiently increasing the constellation order. This is due to the fact that if the input alphabet is chosen to be sufficiently large, no saturation of $I_x^{\log}$ will occur at any subcarrier frequency, and therefore no significant difference between achievable rates is to be expected.

Thus, the potential performance gain of moving from OFDM to SC, and maintaining the same constellation order, may also be realized by using OFDM with a higher-order constellation. However, there are two system design considerations that may not allow for arbitrary increase in constellation order. First, increasing the number of bits per symbol necessitates using lower code rates and perhaps more sophisticated coded modulation schemes. For example, in the setting depicted in Figure 9, for QPSK to be used with code rate 1/2 there is a difference of about 0.5dB, or 12%, between SC and OFDM in terms of SNR. Changing the constellation to 16-QAM will essentially eliminate this difference, but require an unconventional code rate of 1/4.

The second consideration is channel estimation. As the constellation order grows, the requirements on estimation accuracy of the channel gain become more stringent. Conversely, for BPSK and QPSK inputs amplitude estimation in not necessary at all. Hence, when increasing the constellation order in an OFDM system, the overhead of pilot subcarriers might have to grow as well.

In light of the issues above, as well as the sate-of-the-art OFDM wireless communication technology, where code rates below 1/2 and constellations above 256-QAM are uncommon, we conclude that using higher order constellations at low SNR's is not trivial. Therefore, we maintain that fixing the constellation order and using SC in lieu of OFDM is an option well-worth investigating.

VII. CONCLUSTION

In this paper a comparison of the achievable rates of OFDM and single-carrier modulations was performed, under the assumption of a fixed i.i.d. input distribution. In lieu of a tractable expression for the achievable rate of single-carrier modulation, the Shamai-Laroia approximation was used, since it is well known to essentially reflect tight lower bounds on the achievable rate. We revealed an intimate relation between the Shamai-Laroia approximation and the OFDM achievable rate, that stems from the concavity properties of the input-output mutual information in a scalar Gaussian channel with respect to a modified SNR variable — namely, that the Shamai-Laroia approximation is essentially an upper bound on the OFDM achievable rate.

In particular, the upper bound always holds for conventional low order input distributions including BPSK, QAM and 16-QAM. It also holds for all PAM and square QAM inputs, when the SNR exceeds a certain relatively modest threshold. Moreover, we quantified the amount by which the OFDM achievable rate might exceed the Shamai-Laroia approximation, and found it to be very small for any input distribution of interest. In contrast, we demonstrated that the Shamai-Laroia approximation may be arbitrarily larger than the OFDM rate for some ISI channels and any finite-alphabet input distribution and may provide significant improvement in practical scenarios. By similar analysis of a continuous uniform input distribution, it is shown that the difference between achievable rates can be made small by selecting a sufficiently dense input distribution. However, such choice of input might not be practical. Our conclusions extend to the case when linear precoding is allowed, giving additional validity to our assumption of i.i.d. input.

Estimation-theoretic bounds along with Information-Estimation identities were the primary tools used in our analysis. In order to establish our results regarding PAM and square QAM inputs, novel bounds on nonlinear MMSE estimation of PAM inputs in an additive Gaussian channel were developed. They include a "pointwise" bound on the conditional variance of the channel input given the channel output, as well as a tight high-SNR characterization of the MMSE. These bounds might be more widely useful.

We conclude that single-carrier modulation offers a fundamental, possibly large, improvement in spectral efficiency over OFDM when the input alphabet is constrained. However, virtually all state-of-the-art high-performance communication systems over ISI use OFDM. This is mainly due to the fact that implementing optimal joint equalization and decoding is straightforward in OFDM, but difficult in single-carrier modulation. However, practical iterative schemes that approach the SC achievable may be within reach. We believe that this work provides motivation for research and development of such schemes.

APPENDIX A
HIGH-SNR UPPER BOUND ON MMSE DERIVATIVE

Lemma 2. *For any finite-alphabet unit-power input distribution x, there exists $C > 0$ such that*

$$\mathrm{mmse}'_X(\gamma) \leq -C \frac{e^{-(d_{\min}/2)^2 \gamma}}{\sqrt{\gamma}} \tag{88}$$

for sufficiently large γ, where $d_{\min}$ is the minimum distance between any two input values.

Proof: Let $\mathcal{X}$ be the input alphabet and let $K = |\mathcal{X}|$. The derivative of the MMSE function in the complex scalar channels can be read from the results of [21],

$$\mathrm{mmse}'_X(\gamma) = -\mathrm{E}_{Y_\gamma}\left[\phi_X(Y_\gamma; \gamma) + |\psi_X(Y_\gamma; \gamma)|^2\right] \tag{89}$$

where $Y_\gamma = X + \frac{1}{\sqrt{\gamma}}N$ with N standard complex Gaussian and independent of X, and

$$\phi_X(y; \gamma) = \mathrm{E}_X\left[|X - \mathrm{E}\left[X|Y_\gamma = y\right]|^2 \mid Y_\gamma = y\right] \tag{90}$$

$$\psi_X(y; \gamma) = \mathrm{E}_X\left[(X - \mathrm{E}\left[X|Y_\gamma = y\right])^2 \mid Y_\gamma = y\right] \tag{91}$$

$\phi_X(y; \gamma)$ can be thought of as a point-wise MMSE function, given channel outcome y, but $\psi_X(y; \gamma)$ is complex and does not posses much intuitive meaning. Let x_+ and x_- be two input values such that $|x_+ - x_-| = d_{\min}$. We may assume without loss of generality that

$$x_\pm = \pm d_{\min}/2 \tag{92}$$

because the input alphabet can always be shifted and rotated so that the above relation holds. Let p_+ and p_- denote the probabilities of x_+ and x_- respectively and assume without loss of generality that $p_+ \leq p_-$. Let U be random variable independent of X and distributed on $\{0, 1\}$ with $\Pr(U = 1) = p_+/p_-$. Define the random variable $I = 1_{\{X = x_+\}} + 1_{\{X = x_- \wedge U = 1\}}$, so that given $I = 1$, X is distributed equiprobably on $\{x_+, x_-\}$. We have,

$$\phi_X(y; \gamma) = \Pr(I = 1|Y_\gamma = y)\mathrm{E}_X\left[|X - \mathrm{E}\left[X|Y_\gamma = y\right]|^2 \mid Y_\gamma = y, \ I = 1\right] \tag{93}$$

$$+ \Pr(I = 0|Y_\gamma = y)\mathrm{E}_X\left[|X - \mathrm{E}\left[X|Y_\gamma = y\right]|^2 \mid Y_\gamma = y, \ I = 0\right] \tag{94}$$

Notice that

$$\mathrm{E}_X\left[|X - \mathrm{E}\left[X|Y_\gamma = y\right]|^2 \mid Y_\gamma = y, \ I = 1\right] \geq \mathrm{E}_X\left[|X - \mathrm{E}\left[X|Y_\gamma = y, \ I = 1\right]|^2 \mid Y_\gamma = y, \ I = 1\right] \tag{95}$$

since we add the information $I = 1$ to the MMSE estimator. Since the input is binary and symmetric given $I = 1$, the RHS of the above inequality is the pointwise MMSE for symmetric binary input with variance $(d_{\min}/2)^2$ at SNR $\rho \triangleq (d_{\min}/2)^2\gamma$:

$$\mathrm{E}_X\left[|X - \mathrm{E}\left[X|Y_\gamma = y, \ I = 1\right]|^2 \mid Y_\gamma = y, \ I = 1\right] = \left(\frac{d_{\min}}{2}\right)^2 \phi_{\mathrm{BPSK}}\left(\left(\frac{d_{\min}}{2}\right)^{-1}y; \rho\right) \tag{96}$$

with

$$\phi_{\mathrm{BPSK}}(z; \gamma) = 1 - \tanh^2(2\gamma\mathrm{Re}\{z\}) = \frac{1}{\cosh^2(2\gamma\mathrm{Re}\{z\})} \tag{97}$$

therefore

$$\phi_X(y; \gamma) \geq \Pr(I = 1|Y_\gamma = y)\left(\frac{d_{\min}}{2}\right)^2 \phi_{\mathrm{BPSK}}\left(\left(\frac{d_{\min}}{2}\right)^{-1}y; \rho\right) \tag{98}$$

and so,

$$\mathrm{E}_{Y_\gamma}\left[\phi_X^2(Y_\gamma;\gamma)\right] \geq \left(\frac{d_{\min}}{2}\right)^4 \int_{\mathbb{C}} f_{Y_\gamma}(y) \left[\mathrm{Pr}(I = 1|Y_\gamma = y)\phi_{\mathrm{BPSK}}\left(\left(\frac{d_{\min}}{2}\right)^{-1} y;\rho\right)\right]^2 dy \tag{99}$$

We have

$$\mathrm{Pr}(I = 1|Y_\gamma = y)f_{Y_\gamma}(y) = \mathrm{Pr}(I = 1)f_{Y_\gamma|I}(y|I = 1) = p_+ \frac{\gamma}{\pi}\left(e^{-\gamma\left|y - \frac{d_{\min}}{2}\right|^2} + e^{-\gamma\left|y + \frac{d_{\min}}{2}\right|^2}\right) \tag{100}$$

and also

$$\mathrm{Pr}(I = 1|Y_\gamma = y) = \mathrm{Pr}(X = x_+|Y_\gamma = y) + (p_+/p_-)\mathrm{Pr}(X = x_-|Y_\gamma = y) \tag{101}$$

with

$$\mathrm{Pr}(X = x|Y_\gamma = y) = \frac{\mathrm{Pr}\left(X = x\right)e^{-\gamma|y-x|^2}}{\sum_{x'\in\mathcal{X}}\mathrm{Pr}\left(X = x'\right)e^{-\gamma|y-x'|^2}} \tag{102}$$

Let $\mathcal{D} \subseteq \mathbb{C}$ denote the set of points for which $\arg\min_{x\in\mathcal{X}}|y - x|$ is either x_+ or x_-. Clearly, for every $y \in \mathcal{D}$, either $\mathrm{Pr}(X = x_-|Y_\gamma = y) > p_-$ or $\mathrm{Pr}(X = x_+|Y_\gamma = y) > p_+$ and so

$$\mathrm{Pr}(I = 1|Y_\gamma = y) > p_+ \;\; \forall y \in \mathcal{D} \tag{103}$$

