Computing Lower Bounds on the Information Rate of Intersymbol Interference Channels

Seongwook Jeong, *Student Member, IEEE* and Jaekyun Moon[†], *Fellow, IEEE*

Dept. of Electrical and Computer Engineering

University of Minnesota

Minneapolis, Minnesota 55455, U.S.A.

Email: jeong030@umn.edu

[†] Dept. of Electrical Engineering

Korea Advanced Institute of Science and Technology

Daejeon, 305-701, Republic of Korea

Email: jmoon@kaist.edu

Abstract

Provable lower bounds are presented for the information rate $I(X; X+S+N)$ where X is the symbol drawn independently and uniformly from a fixed, finite-size alphabet, S a discrete-valued random variable (RV) and N a Gaussian RV. When S represents the precursor intersymbol interference (ISI) after the minimum mean-squared error (MMSE) decision feedback equalizer (DFE) is applied at the channel output, $I(X; X + S + N)$ serves as a tight lower bound for the symmetric information rate (SIR) as well as capacity of the ISI channel corrupted by Gaussian noise. The new lower bounds are obtained by first introducing a "mismatched" mutual information function that can be proved as a lower bound to $I(X; X + S + N)$ and then further lower-bounding this function with expressions that can be computed via a few single-dimensional integrations with a small computational load. The new bounds provide a similar level of tightness as the well-known conjectured lower bound by Shamai and Laroia for a wide variety of ISI channels of practical interest.

I. INTRODUCTION

The computation of the symmetric information rate (SIR) of the classical discrete-time intersymbol interference (ISI) channel is of great interest in digital communication. The SIR represents the mutual information between the channel input and output while the input is constrained to be independently and uniformly distributed (i.u.d.) over the given alphabet. In this sense, the SIR is also known as capacity with uniform, independent input distribution and itself represents a tight lower bound to unconstrained channel capacity. During recent years, a number of researchers have worked on estimating or bounding the information rate via simulation of the Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm [1]. The information rate with a given input distribution can be closely estimated for finite ISI channels with moderate input alphabet size and channel impulse response length, by running the forward-recursion portion of the BCJR algorithm on long (pseudo) randomly generated input and noise samples [2], [3], [4]. The simulation-based method has been further generalized, and lower and upper bounds based on auxiliary ISI channels with reduced states were introduced for long ISI channels, as well as some non-finite state ISI channels in [5]. The tightness of these bounds is highly related to the optimality of auxiliary channels, but the general rule to find the optimal or near-optimal auxiliary channel has not been provided in [5]. The work of [5] has been recently extended in [6] to further tighten the lower and upper bounds by using an iterative expectation-maximization type algorithm to optimize the parameters of the auxiliary ISI channels. It is noted, however, that the global optimality of the bounds in [6] is neither guaranteed, nor the lower bound is proven to converge to a stationary point as iteration progresses. Another approach based on auxiliary ISI channels is also proposed to obtain a lower bound utilizing a mismatched Ungerboeck-type channel response to achieve improved tightness for a given level of computational complexity [7]. In the context of the work of [7], the Ungerboeck-type response is the channel's response observed at the output of the matched filter front-end. As such, the trellis search detection algorithms driven by the channel observations off the Ungerboeck model must be designed so that they can handle correlated noise samples [8].

An entirely different direction in estimating or bounding the information rate is based on finding an analytical expression that can easily be evaluated or numerically computed (in contrast to the methods based on Monte-Carlo simulation that relies on generating pseudo-random signal and noise samples). An early work in this direction is the lower bound on the SIR by Hirt [9] based on carving a fixed block out of the channel input/output sequences and performing a single multi-dimensional integration (or running Monte-Carlo simulation for estimating the integral) with the dimensionality equal to the

block size. However, this method is also computationally intense unless the size of the block gets small. Unfortunately the lower bound of [9] is not tight unless the block size is very large compared to the channel ISI length.