The set $\mathcal{D}$ depends on other points in $\mathcal{X}$, but can be lower bounded by $\mathcal{D}' \subseteq \mathcal{D}$ which is formed by adding to $\mathcal{X}$ all the points with distance greater than $d_{\min}$ from both x_+ and x_-. Figure 10 illustrates the construction of $\mathcal{D}'$, which is given by

$$\mathcal{D}' = \mathcal{A} \cup \mathcal{B}_+ \cup \mathcal{B}_- \tag{104}$$

where

$$\mathcal{B}_\pm = \left\{y \in \mathbb{C} \;\Big|\; \left|y \mp \frac{d_{\min}}{2}\right|^2 < \left(\frac{d_{\min}}{2}\right)^2\right\} \tag{105}$$

and

$$\mathcal{A} = \left\{y \in \mathbb{C} \;\Big|\; |\mathrm{Im}\{y\}| < \frac{|\mathrm{Re}\{y\}| + d_{\min}/2}{\sqrt{3}} \wedge |\mathrm{Re}\{y\}| < \frac{d_{\min}}{4}\right\} \tag{106}$$

Finally, the set $\mathcal{D}'$ contains the rectangular subset $\mathcal{R} \subset \mathcal{D}'$ given by[6]

$$\mathcal{R} = \left\{y \in \mathbb{C} \;\Big|\; |\mathrm{Im}\{y\}| < \frac{d_{\min}}{\sqrt{12}} \wedge |\mathrm{Re}\{y\}| < \frac{d_{\min}}{2}\right\} \tag{107}$$

Limiting the integration in (99) to $\mathcal{R}$ and substituting (100) and (103) we obtain,

$$\mathrm{E}_{Y_\gamma}\phi_X^2(Y_\gamma;\gamma) \;\geq\; p_+^2 \left(\frac{d_{\min}}{2}\right)^4 \int_{\mathcal{R}} \frac{\gamma}{\pi}\left(e^{-\gamma\left|y - \frac{d_{\min}}{2}\right|^2} + e^{-\gamma\left|y + \frac{d_{\min}}{2}\right|^2}\right)\phi_{\mathrm{BPSK}}^2\left(\frac{y}{d_{\min}/2};\rho\right) dy \tag{108}$$

$$=\; p_+^2 \left(1 - 2Q\left(\sqrt{2\rho/3}\right)\right)\left(\frac{d_{\min}}{2}\right)^4 2M(\rho) \tag{109}$$

---

[6]The real axis border of $\mathcal{R}$ can be extended to $d_{\min}\sqrt{11/12}$, but this doesn't change the leading exponent in the bound nor does it change its coefficient. It only changes the faster decreasing exponents.

with

$$M(\rho) = \frac{1}{\sqrt{\pi}} \int_{-\sqrt{\rho}}^{\sqrt{\rho}} dz \left(\frac{1}{2} e^{-(z-\sqrt{\rho})^2} + \frac{1}{2} e^{-(z+\sqrt{\rho})^2} \right) \phi_{\text{BPSK}}^2 \left(\frac{z}{\sqrt{\rho}}; \rho \right) \tag{110}$$

$$= \frac{e^{-\rho}}{\sqrt{\pi}} \int_{-\sqrt{\rho}}^{\sqrt{\rho}} dz e^{-z^2} \left[\cosh \left(2\sqrt{\rho} z \right) \right]^{-3} \tag{111}$$

$$\geq \frac{e^{-\rho}}{\sqrt{\pi}} \int_{-\sqrt{\rho}}^{\sqrt{\rho}} dz e^{-z^2(1+6\rho)} = \frac{e^{-\rho}}{\sqrt{1+6\rho}} \left(1 - 2Q \left(\sqrt{2\rho(1+6\rho)} \right) \right) \tag{112}$$

where we have used the expression (97) for ϕ_{BPSK} along with $\cosh x \leq e^{x^2/2}$ to establish the above bound. Using $\sqrt{2\pi} x Q(x) \leq e^{-x^2/2}$, we find that $Q\left(\sqrt{2\rho(1+6\rho)}\right) = o(e^{-6\rho^2})$ and so,

$$M(\rho) \geq C' \frac{e^{-\rho}}{\sqrt{\rho}} \tag{113}$$

for some $C' > 0$ and for sufficiently large ρ. Similarly, noticing that $Q\left(\sqrt{2\rho/3}\right) = o(e^{-\rho/3})$ we have also

$$\mathrm{E}_{Y_\gamma} \phi_X^2(Y_\gamma; \gamma) \geq C \frac{e^{-(d_{\min}/2)^2 \gamma}}{\sqrt{\gamma}} \tag{114}$$

for some $C > 0$ and for sufficiently large γ, where we have substituted back $\rho = (d_{\min}/2)^2 \gamma$. Finally,

$$\mathrm{mmse}_x'(\gamma) \geq -\mathrm{E}_{Y_\gamma} \phi_X^2(Y_\gamma; \gamma) \geq -C \frac{e^{-(d_{\min}/2)^2 \gamma}}{\sqrt{\gamma}} \tag{115}$$

under the same conditions. ∎

APPENDIX B
PROOF OF THEOREM 5

We begin by establishing some notation. The input alphabet will be denoted by $\mathcal{X} = \{x_1, ..., x_M\}$ and we assume that $x_{m+1} - x_m = d$ for every $1 \leq m < M$. Let

$$p_{m|y} = \Pr\left(X = x_m | Y_\gamma = y \right) = \frac{e^{-\gamma(y-x_m)^2}}{\sum_{m'=1}^{M} e^{-\gamma(y-x_{m'})^2}} \tag{116}$$

denote the probability of symbol x_m given observation y, and let

$$s(y) = \mathrm{E}\left[X | Y_\gamma = y \right] = \sum_{m=1}^{M} p_{m|y} x_m \tag{117}$$

be the expectation of X conditioned on the observation $Y_\gamma = y$, so that

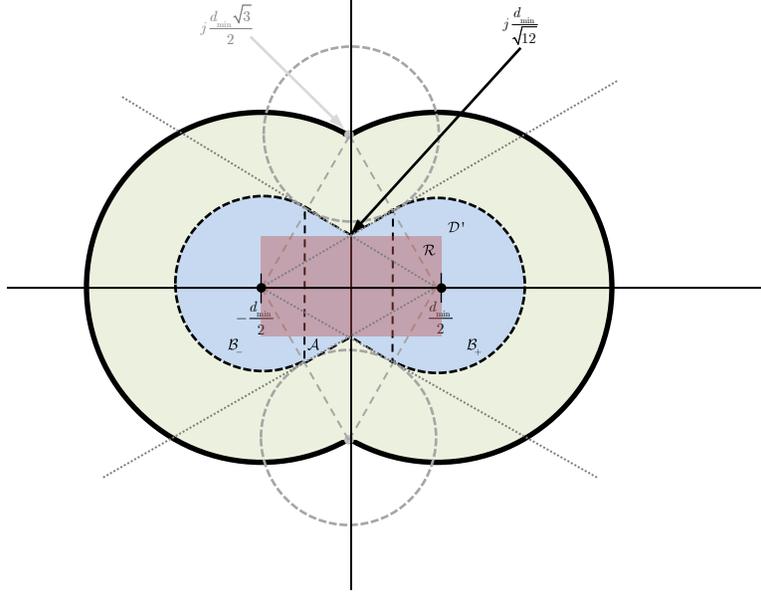$$\phi_X(y; \gamma) = \sum_{m=1}^{M} p_{m|y} \left(x_m - s(y) \right)^2 \tag{118}$$

Figure 10. Illustration of the sets $\mathcal{D}'$ (blue) and $\mathcal{R}$ (red). The black dots indicate the location of the points $x_\pm = \pm d_{\min}/2$. The region $\mathcal{D}'$ is formed by assuming the existence of other input points on the curve formed by points that have distance $d_{\min}$ from either x_+ or x_- and distance greater than $d_{\min}$ from the other point (the edge of the green region in the figure).

Notice that $s(y) = \arg\min_s \sum_{m=1}^M p_{m|y}(x_m - s)^2$ — *i.e.* the conditional expectation is the point-wise optimal estimator of X given $Y_\gamma = y$. Finally, recall that x_J, x_{J+1} denote the two nearest neighbors to y in $\mathcal{X}$.

The upper bound in Theorem 5 is derived by considering the sub-optimal estimator that assumes the input has the same distribution as B_J (uniform on $\{x_J, x_{J+1}\}$). This estimator is given by

$$\tilde{s}(y) = \left(p_{J|y}x_J + p_{J+1|y}x_{J+1}\right) / \left(p_{J|y} + p_{J+1|y}\right) \tag{119}$$

and the resulting bound reads:

$$\begin{aligned} \phi_X(y;\gamma) &\leq \sum_{m=1}^M p_{m|y}\left(x_m - \tilde{s}(y)\right)^2 \\ &\leq \sum_{j=J,J+1} \frac{p_{j|y}}{p_{J|y} + p_{J+1|y}}\left(x_j - \tilde{s}(y)\right)^2 + \sum_{m\neq J,J+1} p_{m|y}\left(x_m - \tilde{s}(y)\right)^2 \end{aligned} \tag{120}$$

The following bound is seen to hold,

$$\sum_{m=1}^{J-1} (x_m - \tilde{s}(y))^2 e^{-\gamma(y-x_m)^2} \leq d^2 e^{-\gamma(y-x_J)^2} \sum_{m=1}^{J-1} (J - m + 1)^2 e^{-\gamma(x_J - x_m)^2}$$

$$= d^2 e^{-\gamma(y-x_J)^2} \sum_{k=1}^{J-1} (k+1)^2 e^{-\gamma d^2 k^2} \leq d^2 e^{-\gamma(y-x_J)^2} \sum_{k=1}^{\infty} (k+1)^2 e^{-\gamma d^2 k^2} \qquad (121)$$

where the first transition follows from $(y - x_m)^2 \geq (y - x_J)^2 + (x_J - x_m)^2$ which holds since $x_m < x_J \leq y$, and from $(x_m - \tilde{s}(y))^2 \geq (x_m - x_J)^2 = d^2 (J - m + 1)^2$, which holds since $x_m < x_J \leq \tilde{s}(y)$. Similarly, we have

$$\sum_{m=J+2}^{M} (x_m - \tilde{s}(y))^2 e^{-\gamma(y-x_m)^2} \leq d^2 e^{-\gamma(y-x_{J+1})^2} \sum_{k=1}^{\infty} (k+1)^2 e^{-\gamma d^2 k^2} \qquad (122)$$

Using the above bounds and observing (116), we find that

$$\sum_{m \neq J, J+1} p_{m|y} (x_m - \tilde{s}(y))^2 \leq \frac{e^{-\gamma(y-x_J)^2} + e^{-\gamma(y-x_{J+1})^2}}{\sum_{m'=1}^{M} e^{-\gamma(y-x_{m'})^2}} d^2 \sum_{k=1}^{\infty} (k+1)^2 e^{-\gamma d^2 k^2}$$

$$\leq d^2 \sum_{k=1}^{\infty} (k+1)^2 e^{-\gamma d^2 k^2} \leq d^2 \sum_{k=1}^{\infty} (k+1)^2 e^{-\gamma d^2 k} \leq \frac{4 d^2 e^{-\gamma d^2}}{\left(1 - e^{-\gamma d^2}\right)^3} \qquad (123)$$

where the last inequality is due to,

$$\sum_{k=1}^{\infty} (k+1)^2 x^k \leq 2x \sum_{k=0}^{\infty} (k+2)(k+1) x^k = 2x \left(\sum_{k=0}^{\infty} x^k\right)'' = 2x \left(\frac{1}{1-x}\right)'' = \frac{4x}{(1-x)^3} \qquad (124)$$

Identifying $\sum_{j=J, J+1} \frac{p_{j|y}}{p_{J|y} + p_{J+1|y}} (x_j - \tilde{s}(y))^2$ with $\phi_{B_J}(y; \gamma)$, the upper bound follows from (120) and (123).

To prove the lower bound in Theorem 5, we first prove the following,

Lemma 3. *Let X be uniformly distributed on $\mathcal{X} = \{x_1, x_2, ..., x_M\}$ such that $x_{m+1} - x_m = d$ for all $1 \leq m < M$. For any $y \in \mathbb{R}$, let $\tilde{x}_y$ be the point in $\mathcal{X}$ with maximum distance from y. Let $\hat{X}$ be uniformly distributed on $\hat{\mathcal{X}} = \mathcal{X} \setminus \{\tilde{x}_y\}$. For every $\gamma > 0$,*

$$\phi_X(y; \gamma) \geq \phi_{\hat{X}}(y; \gamma) \qquad (125)$$

Proof: Without loss of generality, assume $y \leq (x_1 + x_M)/2$ so that $\tilde{x}_y \equiv x_M$ and $\hat{\mathcal{X}} = \{x_1, x_2, ..., x_{M-1}\}$. Let

$$\hat{s}(y) = \mathrm{E}[X | Y_\gamma = y, X \neq \tilde{x}_y] = \mathrm{E}\left[\hat{X} | \hat{Y}_\gamma = y\right] = \sum_{m=1}^{M-1} \frac{p_{m|y}}{1 - p_{M|y}} x_m \qquad (126)$$

denote the expectation of X given $Y = y$ and $X \neq \tilde{x}_y$ or equivalently the expectation of $\hat{X}$ given $\hat{Y}_\gamma = \hat{X} + \frac{1}{\sqrt{\gamma}} \hat{N} = y$, with $\hat{N} \sim \mathcal{N}(0, 1/2)$ and independent of $\hat{X}$. Notice that

$$s(y) - \hat{s}(y) = p_{M|y} (x_M - \hat{s}(y)) \qquad (127)$$

Using the orthogonality principle, we may therefore write,

$$
\begin{aligned}
\phi_X(y;\gamma) &= \sum_{m=1}^{M} p_{m|y} \left(x_m - s(y)\right)^2 = \sum_{m=1}^{M} p_{m|y} \left(x_m - \hat{s}(y)\right)^2 - \left(s(y) - \hat{s}(y)\right)^2 \\
&= \left(1 - p_{M|y}\right) \sum_{m=1}^{M-1} \frac{p_{m|y}}{1 - p_{M|y}} \left(x_m - \hat{s}(y)\right)^2 + \left(p_{M|y} - p_{M|y}^2\right)\left(x_M - \hat{s}(y)\right)^2 \\
&= \phi_{\hat{X}}(y;\gamma) + p_{M|y} \left[\left(1 - p_{M|y}\right)\left(x_M - \hat{s}(y)\right)^2 - \phi_{\hat{X}}(y;\gamma)\right]
\end{aligned}
\tag{128}
$$

By our assumption that $y \leq (x_1 + x_M)/2$ we have $p_{m|y} \geq p_{M-m+1|y}$ for every $1 \leq m \leq M/2$ and therefore $\hat{s}(y) \leq s(y) \leq (x_1 + x_M)/2$. Thus,

$$
\left(x_M - \hat{s}(y)\right)^2 \geq \left(\frac{x_M - x_1}{2}\right)^2 = \left(\frac{d}{2}\right)^2 (M-1)^2
\tag{129}
$$

We obtain the following crude upper bound for $\phi_{\hat{X}}(y;\gamma)$ by considering the suboptimal estimator $(x_1 + x_{M-1})/2$,

$$
\begin{aligned}
\phi_{\hat{X}}(y;\gamma) &\leq \sum_{m=1}^{M-1} \frac{p_{m|y}}{1 - p_{M|y}} \left(x_m - \frac{x_1 + x_{M-1}}{2}\right)^2 \\
&\leq \left(\frac{x_{M-1} - x_1}{2}\right)^2 = \left(\frac{d}{2}\right)^2 (M-2)^2
\end{aligned}
\tag{130}
$$

The second inequality follows from the fact that x_1 is the farthest point from $(x_1 + x_{M-1})/2$ in $\hat{\mathcal{X}}$ and therefore moving all probability mass to $m = 1$ increases the sum. Since x_M is farthest from y in $\mathcal{X}$, we have $p_{M|y} \leq p_{m|y}$ for any $1 \leq m < M$, and consequently

$$
p_{M|y} \leq \frac{1}{M}
\tag{131}
$$

Combining (129), (130) and (131) we find that

$$
\left(1 - p_{M|y}\right)\left(x_M - \hat{s}(y)\right)^2 - \phi_{\hat{X}}(y;\gamma) \geq \left(\frac{d}{2}\right)^2 \left(M - 1 - \frac{1}{M}\right) \geq 0
\tag{132}
$$

for every $M \geq 2$. We therefore conclude by (128) that $\phi_X(y;\gamma) \geq \phi_{\hat{X}}(y;\gamma)$ for every y and every γ. ∎

The lower bound in Theorem 5 follows immediately from Lemma 3 by applying it $M - 2$ times and obtaining a chain of inequalities, starting from $\phi_X(y;\gamma)$ and ending in $\phi_{B_J}(y;\gamma)$.

APPENDIX C
PROOF OF THEOREMS 6 AND 7

A. Lower bound on $\mathrm{mmse}_{d,M\text{-}PAM}(\gamma)$

Using the notation of Section IV, we have

$$
\mathrm{mmse}_{d,M\text{-}PAM}(\gamma) = \mathrm{E}_{Y_\gamma} \phi_X(Y_\gamma;\gamma) = \sum_{m=1}^{M} \frac{1}{M} \int_{-\infty}^{\infty} \sqrt{\frac{\gamma}{\pi}} e^{-\gamma \nu^2} \phi_X(x_m + \nu;\gamma)\, d\nu
\tag{133}
$$

Using Theorem 5, we find that for $m < M$ and $\nu \geq 0$

$$
\phi_X\left(x_m + \nu; \gamma\right) \geq \left(\frac{d}{2}\right)^2 \phi_{\text{BPSK}}\left(\left(\frac{d}{2}\right)^{-1}\left[\frac{x_m - x_{m+1}}{2} + \nu\right]; \left(\frac{d}{2}\right)^2 \gamma\right)
$$

$$
= \left(\frac{d}{2}\right)^2 \phi_{\text{BPSK}}\left(-1 + \left(\frac{d}{2}\right)^{-1}\nu; \left(\frac{d}{2}\right)^2 \gamma\right) \tag{134}
$$

for every $\nu \in \mathbb{R}$. Writing $\rho = (d/2)^2 \gamma$, and integrating the above inequality, we have

$$
\int_0^\infty \sqrt{\frac{\gamma}{\pi}} e^{-\gamma \nu^2} \phi_X\left(x_m + \nu; \gamma\right) d\nu \geq \left(\frac{d}{2}\right)^2 \int_0^\infty \sqrt{\frac{\rho}{\pi}} e^{-\rho \nu^2} \phi_{\text{BPSK}}\left(-1 + \nu; \rho\right) d\nu
$$

$$
= \left(\frac{d}{2}\right)^2 \text{mmse}_{\text{BPSK}}\left(\rho\right) - \left(\frac{d}{2}\right)^2 \int_{-\infty}^0 \sqrt{\frac{\rho}{\pi}} e^{-\rho \nu^2} \phi_{\text{BPSK}}\left(-1 + \nu; \rho\right) d\nu \tag{135}
$$

where the first transition follows from applying (134), scaling the integration variable by $d/2$ and using $\rho = (d/2)^2 \gamma$.