A number of more computationally efficient and analytically evaluated lower bounds for the SIR has been discussed in [10], [11]. Unfortunately, however, the only bound presented in [11] that is reasonably tight throughout the entire signal-to-noise ratio (SNR) region (i.e., both low and high code rate regimes) is the one that could not be proved as a lower bound. This particular bound is now widely known as the Shamai-Laroia conjecture (SLC) and, although unproven, is a popular tool for quickly estimating the SIR of ISI channels. At high code rates, the SIR is generally very close to capacity, so an easily computed tight SIR lower bound is also useful to quickly estimating channel capacity for high code rate applications, such as data storage channels and optical fiber channels. Consider the random variable (RV) $Y = X + S + N$, where X is a symbol drawn independently and uniformly from a fixed, finite-size alphabet set symmetrically positioned around the origin, S a zero-mean discrete-valued RV and N a zero-mean Gaussian RV. The SLC is concerned with the special case where S is a linear sum of symbols drawn independently and uniformly from the same symbol set where X was taken. As the number of symbols forming S grows, finding an analytical expression for the probability density function of $S + N$ (and thus one for $I(X;Y)$) is a long-standing problem [13], [14], as pointed out in [11]. The SLC of [11] can be stated as $I(X; X + S + N) \geq I(X; X + G)$, where G is a Gaussian RV with variance matching that of $S + N$. The information rate $I(X; X + G)$ is easily obtained by numerically calculating a single one-dimensional integral, and is generally observed to be reasonably tight to $I(X; X + S + N)$ in most cases. Unfortunately, $I(X; X + G)$ remains as a conjectured bound with no proof available to date. One difficulty of proving the SLC stems from the fact that for the channels driven by the inputs from a finite alphabet, Gaussian noise is not the worst-case noise in terms of the achievable information rate [11], [12]. Another difficulty is that the power contribution of a single individual weight involved in constructing S could remain a significant portion of the total power associated with all weights, even if the number of weights approaches infinity. This is to say that the Lindberg condition for the central limit theorem does not hold for this problem, and the Gaussian approximation of S cannot be justified [11].

In this paper, we are also interested in the easily computable analytical expressions for lower bounds to the SIR $I(X;Y)$. The bounds we develop here are fairly tight, with their tightness generally enhanced with increasing computational load (which in the end still remains small). Our approach is to first define a "mismatched" mutual information (MI) function based on the "mismatched" entropy that takes the log

operation not on the actual underlying probability density but on the Gaussian density with the same variance. We then prove that this "mismatched" MI is always less than or equal to the SIR $I(X;Y)$. We further bound this function from below so that the final bound can be evaluated using numerical integration. The bound is basically evaluated by computing a few single-dimensional integrals. This is in contrast to the Hirt bound that computes a single multi-dimensional integral of very high dimension. Our bound computation also requires the evaluation of sum of the absolute values of the linear coefficients that form S as well as the identification of dominant coefficient values, if they exist. In the context of the well-known minimum mean-squared error decision feedback equalizer (MMSE-DFE) filter application, S represents the collection of precursor ISI contributions at the forward filter output and the linear coefficients correspond to the weights on the interfering symbols after ideal postcursor ISI cancellation. These linear coefficients can easily be obtained with a small amount of computation. At a reasonable overall computational load, our developed bounds are shown to be as tight as the Shamai-Loria conjecture for many practical ISI channels.

Section II presents the provable bound to $I(X;Y)$ and numerically compares it with the SLC for some example distributions for the linear coefficients that form S. Section III develops upper and lower bounds on the provable bound itself, based on identifying clusters in the distribution of $S+N$. Finding clusters in the $S + N$ distribution is the same as identifying dominant coefficient values from the linear coefficient set that is used to construct S. Section IV generates and discusses numerical results. In all finite-ISI channels examined, our bound provides the same level of tightness as the SLC to the SIR (while being actually tighter than SLC at high SNRs when viewed closed up) with a very reasonable computation load. In particular, our lower bound is presented on the same channel employed in [6]. This provides an indirect means to compare the computational loads of the two methods. As expected, our analytical method is much better in quickly producing a tight bound than the simulation-based method of [6] in terms of complexity/accuracy tradeoffs. Note that the method of [6] represents a latest development in simulation-based SIR bounds. Section V concludes the paper.

II. A PROVABLE LOWER BOUND TO THE SYMMETRICAL INFORMATION RATE

We first present a provable lower bound to $I(X;Y)$ where $Y = X + \sum_{k=1}^{L} d_{-k}X_k + N = X + S + N$. The symbols X and X_k are all independently and uniformly drawn. The linear coefficients d_{-k}'s are related to the channel impulse response and will be specified in Section IV. Let $V = S + N$ so we can write $Y = X + V$. Note that V is a Gaussian mixture. Also let $Z = X + G$ where G is a zero mean Gaussian with variance matching that of V, i.e., $\sigma_G^2 = \sigma_V^2$.