Since $\phi_{\text{BPSK}}\left(y; \rho\right) = \left[\cosh\left(2y\rho\right)\right]^{-2} \leq 4 e^{-4|y|\rho}$, we have the following upper bound

$$
\int_{-\infty}^0 \sqrt{\frac{\rho}{\pi}} e^{-\rho \nu^2} \phi_{\text{BPSK}}\left(-1 + \nu; \rho\right) d\nu \leq 4 \int_{-\infty}^0 \sqrt{\frac{\rho}{\pi}} e^{-\rho(\nu - 2)^2} d\nu = 4Q\left(\sqrt{8\rho}\right) = \underline{B}\left(\rho\right) \tag{136}
$$

where $Q\left(\cdot\right)$ is the standard error function (47). Therefore,

$$
\int_{-\infty}^\infty \sqrt{\frac{\gamma}{\pi}} e^{-\gamma \nu^2} \phi_X\left(x_m + \nu; \gamma\right) d\nu \geq \left(\frac{d}{2}\right)^2 \text{mmse}_{\text{BPSK}}\left(\left(\frac{d}{2}\right)^2 \gamma\right) - \underline{B}\left(\left(\frac{d}{2}\right)^2 \gamma\right) \tag{137}
$$

for every $m < M$. Similarly, for every $m > 1$ and every $\nu \leq 0$,

$$
\phi_X\left(x_m + \nu; \gamma\right) \geq \left(\frac{d}{2}\right)^2 \phi_{\text{BPSK}}\left(\left(\frac{d}{2}\right)^{-1}\left[\frac{x_m - x_{m-1}}{2} + \nu\right]; \left(\frac{d}{2}\right)^2 \gamma\right) \tag{138}
$$

and so

$$
\int_{-\infty}^0 \sqrt{\frac{\gamma}{\pi}} e^{-\gamma \nu^2} \phi_X\left(x_m + \nu; \gamma\right) d\nu \geq \left(\frac{d}{2}\right)^2 \text{mmse}_{\text{BPSK}}\left(\left(\frac{d}{2}\right)^2 \gamma\right) - \underline{B}\left(\left(\frac{d}{2}\right)^2 \gamma\right) \tag{139}
$$

Consequently, we find that

$$
\int_{-\infty}^\infty \sqrt{\frac{\gamma}{\pi}} e^{-\gamma \nu^2} \phi_X\left(x_m + \nu; \gamma\right) d\nu \geq 2\left(\frac{d}{2}\right)^2 \text{mmse}_{\text{BPSK}}\left(\left(\frac{d}{2}\right)^2 \gamma\right) - 2\underline{B}\left(\left(\frac{d}{2}\right)^2 \gamma\right) \quad \forall 1 < m < M \tag{140}
$$

while for $m = 1$ and $m = M$ it is easily seen that

$$
\int_{-\infty}^\infty \sqrt{\frac{\gamma}{\pi}} e^{-\gamma \nu^2} \phi_X\left(x_m + \nu; \gamma\right) d\nu \geq \left(\frac{d}{2}\right)^2 \text{mmse}_{\text{BPSK}}\left(\left(\frac{d}{2}\right)^2 \gamma\right) \tag{141}
$$

Substituting back to (133), we find that,

$$\mathrm{mmse}_{d,M\text{-PAM}}(\gamma) \geq 2\frac{M-1}{M}\left(\frac{d}{2}\right)^2\left[\mathrm{mmse}_{\mathrm{BPSK}}\left(\left(\frac{d}{2}\right)^2\gamma\right) - \underline{B}\left(\left(\frac{d}{2}\right)^2\gamma\right)\right] \tag{142}$$

as required.

B. Upper bound on $\mathrm{mmse}'_{d,M\text{-PAM}}(\gamma)$

Similarly to (133), we have

$$\mathrm{E}_{Y_\gamma}\phi_X^2(Y_\gamma;\gamma) = \sum_{m=1}^{M}\frac{1}{M}\int_{-\infty}^{\infty}\sqrt{\frac{\gamma}{\pi}}e^{-\gamma\nu^2}\phi_X^2(x_m+\nu;\gamma)\,d\nu \tag{143}$$

Thus, the upper bound on $\mathrm{mmse}'_{d,M\text{-PAM}}(\gamma) = -2\mathrm{E}_{Y_\gamma}\phi_X^2(Y_\gamma;\gamma)$ is obtained by applying the procedure of C-A on ϕ_X^2. In particular, similarly to (140), for $1 < m < M$ we have

$$2\int_{-\infty}^{\infty}\sqrt{\frac{\gamma}{\pi}}e^{-\gamma\nu^2}\phi_X^2(x_m+\nu;\gamma)\,\nu$$
$$\geq -2\left(\frac{d}{2}\right)^4\mathrm{mmse}'_{\mathrm{BPSK}}\left(\left(\frac{d}{2}\right)^2\gamma\right) - 2\left(\frac{d}{2}\right)^4\int_{-\infty}^{0}\sqrt{\frac{\rho}{\pi}}e^{-\rho\nu^2}\phi_{\mathrm{BPSK}}^2(-1+\nu;\rho)\,d\nu \tag{144}$$

with $\rho = (d/2)^2\gamma$. Using $\phi_{\mathrm{BPSK}}^2(y;\rho) = [\cosh(2y\rho)]^{-4} \leq 16e^{-8|y|\rho}$ we find that

$$2\int_{-\infty}^{0}\sqrt{\frac{\rho}{\pi}}e^{-\rho\nu^2}\phi_{\mathrm{BPSK}}^2(-1+\nu;\rho)\,d\nu \leq 32e^{8\rho}\int_{-\infty}^{0}\sqrt{\frac{\rho}{\pi}}e^{-\rho(\nu-4)^2}\,d\nu = 32e^{8\rho}Q\left(\sqrt{32\rho}\right) = \bar{C}(\rho) \tag{145}$$

where $Q(\cdot)$ is the standard error function (47). Moreover, similarly to (141), for $m = 1$ and $m = M$, we have

$$2\int_{-\infty}^{\infty}\sqrt{\frac{\gamma}{\pi}}e^{-\gamma\nu^2}\phi_X^2(x_m+\nu;\gamma)\,d\nu \geq -\left(\frac{d}{2}\right)^4\mathrm{mmse}'_{\mathrm{BPSK}}\left(\left(\frac{d}{2}\right)^2\gamma\right) \tag{146}$$

We thus conclude that,

$$\mathrm{mmse}'_{d,M\text{-PAM}}(\gamma) \leq 2\frac{M-1}{M}\left(\frac{d}{2}\right)^4\left[\mathrm{mmse}'_{\mathrm{BPSK}}\left(\left(\frac{d}{2}\right)^2\gamma\right) + \bar{C}\left(\left(\frac{d}{2}\right)^2\gamma\right)\right] \tag{147}$$

C. Lower bound on $\mathrm{mmse}'_{d,M\text{-PAM}}(\gamma)$

'

We apply the pointwise upper bound of theorem 5 to obtain (similarly to (134)),

$$\frac{\phi_X(x_m+\nu;\gamma)}{(d/2)^2} \leq \bar{D}\left(\left(\frac{d}{2}\right)^2\gamma\right) + \begin{cases}\phi_{\mathrm{BPSK}}\left(-1+\frac{\nu}{d/2};\left(\frac{d}{2}\right)^2\gamma\right) & 0 \leq \nu \leq d \\ 1 & \nu \geq d\end{cases} \tag{148}$$

with $\bar{D}(\gamma) = 4\sum_{k=1}^{\infty}(k+1)^2e^{-4\gamma k^2}$ and we have used the fact that $\phi_{\mathrm{BPSK}}(y;\rho) \leq 1$. Squaring this inequality,

we have

$$\frac{\phi_X^2\left(x_m+\nu;\gamma\right)}{\left(d/2\right)^4} \leq c\left(\left(\frac{d}{2}\right)^2\gamma\right) + \begin{cases} \phi_{\mathrm{BPSK}}^2\left(-1+\frac{\nu}{d/2};\left(\frac{d}{2}\right)^2\gamma\right) & 0 \leq \nu \leq d \\ 1 & \nu \geq d \end{cases} \tag{149}$$

where $c\left(\rho\right) = 2\bar{D}\left(\rho\right) + \bar{D}^2\left(\rho\right)$. Letting $\rho = \left(d/2\right)^2\gamma$, we have for $m < M$,

$$\left(\frac{d}{2}\right)^{-4}\int_0^\infty \sqrt{\frac{\gamma}{\pi}}e^{-\gamma\nu^2}\phi_X^2\left(x_m+\nu;\gamma\right)d\nu \leq$$

$$\int_0^2 \sqrt{\frac{\rho}{\pi}}e^{-\rho\nu^2}\phi_{\mathrm{BPSK}}^2\left(-1+\nu;\rho\right)d\nu + \int_2^\infty \sqrt{\frac{\rho}{\pi}}e^{-\rho\nu^2}d\nu + c\left(\rho\right) \leq$$

$$\int_{-\infty}^\infty \sqrt{\frac{\rho}{\pi}}e^{-\rho\nu^2}\phi_{\mathrm{BPSK}}^2\left(\nu;\rho\right)d\nu + Q\left(\sqrt{8\rho}\right) + c\left(\rho\right) =$$

$$-\frac{1}{2}\mathrm{mmse}_{\mathrm{BPSK}}'\left(\rho\right) + \frac{1}{2}\underline{C}\left(\rho\right) \tag{150}$$

with $\underline{C}\left(\rho\right) = 2\left[c\left(\rho\right) + Q\left(\sqrt{8\rho}\right)\right]$ and $Q\left(\cdot\right)$ the standard error function (47). The first transition in the above equation follows from integrating (149) and scaling the integration variable by $d/2$ as in (135). The second transition is obtained by extending the integration limits of the first term, and evaluating the integral in the second term. Similarly, for $m > 1$ we have

$$\left(\frac{d}{2}\right)^{-4}\int_{-\infty}^0 \sqrt{\frac{\gamma}{\pi}}e^{-\gamma\nu^2}\phi_X^2\left(x_m+\nu;\gamma\right)d\nu \leq -\frac{1}{2}\mathrm{mmse}_{\mathrm{BPSK}}'\left(\rho\right) + \frac{1}{2}\underline{C}\left(\rho\right) \tag{151}$$

and for $m = 1, M$ it is simple to show that

$$\left(\frac{d}{2}\right)^{-4}\int_{-\infty}^\infty \sqrt{\frac{\gamma}{\pi}}e^{-\gamma\nu^2}\phi_X^2\left(x_m+\nu;\gamma\right)d\nu \leq -\frac{1}{2}\mathrm{mmse}_{\mathrm{BPSK}}'\left(\rho\right) + \frac{1}{2}\underline{C}\left(\rho\right) \;,\; m = 1, M \tag{152}$$