Definition 1 (“Mismatched” Mutual Information (MMI) Function): Define

$$I'(X;Y) \triangleq H'(Y) - H'(V) \tag{1}$$

where

$$H'(Y) \triangleq -\int_{-\infty}^{\infty} f_Y(t) \log f_Z(t) dt, \quad H'(V) \triangleq -\int_{-\infty}^{\infty} f_V(t) \log f_G(t) dt$$

and $f_Y(t)$, $f_V(t)$, $f_Z(t)$, and $f_G(t)$ are the probability density functions (pdfs) of the RVs, Y, V, Z, and G, respectively. Note that the “mismatched” entropy functions $H'(Y)$ and $H'(V)$ are defined based the log operation applied not to the actual underlying pdf $f_V(t)$ but rather to the “mismatched” Gaussian pdf $f_G(t)$.

Lemma 1: Given the MMI function defined as above, we have

$$I'(X;Y) \leq I(X;Y). \tag{2}$$

Proof: See Appendix A. ∎

Let us now take a close look at this MMI function $I'(X;Y)$ and develop some insights into its behaviour. Let the variances of V, S, and N be σ_V^2, σ_S^2, and σ_N^2 respectively. Further assume that the RVs, X, V, S, and N are all real-valued. We will also assume a binary input alphabet. These assumptions are not necessary for our development but make the presentation clearer as well as less cluttered. We will simply state the results in Section III.C for a non-binary/complex-valued example. We also denote $m_i = \sum_{k=1}^{L} d_{-k} X_k$ for $i = 1, 2, \ldots, 2^L$ since $\{X_k\}_{k=1}^{L}$ can have 2^L different sequences. Naturally, the pdfs of RVs V and G can be written as

$$f_V(t) = 2^{-L} \sum_{i=1}^{2^L} \frac{1}{\sqrt{2\pi\sigma_N^2}} \exp\left(-\frac{(t-m_i)^2}{2\sigma_N^2}\right)$$

$$f_G(t) = \frac{1}{\sqrt{2\pi\sigma_V^2}} \exp\left(-\frac{t^2}{2\sigma_V^2}\right).$$

Proposition 1: Denoting $\rho_i \triangleq m_i/\sqrt{P_X}$, letting ρ_k^+’s to mean the positive-half subset of ρ_i’s, and defining $R \triangleq P_X/\sigma_V^2$ and $\phi \triangleq \sigma_N/\sigma_V$, the MMI function can be rewritten as $I'(X;Y) = \log 2 - F$

with the new definition

$$
F \triangleq 2^{-L} \sum_{i=1}^{2^L} \mathrm{E}_\tau \left[\log \left\{ 1 + e^{-2R\rho_i} e^{-2\phi\sqrt{R}\tau - 2R} \right\} \right]
$$

$$
= \mathrm{E}_{\rho,\tau} \left[\log \left\{ 1 + e^{-2R\rho} e^{-2\phi\sqrt{R}\tau - 2R} \right\} \right] \tag{3a}
$$

$$
= 2^{-(L-1)} \sum_{k=1}^{2^{L-1}} \mathrm{E}_\tau \left[\frac{1}{2} \log \left\{ 1 + 2 \cosh \left(2R\rho_k^+ \right) e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4\phi\sqrt{R}\tau - 4R} \right\} \right]
$$

$$
= \mathrm{E}_{\rho^+,\tau} \left[\frac{1}{2} \log \left\{ 1 + 2 \cosh \left(2R\rho^+ \right) e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4\phi\sqrt{R}\tau - 4R} \right\} \right]. \tag{3b}
$$

The detail derivation is given in Appendix B. The position m_i of the ith Gaussian pdf of the mixture $f_V(t)$ is expressed as a dimensionless quantity: $\rho_i = m_i/\sqrt{P_X}$, with the normalization by the square root of the input power. Because of the symmetric nature of $f_V(t)$, ρ_i occurs in equal-magnitude, opposite-polarity pairs. The expectation is initially over the τ variable, which is considered a zero-mean unit-variance Gaussian random variable when contained inside the argument of the expectation operator. The expectation operator in this case can simply be viewed as a short-hand notation as in

$$
\mathrm{E}_\tau \left[p(\tau) \right] = \int_{-\infty}^{\infty} \frac{e^{-\tau^2/2}}{\sqrt{2\pi}} p(\tau) d\tau.
$$

In (3a) and (3b), however, ρ (or ρ^+) is also treated as a RV and the expectation is over both τ and ρ (or τ and ρ^+) as the double subscripts indicate. Given the pdfs of τ, ρ, and ρ^+, the computation of the expectation now involves numerical evaluation of a double integral. Note that in (3a) ρ is a discrete-valued random variable distributed according to $f_\rho(t)$, which denotes the probability distribution of $\rho = (1/\sqrt{P_X}) \sum_{k=1}^{L} d_{-k} X_k$ and ρ^+ is a discrete-valued random variable distributed according to $2f_\rho(t)u(t)$ where $u(t)$ is a step function. Also, notice that $\cosh(2R\rho^+) \geq 1$ and $\phi \leq 1$.

It is insightful to compare F with

$$
F_{SLC} \triangleq \log 2 - C_{SLC}(R)
$$

$$
= \int_{-\infty}^{\infty} \frac{e^{-\tau^2/2}}{\sqrt{2\pi}} \log \left\{ 1 + e^{-2\sqrt{R}\tau - 2R} \right\} d\tau = \mathrm{E}_\tau \left[\log \left\{ 1 + e^{-2\sqrt{R}\tau - 2R} \right\} \right] \tag{4a}
$$

$$
= \mathrm{E}_\tau \left[\frac{1}{2} \log \left\{ 1 + 2e^{-2\sqrt{R}\tau - 2R} + e^{-4\sqrt{R}\tau - 4R} \right\} \right] \tag{4b}
$$

where $C_{SLC}(R)$ is the SIR of the binary-input Gaussian channel with SNR given by $R \triangleq P_X/\sigma_V^2$ and is the well-known SLC. The function F_{SLC} quantifies the gap between the SLC and the maximum attainable capacity for any binary channel with no constraint on SNR, namely, 1 bit/channel use. Comparing the expressions for F in (3b) and F_{SLC} in (4b), we see that if $\rho^+ = 0$ so that $\phi = 1$, then $F = F_{SLC}$, and

$I'(X; Y)$ and the SLC both become equal to $I(X; Y)$. Also, if the discrete RV ρ converges to a Gaussian random variable (in cumulative distribution), then again we get $F = F_{SLC}$ and $I'(X; Y) = C_{SLC}(R) = I(X; Y)$.

Furthermore, that $\rho^+ \geq 0$ in (3b) makes F larger while the factor ϕ being less than 1 has an effect of decreasing F as it increases. If $I'(X; Y) = \log 2 - F$ is to be a tight lower bound to $I(X; Y)$, then F needs to be small. The important question is: how does F overall compare with F_{SLC}, over all interested range of SNR? Since it is already proved that $I'(X; Y) = \log 2 - F$, if $F \leq F_{SLC}$ for some R values, then clearly $C_{SLC}(R) = \log 2 - F_{SLC} \leq I(X; Y)$ at those SNRs, i.e., the SLC holds true at least at these SNRs.

While exact computation of (3b) requires in general obtaining all possible positive-side values of $\rho = (1/\sqrt{P_X}) \sum_{k=1}^{L} d_{-k} X_k$ and thus can be computationally intense for large L, in the cases where we know the functional form of the distribution for ρ, evaluation of (3a) or (3b) is easy; the behaviour of F under different ρ distributions offer useful insights.

First try a uniform distribution for ρ. For a uniformly distributed discrete random variable ρ from $-K\Delta = -|\rho|_{\max}$ to $K\Delta = |\rho|_{\max}$ with a gap Δ between delta functions in the pdf, we have

$$\sigma_S^2 = \frac{2P_X\Delta^2}{2K+1} \sum_{i=1}^{K} i^2 = \frac{P_X\Delta^2 K(K+1)}{3} = \frac{P_X|\rho|_{\max}(|\rho|_{\max} + \Delta)}{3}$$

which makes

$$\phi^2 = \frac{\sigma_N^2}{\sigma_N^2 + \sigma_S^2} = 1 - \frac{\sigma_S^2}{\sigma_V^2} = 1 - \frac{R\Delta^2 K(K+1)}{3} = 1 - \frac{R|\rho|_{\max}(|\rho|_{\max} + \Delta)}{3}.$$

Fig. 1 shows F and F_{SLC} plotted with $K = 1000$ as functions of R for various values of ϕ.

We also consider a simple case involving only a single coefficient d_{-1}, in which case ρ takes only two possible values, e.g., $\rho = \pm\sqrt{(1 - \phi^2)/R}$. The plots of F and F_{SLC} for this case are shown against R for different values of ϕ in Fig. 2. Figs. 1 and 2 point to similar behaviours of F versus F_{SLC}. Namely, F becomes smaller than F_{SLC} as ϕ increases for a range of R values. At these R values, the provable lower bound $I'(X; Y)$ is apparently tighter than the SLC, with respect to the SIR.

III. BOUNDING F

Exact computation of F in general is not easy, especially when L goes to infinity. We thus resort to bounding F with expressions that can easily be computed. An upper bound on F will provide a lower bound on $I'(X; Y)$ and thus on $I(X; Y)$. Lower bounds on F are also derived to see if they can get smaller than F_{SLC}. If so, this would mean $I'(X; Y) = \log 2 - F$ is larger than $C_{SLC}(R) = \log 2 - F_{SLC}$, i.e., our bound is tighter than the SLC.