Therefore

$$\mathrm{mmse}_{d,M\text{-PAM}}'\left(\gamma\right) = -2\sum_{m=1}^M \frac{1}{M}\int_{-\infty}^\infty \sqrt{\frac{\gamma}{\pi}}e^{-\gamma\nu^2}\phi_X^2\left(x_m+\nu;\gamma\right)d\nu$$

$$\geq 2\frac{M-1}{M}\left(\frac{d}{2}\right)^4\left[\mathrm{mmse}_{\mathrm{BPSK}}'\left(\left(\frac{d}{2}\right)^2\gamma\right) - \underline{C}\left(\left(\frac{d}{2}\right)^2\gamma\right)\right] \tag{153}$$

D. Upper bound on $\mathrm{mmse}_{d,M\text{-PAM}}\left(\gamma\right)$

The upper bound on $\mathrm{mmse}_{d,M\text{-PAM}}\left(\gamma\right)$ may be derived in the same way as the lower bound on $\mathrm{mmse}_{d,M\text{-PAM}}'\left(\gamma\right)$. However, we will take a slightly different approach in order to obtain a better expression for the slackness term $\bar{B}\left(\gamma\right)$. Let $\tilde{s}\left(y\right)$ be the sub-optimal estimator for X that assumes X is uniformly distributed on the two nearest

neighbors to y in $\mathcal{X}$. We have

$$\mathrm{mmse}_{d, M\text{-PAM}}\left(\gamma\right) = \mathrm{E}\left(X - \mathrm{E}\left[X|Y_\gamma\right]\right)^2$$

$$\leq \mathrm{E}\left(X - \tilde{s}\left(Y_\gamma\right)\right)^2 = \sum_{m=1}^{M} \frac{1}{M} \int_{-\infty}^{\infty} \sqrt{\frac{\gamma}{\pi}} e^{-\gamma\nu^2} \left(x_m - \tilde{s}\left(x_m + \nu\right)\right)^2 d\nu \tag{154}$$

For convenience denote $x_{M+1} \equiv \infty$ and $x_0 = -\infty$. We observe that for any $m' \geq m$ and any $x_{m'} \leq \nu \leq x_{m'+1}$,

$$\left(x_m - \tilde{s}\left(x_m + \nu\right)\right)^2 \leq d^2 \left(m' - m + 1\right)^2 \tag{155}$$

Therefore, for any $m < M$,

$$\int_0^{\infty} \sqrt{\frac{\gamma}{\pi}} e^{-\gamma\nu^2} \left(x_m - \tilde{s}\left(x_m + \nu\right)\right)^2 d\nu \leq \int_0^d \sqrt{\frac{\gamma}{\pi}} e^{-\gamma\nu^2} \left(x_m - \tilde{s}\left(x_m + \nu\right)\right)^2 d\nu$$

$$+ \sum_{m'=m+1}^{M} d^2 \left(m' - m + 1\right)^2 \int_{x_{m'}-x_m}^{x_{m'+1}-x_m} \sqrt{\frac{\gamma}{\pi}} e^{-\gamma\nu^2} d\nu \tag{156}$$

The first term is clearly upper bounded by $\left(\frac{d}{2}\right)^2 \mathrm{mmse}_{\mathrm{BPSK}}\left(\left(\frac{d}{2}\right)^2 \gamma\right)$:

$$\int_0^d \sqrt{\frac{\gamma}{\pi}} e^{-\gamma\nu^2} \left(x_m - \tilde{s}\left(x_m + \nu\right)\right)^2 d\nu = \left(\frac{d}{2}\right)^2 \int_0^d \sqrt{\frac{\gamma}{\pi}} e^{-\gamma\nu^2} \phi_{\mathrm{BPSK}}\left(-1 + \left(\frac{d}{2}\right)^{-1}\nu; \left(\frac{d}{2}\right)^2 \gamma\right) d\nu$$

$$\leq \left(\frac{d}{2}\right)^2 \int_{-\infty}^{\infty} \sqrt{\frac{(d/2)^2 \gamma}{\pi}} e^{-(d/2)^2\gamma\nu^2} \phi_{\mathrm{BPSK}}\left(-1 + \nu; \left(\frac{d}{2}\right)^2 \gamma\right) d\nu$$

$$= \left(\frac{d}{2}\right)^2 \mathrm{mmse}_{\mathrm{BPSK}}\left(\left(\frac{d}{2}\right)^2 \gamma\right) \tag{157}$$

The second term can be upper bounded as follows

$$\sum_{m'=m+1}^{M} d^2 \left(m' - m + 1\right)^2 \int_{x_{m'}-x_m}^{x_{m'+1}-x_m} \sqrt{\frac{\gamma}{\pi}} e^{-\gamma\nu^2} d\nu$$

$$= \sum_{k=1}^{M-m-1} d^2 \left(k+1\right)^2 \int_{kd}^{(k+1)d} \sqrt{\frac{\gamma}{\pi}} e^{-\gamma\nu^2} d\nu + d^2 \left(M - m + 1\right)^2 \int_{(M-m)d}^{\infty} \sqrt{\frac{\gamma}{\pi}} e^{-\gamma\nu^2} d\nu$$

$$= \sum_{k=1}^{M-m-1} d^2 \left(k+1\right)^2 \left[Q\left(kd\sqrt{2\gamma}\right) - Q\left((k+1) d\sqrt{2\gamma}\right)\right] + d^2 \left(M - m + 1\right)^2 Q\left((M - m) d\sqrt{2\gamma}\right)$$

$$= 4d^2 Q\left(\sqrt{2d^2\gamma}\right) + d^2 \sum_{k=2}^{M-m} \left(2k+1\right) Q\left(k\sqrt{2d^2\gamma}\right)$$

$$\leq 4d^2 Q\left(\sqrt{2d^2\gamma}\right) + d^2 \sum_{k=2}^{\infty} \left(2k+1\right) Q\left(k\sqrt{2d^2\gamma}\right) = \left(\frac{d}{2}\right)^2 \bar{B}\left(\left(\frac{d}{2}\right)^2 \gamma\right) \tag{158}$$

where $Q\left(\cdot\right)$ is the standard error function (47). This upper bound can be slightly relaxed to obtain a more manageable expression, using the inequality $\sqrt{2\pi}xQ\left(x\right) \leq e^{-x^2/2}$:

$$
\begin{aligned}
\bar{B}\left(\rho\right) &\leq \frac{16}{\sqrt{16\pi\rho}}e^{-4\rho} + \frac{4}{\sqrt{16\pi\rho}}\sum_{k=2}^{\infty}\frac{2k+1}{k}e^{-4k^2\rho} \\
&\leq \frac{1}{\sqrt{\pi\rho}}\left(4e^{-4\rho} + \frac{5}{2}e^{-16\rho}\sum_{k=0}^{\infty}e^{-4k(k+4)\rho}\right) \\
&\leq \frac{1}{\sqrt{\pi\rho}}\left(4e^{-4\rho} + \frac{5}{2}e^{-16\rho}\sum_{k=0}^{\infty}e^{-20k\rho}\right) = \frac{1}{2\sqrt{\pi\rho}}\left(8e^{-4\rho} + 5\frac{e^{-16\rho}}{1-e^{-20\rho}}\right)
\end{aligned}
\tag{159}
$$

where we used $\left(2k+1\right)/k \leq 5/2$ for every $k \geq 2$ and $4k\left(k+4\right) \geq 20k$ for every $k \geq 0$. We conclude that, with $\rho = \left(d/2\right)^2\gamma$,

$$
\int_{0}^{\infty}\sqrt{\frac{\gamma}{\pi}}e^{-\gamma\nu^2}\left(x_m - \tilde{s}\left(x_m + \nu\right)\right)^2 d\nu \leq \left(\frac{d}{2}\right)^2\left[\mathrm{mmse_{BPSK}}\left(\rho\right) + \bar{B}\left(\rho\right)\right], \; \forall m < M
\tag{160}
$$

and it may similarly be shown that,

$$
\int_{-\infty}^{0}\sqrt{\frac{\gamma}{\pi}}e^{-\gamma\nu^2}\left(x_m - \tilde{s}\left(x_m + \nu\right)\right)^2 d\nu \leq \left(\frac{d}{2}\right)^2\left[\mathrm{mmse_{BPSK}}\left(\rho\right) + \bar{B}\left(\rho\right)\right], \; \forall m > 1
\tag{161}
$$

It is also simple to show that for $m = 1, M$,

$$
\int_{-\infty}^{\infty}\sqrt{\frac{\gamma}{\pi}}e^{-\gamma\nu^2}\left(x_m - \tilde{s}\left(x_m + \nu\right)\right)^2 d\nu \leq \left(\frac{d}{2}\right)^2\left[\mathrm{mmse_{BPSK}}\left(\rho\right) + \bar{B}\left(\rho\right)\right], \; \forall m = 1, M
\tag{162}
$$

and so

$$
\begin{aligned}
\mathrm{mmse}_{d,M\text{-PAM}}\left(\gamma\right) &\leq \sum_{m=1}^{M}\frac{1}{M}\int_{-\infty}^{\infty}\sqrt{\frac{\gamma}{\pi}}e^{-\gamma\nu^2}\left(x_m - \tilde{s}\left(x_m + \nu\right)\right)^2 d\nu \\
&\leq 2\frac{M-1}{M}\left(\frac{d}{2}\right)^2\left[\mathrm{mmse_{BPSK}}\left(\left(\frac{d}{2}\right)^2\gamma\right) + \bar{B}\left(\left(\frac{d}{2}\right)^2\gamma\right)\right]
\end{aligned}
\tag{163}
$$

Appendix D
Proof of Theorem 8

Using $\phi_{\mathrm{BPSK}}(y;\gamma) = 1/\cosh^2\left(2\gamma y\right)$, we find that

$$
\begin{aligned}
\mathrm{mmse_{BPSK}}\left(\gamma\right) &= \sqrt{\frac{\gamma}{\pi}}\int_{-\infty}^{\infty}\phi_{\mathrm{BPSK}}(y;\gamma)\left(\frac{e^{-\gamma(y-1)^2} + e^{-\gamma(y+1)^2}}{2}\right)dy \\
&= \sqrt{\frac{\gamma}{\pi}}e^{-\gamma}\int_{-\infty}^{\infty}\frac{1}{\cosh\left(2\gamma y\right)}e^{-\gamma y^2}dy \\
&= \frac{1}{\sqrt{\pi\gamma}}e^{-\gamma}\int_{-\infty}^{\infty}\frac{1}{\cosh\left(2z\right)}e^{-\frac{z^2}{\gamma}}dz
\end{aligned}
\tag{164}
$$

and (56) is readily found by substituting $1 - \frac{z^2}{\gamma} \leq e^{-\frac{z^2}{\gamma}} \leq 1$ and integrating. Note that by substituting $e^{-z^2/\gamma} = \sum_{k=0}^{\infty} \frac{1}{k!} \left(-z^2/\gamma\right)^k$, the high-SNR asymptotic expansion of $\mathrm{mmse}_{\mathrm{BPSK}}(\gamma)$ is obtained. A different change of variables yields the equality,

$$\mathrm{mmse}_{\mathrm{BPSK}}(\gamma) = \frac{1}{\sqrt{\pi}} e^{-\gamma} \int_{-\infty}^{\infty} \frac{1}{\cosh\left(2\sqrt{\gamma}z\right)} e^{-z^2} dz \tag{165}$$

and substituting $1 \leq \cosh\left(2\sqrt{\gamma}z\right) \leq e^{2\gamma z^2}$ yields the bounds in (57). Since $\mathrm{mmse}'_{\mathrm{BPSK}}(\gamma) = -2\mathrm{E}_{Y_\gamma} \phi^2_{\mathrm{BPSK}}(Y_\gamma; \gamma)$, we find the bounds for $\mathrm{mmse}'_{\mathrm{BPSK}}(\gamma)$ by replacing $\cosh(\cdot)$ with $\cosh^3(\cdot)$ in the derivations above.

APPENDIX E

PROOF OF PROPOSITION 4

Let X be a real-valued RV uniformly distributed in $[-A/2, A/2]$, and using the notation of Section IV let $Y_\gamma = X + \frac{1}{\sqrt{\gamma}} N$ with $N \sim \mathcal{N}(0, \frac{1}{2})$ and independent of X. Using the orthogonality principle and considering the measurement Y_γ as a suboptimal estimator, we may express the MMSE as

$$\mathrm{mmse}_X(\gamma) = \frac{1}{2\gamma} - \mathrm{E}\left(Y_\gamma - \mathrm{E}\left[X|Y_\gamma\right]\right)^2 \tag{166}$$

Straightforward calculation of $\mathrm{E}\left(Y_\gamma - \mathrm{E}\left[X|Y_\gamma\right]\right)^2$ shows that we may write

$$\mathrm{mmse}_X(\gamma) = \frac{1}{2\gamma}\left(1 - \int_{-\infty}^{\infty} g(y;\gamma) dy\right) \tag{167}$$

with

$$g(y;\gamma) \triangleq \frac{1}{2\pi A} \frac{\left(e^{-\gamma(y-A/2)^2} - e^{-\gamma(y+A/2)^2}\right)^2}{Q\left(\sqrt{2\gamma}\left[y - A/2\right]\right) - Q\left(\sqrt{2\gamma}\left[y + A/2\right]\right)} \tag{168}$$

and with the error function $Q(\cdot)$ defined in (47).

Differentiating (167), we have

$$\mathrm{mmse}'_X(\gamma) = -\frac{1}{\gamma}\mathrm{mmse}_X(\gamma) + \frac{1}{2\gamma}\int_{-\infty}^{\infty}\left[h_1(y;\gamma) - h_2(y;\gamma)\right] dy \tag{169}$$

where

$$h_1(y;\gamma) \triangleq \frac{1}{\pi A} \frac{\left((y-A/2)^2 e^{-\gamma(y-A/2)^2} - (y+A/2)^2 e^{-\gamma(y+A/2)^2}\right)\left(e^{-\gamma(y-A/2)^2} - e^{-\gamma(y+A/2)^2}\right)}{Q\left(\sqrt{2\gamma}\left[y - A/2\right]\right) - Q\left(\sqrt{2\gamma}\left[y + A/2\right]\right)} \tag{170}$$

and

$$h_2(y;\gamma) \triangleq \frac{1}{4A\pi\sqrt{\pi\gamma}} \frac{\left((y-A/2) e^{-\gamma(y-A/2)^2} - (y+A/2) e^{-\gamma(y+A/2)^2}\right)\left(e^{-\gamma(y-A/2)^2} - e^{-\gamma(y+A/2)^2}\right)^2}{\left(Q\left(\sqrt{2\gamma}\left[y - A/2\right]\right) - Q\left(\sqrt{2\gamma}\left[y + A/2\right]\right)\right)^2} \tag{171}$$

For $|y| \leq A/2$ we have $h_2(y; \gamma) \leq 0$. For $y > A/2$ we find that

$$h_2(y; \gamma) \leq \frac{1}{4A\pi\sqrt{\pi\gamma}} \frac{(y - A/2)e^{-3\gamma(y-A/2)^2}}{Q\left(\sqrt{2\gamma}[y - A/2]\right)^2} \tag{172}$$

where we have used

$$\frac{e^{-\gamma(y-A/2)^2} - e^{-\gamma(y+A/2)^2}}{Q\left(\sqrt{2\gamma}[y - A/2]\right) - Q\left(\sqrt{2\gamma}[y + A/2]\right)} \leq \frac{e^{-\gamma(y-A/2)^2}}{Q\left(\sqrt{2\gamma}[y - A/2]\right)} \tag{173}$$

for every γ, A and y. Integrating, we have

$$\int_{-\infty}^{\infty} h_2(y; \gamma) \, dy = 2\int_0^{\infty} h_2(y; \gamma) \, dz \leq 2\int_{A/2}^{\infty} h_2(y; \gamma) \, dz = \frac{c_2}{2A\pi\sqrt{\pi}\gamma\sqrt{\gamma}} \tag{174}$$

where

$$c_2 = \int_0^{\infty} \frac{xe^{-3x^2}}{Q\left(\sqrt{2}x\right)^2} dx \approx 10.6 \tag{175}$$

Turning to h_1, we find that for $y > 0$,

$$h_1(y; \gamma) \geq \frac{1}{\pi A} \frac{(y - A/2)^2 e^{-2\gamma(y-A/2)^2} - \left[(y - A/2)^2 + (y + A/2)^2\right] e^{-\gamma(y-A/2)^2} e^{-\gamma(y+A/2)^2}}{Q\left(\sqrt{2\gamma}[y - A/2]\right) - Q\left(\sqrt{2\gamma}[y + A/2]\right)}$$

$$\geq \frac{1}{\pi A\gamma} \frac{\gamma(y - A/2)^2 e^{-2\gamma(y-A/2)^2}}{Q\left(\sqrt{2\gamma}[y - A/2]\right)} - \frac{1}{\pi A} \frac{\left(A^2/2 + 2y^2\right)e^{-2\gamma y^2}}{Q\left(\sqrt{2\gamma}[y - A/2]\right) - Q\left(\sqrt{2\gamma}[y + A/2]\right)} e^{-A^2\gamma/2} \tag{176}$$

Therefore,

$$\int_{-\infty}^{\infty} h_1(y; \gamma) \, dy = 2\int_0^{\infty} h_1(y; \gamma) \, dz \geq \frac{2c_1}{\pi A\gamma\sqrt{\gamma}} - \left[\frac{A}{\pi\sqrt{\gamma}}K_0(A) + \frac{4}{\pi A\gamma\sqrt{\gamma}}K_2(A)\right] e^{-A^2\gamma/2} \tag{177}$$

with

$$c_1 = \int_0^{\infty} \frac{x^2 e^{-2x^2}}{Q\left(\sqrt{2}x\right)} dz \approx 2.26 \tag{178}$$

and

$$K_i(A) = \int_0^{\infty} \frac{x^i e^{-2x^2} dx}{Q\left(\sqrt{2}(x - A/2)\right) - Q\left(\sqrt{2}(x + A/2)\right)} \tag{179}$$

Putting the bounds together, and simplifying the exponential term by assuming $\gamma > 1$,

$$\text{mmse}_X'(\gamma) \geq -\frac{1}{\gamma}\text{mmse}_X(\gamma) + \frac{c_0}{\pi A\gamma^2\sqrt{\gamma}} - k(A)e^{-A^2\gamma/2} \tag{180}$$

where

$$c_0 = c_1 - \frac{1}{4\sqrt{\pi}}c_2 \approx 0.77 \tag{181}$$

and

$$k\left(A\right) = \frac{A}{\pi}K_0\left(A\right) + \frac{4}{\pi A}K_2\left(A\right) \tag{182}$$

Consequently,

$$\mathrm{mmse}_X\left(\gamma\right) + \left(1+\gamma\right)\mathrm{mmse}_X'\left(\gamma\right) \geq \frac{c_0}{\pi A\gamma\sqrt{\gamma}} - \frac{1}{2\gamma^2} - k\left(A\right)e^{-A^2\gamma/2} \tag{183}$$

for $\gamma > 1$, where we have used (180) along with $\mathrm{mmse}_X\left(\gamma\right) \leq 1/2\gamma$ which is true for any input.

Let X be the in-phase or quadrature component of a unit power ∞-QAM input, so that $A = \sqrt{6}$ (X has variance $1/2$), and

$$\mathrm{mmse}_{\infty\text{-QAM}}\left(\gamma\right) = 2\mathrm{mmse}_X\left(\gamma\right) \tag{184}$$

Therefore, using (22) and (183), we find that a sufficient condition for $I^{\log}_{\infty\text{-QAM}}$ to be convex is

$$\frac{c_0}{\pi\sqrt{6}\gamma\sqrt{\gamma}} - \frac{1}{2\gamma^2} - k\left(\sqrt{6}\right)e^{-3\gamma} \geq 0 \tag{185}$$

As a result, there must exist a value of γ above which convexity holds. Using $k\left(\sqrt{6}\right) \approx 0.586$, it is seen that the above inequality becomes positive for $\gamma > 25$, or 14 dB, and therefore convexity holds above this value. Numerically examining $I^{\log}_{\infty\text{-QAM}}$ and its derivatives for SNR's below 14 dB, it is seen that the function is concave below $\underline{\gamma}_0 = 8.76$ dB and then becomes convex. The above analysis guarantees that $I^{\log}_{\infty\text{-QAM}}$ never becomes concave again at higher SNR's.

We remark that bounds (174) and (177) could have been made tighter by extending the lower integration limit in (175) and (178) to $-\infty$, at the cost of adding additional exponential factors.

Appendix F
Proof of Proposition 5

First, we show that the concave envelope of $I^{\log}_{\infty\text{-QAM}}$ is $\hat{I}^{\log}_{\infty\text{-QAM}}\left(\zeta\right) = \zeta$. Assume by contradiction that there exists another concave function $\tilde{I}\left(\zeta\right)$ that upper bounds $I^{\log}_{\infty\text{-QAM}}$ and satisfies $\tilde{I}(\zeta_a) < \zeta_a$ for some $\zeta_c \geq 0$. Since $\tilde{I}\left(0\right) \geq I^{\log}_{\infty\text{-QAM}}\left(0\right) = 0$, we must have $\tilde{I}'(\zeta_i) < 1$ for some $\zeta_b \in [0, \zeta_a)$ for $\tilde{I}(\zeta_a) < \zeta_a$ to be possible. By the concavity of $\tilde{I}$, $\tilde{I}'$ is non-increasing, and hence $\tilde{I}\left(\zeta\right) \leq \tilde{I}(\zeta_b) + (\zeta - \zeta_b)\tilde{I}'(\zeta_b)$ for $\zeta \geq \zeta_b$. However, by (70) we clearly have that for any $C \in \mathbb{R}$ and $\alpha < 1$, $I^{\log}_{\infty\text{-QAM}}\left(\zeta\right) > C + \alpha\zeta$ for sufficiently high ζ. There must therefore exist $\zeta_c \geq 0$ such that

$$I^{\log}_{\infty\text{-QAM}}\left(\zeta_c\right) > \tilde{I}(\zeta_b) + (\zeta_c - \zeta_b)\tilde{I}'(\zeta_b) \geq \tilde{I}\left(\zeta_c\right) \tag{186}$$

forming a contradiction. We conclude that the concave envelope satisfies $\hat{I}^{\log}_{\infty\text{-QAM}}\left(\zeta\right) \geq \zeta$. Clearly, ζ is concave and upper bounds $I^{\log}_{\infty\text{-QAM}}$ and therefore $\hat{I}^{\log}_{\infty\text{-QAM}}\left(\zeta\right) = \zeta$.

For any input distribution, $I^{\log\prime}_x\left(\zeta\right) = \left(1+\gamma\right)\mathrm{mmse}_x(\gamma) \leq 1$. Therefore, $\zeta - I^{\log}_{\infty\text{-QAM}}\left(\zeta\right)$ is an increasing function. Thus, given (70) and the expression for $\hat{I}^{\log}_{\infty\text{-QAM}}$, we may easily find the maximum difference between it

and $I_{\infty\text{-QAM}}^{\log}$,

$$\Delta_{\infty\text{-QAM}} = \sup_{\zeta} \left(\hat{I}_{\infty\text{-QAM}}^{\log}(\zeta) - I_{\infty\text{-QAM}}^{\log}(\zeta) \right) = \lim_{\zeta \to \infty} \left(\zeta - I_{\infty\text{-QAM}}^{\log}(\zeta) \right) = \log\left(\frac{\pi e}{6} \right) \tag{187}$$

We now consider the interval $[0, \bar{\zeta}]$ for some $\bar{\zeta} > \underline{\zeta}_0$. Since the constant function $I_{\infty\text{-QAM}}^{\log}(\bar{\zeta})$ is concave and upper bounds $I_{\infty\text{-QAM}}^{\log}$ on the interval, we must have $\hat{I}_{\infty\text{-QAM}}^{\log;[0,\bar{\zeta}]}(\bar{\zeta}) \leq I_{\infty\text{-QAM}}^{\log}(\bar{\zeta})$. By definition, $\hat{I}_{\infty\text{-QAM}}^{\log;[0,\bar{\zeta}]}$ also upper bounds $I_{\infty\text{-QAM}}^{\log}$, and so we must have $\hat{I}_{\infty\text{-QAM}}^{\log;[0,\bar{\zeta}]}(\bar{\zeta}) = I_{\infty\text{-QAM}}^{\log}(\bar{\zeta})$. However, since $\bar{\zeta} > \underline{\zeta}_0$, by Proposition 4 $I_{\infty\text{-QAM}}^{\log}$ is convex around $\bar{\zeta}$, and therefore $\hat{I}_{\infty\text{-QAM}}^{\log;[0,\bar{\zeta}]}(\zeta)$ cannot be identical to $I_{\infty\text{-QAM}}^{\log}(\zeta)$ in a neighborhood of $\bar{\zeta}$. Hence, there exists $\underline{\zeta}_1$ such that $\hat{I}_{\infty\text{-QAM}}^{\log;[0,\bar{\zeta}]}$ is linear on the interval $[\underline{\zeta}_1, \bar{\zeta}]$ and that $\hat{I}_{\infty\text{-QAM}}^{\log;[0,\bar{\zeta}]}(\underline{\zeta}_1) = I_{\infty\text{-QAM}}^{\log}(\underline{\zeta}_1)$. By Proposition 4, $I_{\infty\text{-QAM}}^{\log}{}'$ has only a single minimum, located at $\underline{\zeta}_0$, below which $I_{\infty\text{-QAM}}^{\log}$ is concave. As is easily confirmed from inspection of Figure 7, this implies that $\underline{\zeta}_1 < \underline{\zeta}_0$ and that $\hat{I}_{\infty\text{-QAM}}^{\log;[0,\bar{\zeta}]}$ is given by (73), since $\hat{I}_{\infty\text{-QAM}}^{\log;[0,\bar{\zeta}]}$ may be identical to $I_{\infty\text{-QAM}}^{\log}$ in the interval $[0, \underline{\zeta}_1]$, where the latter is concave. Moreover, $\underline{\zeta}_1$ is uniquely determined by (73) and the condition $\hat{I}_{\infty\text{-QAM}}^{\log;[0,\bar{\zeta}]}(\bar{\zeta}) = I_{\infty\text{-QAM}}^{\log}(\bar{\zeta})$.

Since

$$\frac{d}{d\zeta} \left(\hat{I}_{\infty\text{-QAM}}^{\log;[0,\bar{\zeta}]}(\zeta) - I_{\infty\text{-QAM}}^{\log}(\zeta) \right) = I_{\infty\text{-QAM}}^{\log}{}'(\underline{\zeta}_1) - I_{\infty\text{-QAM}}^{\log}{}'(\zeta) \tag{188}$$

for $\zeta \in [\underline{\zeta}_1, \bar{\zeta}]$, the maximum difference between $I_{\infty\text{-QAM}}^{\log}$ and its concave envelope on $[0, \bar{\zeta}]$ is obtained for ζ_m which satisfies $I_{\infty\text{-QAM}}^{\log}{}'(\zeta_m) = I_{\infty\text{-QAM}}^{\log}{}'(\underline{\zeta}_1)$ and may therefore be easily found numerically.

The construction of the convex envelope of $I_{\infty\text{-QAM}}^{\log}$ follows exactly the same lines as the construction of $\hat{I}_{\infty\text{-QAM}}^{\log;[0,\bar{\zeta}]}$ above. Since the convex function 0 lower bounds $I_{\infty\text{-QAM}}^{\log}$, we must have $\check{I}_{\infty\text{-QAM}}^{\log}(0) \geq 0$. By definition, $\check{I}_{\infty\text{-QAM}}^{\log}$ also lower bounds $I_{\infty\text{-QAM}}^{\log}$, and so we must have $\check{I}_{\infty\text{-QAM}}^{\log}(0) = 0$. However, by Proposition 4 $I_{\infty\text{-QAM}}^{\log}$ is concave around $\zeta = 0$, and therefore $\check{I}_{\infty\text{-QAM}}^{\log}(\zeta)$ cannot be identical to $I_{\infty\text{-QAM}}^{\log}(\zeta)$ in a neighborhood of 0. Hence, there exists $\tilde{\zeta}_2$ such that $\check{I}_{\infty\text{-QAM}}^{\log}$ is linear on the interval $[0, \tilde{\zeta}_2]$ and that $\check{I}_{\infty\text{-QAM}}^{\log}(\tilde{\zeta}_2) = I_{\infty\text{-QAM}}^{\log}(\tilde{\zeta}_2)$. By Proposition 4, $I_{\infty\text{-QAM}}^{\log}{}'$ has only a single minimum, located at $\underline{\zeta}_0$, above which $I_{\infty\text{-QAM}}^{\log}$ is convex. As is easily confirmed from inspection of Figure 7, this implies that $\tilde{\zeta}_2 > \underline{\zeta}_0$ and that $\check{I}_{\infty\text{-QAM}}^{\log}$ is given by (75), since $\check{I}_{\infty\text{-QAM}}^{\log}$ may be identical to $I_{\infty\text{-QAM}}^{\log}$ in the interval $[\tilde{\zeta}_2, \infty)$, where the latter is convex. Moreover, $\tilde{\zeta}_2 \approx 5.52$ [bits] is uniquely determined by (73) and the condition $\check{I}_{\infty\text{-QAM}}^{\log}(\tilde{\zeta}_2) = I_{\infty\text{-QAM}}^{\log}(\tilde{\zeta}_2)$.

Since

$$\frac{d}{d\zeta} \left(I_{\infty\text{-QAM}}^{\log}(\zeta) - \check{I}_{\infty\text{-QAM}}^{\log}(\zeta) \right) = I_{\infty\text{-QAM}}^{\log}{}'(\zeta) - I_{\infty\text{-QAM}}^{\log}{}'(\tilde{\zeta}_2) \tag{189}$$

for $\zeta \in [0, \tilde{\zeta}_2]$, the maximum difference between $I_{\infty\text{-QAM}}^{\log}$ and its convex envelope is obtained for $\tilde{\zeta}_m < \tilde{\zeta}_2$ which satisfies $I_{\infty\text{-QAM}}^{\log}{}'(\zeta_m) = I_{\infty\text{-QAM}}^{\log}{}'(\tilde{\zeta}_2)$. Simple numerical computation shows that $\tilde{\zeta}_m \approx 1.70$ [bits] and that

$$\tilde{\Delta}_{\infty\text{-QAM}} = I_{\infty\text{-QAM}}^{\log}\left(\tilde{\zeta}_m \right) - \check{I}_{\infty\text{-QAM}}^{\log}(\tilde{\zeta}_m) \approx 0.0608 \text{ [bit]} \tag{190}$$

REFERENCES

[1] J.G. Proakis. *Digital communications*, volume 1221. McGraw-hill, 1987.

[2] G. Forney Jr. Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference. *Information Theory, IEEE Transactions on*, 18(3):363–378, 1972.

[3] J.M. Cioffi, G.P. Dudevoir, M. Vedat Eyuboglu, and G.D. Forney Jr. MMSE decision-feedback equalizers and coding I: Equalization results. *Communications, IEEE Transactions on*, 43(10):2582–2594, 1995.

[4] Michael Tuchler, Ralf Koetter, and Andrew C Singer. Turbo equalization: principles and new results. *Communications, IEEE Transactions on*, 50(5):754–767, 2002.

[5] Taewon Hwang, Chenyang Yang, Gang Wu, Shaoqian Li, and G Ye Li. OFDM and its wireless applications: a survey. *Vehicular Technology, IEEE Transactions on*, 58(4):1673–1694, 2009.

[6] Junyi Li, Xinzhou Wu, and Rajiv Laroia. *OFDMA Mobile Broadband Communications*. Cambdige University Press, 2013.

[7] John AC Bingham. *ADSL, VDSL, and multicarrier modulation*. Wiley New York, 2000.

[8] Richard Van Nee, VK Jones, Geert Awater, Allert Van Zelst, James Gardner, and Greg Steele. The 802.11n MIMO-OFDM standard for wireless LAN and beyond. *Wireless Personal Communications*, 37(3-4):445–453, 2006.

[9] Arunabha Ghosh, David R Wolter, Jeffrey G Andrews, and Runhua Chen. Broadband wireless access with WiMax/802.16: current performance benchmarks and future potential. *Communications Magazine, IEEE*, 43(2):129–136, 2005.

[10] Ulrich Reimers. DVB-T: the COFDM-based system for terrestrial television. 1996.

[11] Amitava Ghosh, Rapeepat Ratasuk, Bishwarup Mondal, Nitin Mangalvedhe, and Tim Thomas. LTE-advanced: next-generation wireless broadband technology. *Wireless Communications, IEEE*, 17(3):10–22, 2010.

[12] Nevio Benvenuto, Rui Dinis, David Falconer, and Stefano Tomasin. Single carrier modulation with nonlinear frequency domain equalization: an idea whose time has come — again. *Proceedings of the IEEE*, 98(1):69–96, 2010.

[13] Hyung G Myung and David Goodman. *Single carrier FDMA: a new air interface for long term evolution*, volume 8. John Wiley & Sons, 2008.

[14] Eldad Perahia and Michelle X Gong. Gigabit wireless LANs: an overview of IEEE 802.11 ac and 802.11 ad. *ACM SIGMOBILE Mobile Computing and Communications Review*, 15(3):23–33, 2011.

[15] William Ryan and Shu Lin. *Channel codes: classical and modern*. Cambridge University Press, 2009.

[16] Phong S Nguyen, Arvind Yedla, Henry D Pfister, and Krishna R Narayanan. Spatially-coupled codes and threshold saturation on intersymbol-interference channels. *arXiv preprint arXiv:1107.3253*, 2011.

[17] W. Hirt and J.L. Massey. Capacity of the discrete-time gaussian channel with intersymbol interference. *Information Theory, IEEE Transactions on*, 34(3):38–38, 1988.

[18] S. Shamai and R. Laroia. The intersymbol interference channel: Lower bounds on capacity and channel precoding loss. *Information Theory, IEEE Transactions on*, 42(5):1388–1404, 1996.

[19] D. Guo, S. Shamai, and S. Verdú. Mutual information and minimum mean-square error in Gaussian channels. *Information Theory, IEEE Transactions on*, 51(4):1261–1282, 2005.

[20] Dongning Guo, Yihong Wu, Shlomo Shamai, and Sergio Verdú. Estimation in gaussian noise: Properties of the minimum mean-square error. *Information Theory, IEEE Transactions on*, 57(4):2371–2385, 2011.

[21] Miquel Payaró and Daniel P Palomar. Hessian and concavity of mutual information, differential entropy, and entropy power in linear vector gaussian channels. *Information Theory, IEEE Transactions on*, 55(8):3613–3628, 2009.

[22] Sarah Kate Wilson and John M Cioffi. A comparison of a single-carrier system using a DFE and a coded OFDM system in a broadcast Rayleigh-fading channel. In *Information Theory, 1995. Proceedings., 1995 IEEE International Symposium on*, page 335. IEEE, 1995.

[23] Andreas Czylwik. Comparison between adaptive OFDM and single carrier modulation with frequency domain equalization. In *Vehicular Technology Conference, 1997, IEEE 47th*, volume 2, pages 865–869. IEEE, 1997.

[24] Jan Tubbax, Boris Côme, Liesbet Van der Perre, Luc Deneire, Stephane Donnay, and Marc Engels. OFDM versus single carrier with cyclic prefix: a system-based comparison. In *Vehicular Technology Conference, 2001. VTC 2001 Fall. IEEE VTS 54th*, volume 2, pages 1115–1119. IEEE, 2001.

[25] Zhengdao Wang, Xiaoli Ma, and Georgios B Giannakis. OFDM or single-carrier block transmissions? *Communications, IEEE Transactions on*, 52(3):380–394, 2004.

[26] Yuan-Pei Lin and See-May Phoong. BER minimized OFDM systems with channel independent precoders. *Signal Processing, IEEE Transactions on*, 51(9):2369–2380, 2003.

[27] Amanda de Paula and Cristiano Panazio. An uncoded BER comparison between DFE-SCCP and OFDM using a convex analysis framework. In *Circuits and Systems (ISCAS), 2011 IEEE International Symposium on*, pages 2397–2400. IEEE, 2011.

[28] Robert FH Fischer and Johannes B Huber. On the equivalence of single-and multicarrier modulation: A new view. In *Information Theory. 1997. Proceedings., 1997 IEEE International Symposium on*, page 197. IEEE, 1997.

[29] N Zervos and Irving Kalet. Optimized decision feedback equalization versus optimized orthogonal frequency division multiplexing, for high-speed data transmission over the local cable network. In *Communications, 1989. ICC'89, BOSTONICC/89. Conference record.'World Prosperity Through Communications', IEEE International Conference on*, pages 1080–1085. IEEE, 1989.

[30] Nevio Benvenuto and Stefano Tomasin. On the comparison between OFDM and single carrier modulation with a DFE using a frequency-domain feedforward filter. *Communications, IEEE Transactions on*, 50(6):947–955, 2002.

[31] Jiaqi Zhang, Yukui Pei, and Ning Ge. Comparison of achievable rates of OFDM and single carrier communication systems. *Tsinghua Science and Technology*, 17(1):73–77, 2012.

[32] M Franceschini, R Pighi, G Ferrari, and R Raheli. On information theoretic aspects of single-and multi-carrier communications. In *Information Theory and Applications Workshop, 2008*, pages 94–99. IEEE, 2008.

[33] Volker Aue, Gerhard P Fettweis, and Reinaldo Valenzuela. A comparison of the performance of linearly equalized single carrier and coded OFDM over frequency selective fading channels using the random coding technique. In *Communications, 1998. ICC 98. Conference Record. 1998 IEEE International Conference on*, volume 2, pages 753–757. IEEE, 1998.

[34] Amanda de Paula and Cristiano Panazio. A comparison between OFDM and single-carrier with cyclic prefix using channel coding and frequency-selective block fading channels. *Journal of Communication and Information Systems.*, 26(1):19–29, 2011.

[35] Amanda de Paula and Cristiano Panazio. Comparison of OFDM and SC-DFE capacities without channel knowledge at the transmitter. *arXiv preprint arXiv:1306.3440*, 2013.

[36] R.M. Gray. *Entropy and information theory*. Springer Verlag, 2010.

[37] Thomas M Cover and Joy A Thomas. *Elements of Information Theory*. John Wiley & Sons, 2012.

[38] S. Jeong and J. Moon. Easily computed lower bounds on the information rate of intersymbol interference channels. *Information Theory, IEEE Transactions on*, 58(2):864–877, 2012.

[39] Y. Carmon and S. Shamai. Lower bounds and approximations for the information rate of the ISI channel. *arXiv preprint arXiv:1401.1480*, 2014.

[40] Dieter M Arnold, H-A Loeliger, Pascal O Vontobel, Aleksandar Kavcic, and Wei Zeng. Simulation-based computation of information rates for channels with memory. *Information Theory, IEEE Transactions on*, 52(8):3498–3508, 2006.

[41] H.D. Pfister, J.B. Soriaga, and P.H. Siegel. On the achievable information rates of finite state ISI channels. In *Global Telecommunications Conference, 2001. GLOBECOM'01. IEEE*, volume 5, pages 2992–2996. IEEE, 2001.

[42] A. Radosevic, D. Fertonani, T.M. Duman, J.G. Proakis, and M. Stojanovic. Bounds on the information rate for sparse channels with long memory and iud inputs. *Communications, IEEE Transactions on*, 59(12):3343–3352, 2011.

[43] Robert M Gray. *Toeplitz and circulant matrices: A review*. Now Pub, 2006.

[44] A. Lozano, A.M. Tulino, and S. Verdú. Optimum power allocation for parallel gaussian channels with arbitrary input distributions. *Information Theory, IEEE Transactions on*, 52(7):3033–3051, 2006.

[45] C. Xiao, Y.R. Zheng, and Z. Ding. Globally optimal linear precoders for finite alphabet signals over complex vector gaussian channels. *Signal Processing, IEEE Transactions on*, 59(7):3301–3314, 2011.

[46] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2009.

[47] Alex Alvarado, Fredrik Brannstrom, Erik Agrell, and Tobias Koch. High-SNR asymptotics of mutual information for discrete constellations with applications to BICM. 2013.

[48] Joel G Smith. The information capacity of amplitude-and variance-constrained sclar Gaussian channels. *Information and Control*, 18(3):203–219, 1971.

[49] V Erceg, L Schumacher, et al. TGn channel models. *IEEE 802.11-03/940r4*, 2004.