A. Simple Bounds

Since the function $\log\left(1 + 2\cosh(2R\rho^+)e^{-2\phi\sqrt{R}\tau} + e^{-4\phi\sqrt{R}\tau-4R}\right)$ is convex in ρ^+, the function $\mathrm{E}_\tau\left[\frac{1}{2}\log\left(1 + 2\cosh(2R\rho^+)e^{-2\phi\sqrt{R}\tau} + e^{-4\phi\sqrt{R}\tau-4R}\right)\right]$ is also convex in ρ^+. Moreover, the function $\mathrm{E}_\tau\left[\frac{1}{2}\log\left(1 + 2\cosh(2R\rho^+)e^{-2\phi\sqrt{R}\tau} + e^{-4\phi\sqrt{R}\tau-4R}\right)\right]$ increases as ρ^+ increases. Accordingly, we can develop bounds on F. The first simple upper bound is

$$F^{u1} \triangleq T\left(|\rho|_{\max}, \theta\right)\Big|_{\theta=\sigma_\rho} \tag{5}$$

where, for a given $|\rho|_{\max}$, the function $T\left(|\rho|_{\max}, \theta\right)$ represents a straight line passing through two points of the function $\mathrm{E}_\tau\left[\frac{1}{2}\log\left(1 + 2\cosh(2R\theta)e^{-2\phi\sqrt{R}\tau} + e^{-4\phi\sqrt{R}\tau-4R}\right)\right]$ at $\theta = 0$ and at $\theta = |\rho|_{\max}$. Note that $|\rho|_{\max} \triangleq \arg\max_i |\rho_i| = \sum_{k=1}^{L}|d_{-k}|$ and σ_ρ is the standard deviation of RV ρ.

Similarly, $\mathrm{E}_\tau\left[\frac{1}{2}\log\left(1 + 2\alpha e^{-2\phi\sqrt{R}\tau} + e^{-4\phi\sqrt{R}\tau-4R}\right)\right]$ is a concave and increasing function of $\alpha \triangleq \cosh(2R\rho^+)$. Based on this property, we can develop another upper bound.

$$F^{u2} \triangleq \mathrm{E}_\tau\left[\frac{1}{2}\log\left\{1 + 2(s\sigma_\rho + 1)e^{-2\phi\sqrt{R}\tau-2R} + e^{-4\phi\sqrt{R}\tau-4R}\right\}\right] \tag{6}$$

where $s = \left(\cosh(2R|\rho|_{\max}) - 1\right)/|\rho|_{\max}$, the slope of a straight line connecting two points $(0, 1)$ and $(|\rho|_{\max}, \cosh(2R|\rho|_{\max}))$.

A lower bound on F can also be obtained that can help shed lights on how tight the upper bounds on F are. Using the convexity of $\mathrm{E}_\tau\left[\log\left(1 + e^{-2R\rho}e^{-2\phi\sqrt{R}\tau-2R}\right)\right]$ in ρ, the simple lower bound of F is

$$F^l \triangleq \mathrm{E}_\tau\left[\frac{1}{2}\log\left\{1 + 2e^{-2\phi\sqrt{R}\tau-2R} + e^{-4\phi\sqrt{R}\tau-4R}\right\}\right]. \tag{7}$$

The detailed derivations of (5), (6) and (7) are given in Appendix C.

B. Tightened Bounds Based on Cluster Identification

The above bounds can be tightened up by identifying clusters in the Gaussian mixture $f_V(t)$. In practical ISI channels, $f_V(t)$ often consists of clusters. This is due to the fact that the coefficient set d_{-k}'s typically contains a few dominating coefficients plus many small terms. Assuming there are M dominating coefficients among d_{-k}'s, we can let $\rho_k = \lambda_j + \mu_i$ where $j = 1, 2, \ldots, 2^M$, $i = 1, 2, \ldots, 2^{L-M}$, and $k = (j-1)2^{L-M} + i$. Since X_k is an i.u.d. RV, λ and μ are independent so that $\sigma_\rho^2 = \sigma_\lambda^2 + \sigma_\mu^2$ where σ_λ^2 and σ_μ^2 denote the variance of RVs λ and μ, respectively. Notice that λ_j can be viewed as the position of a specific cluster while μ_i points to a specific Gaussian pdf out of 2^{L-M} Gaussian pdf's symmetrically positioned around λ_j.

Therefore, assuming there are 2^M clusters of Gaussian pdfs, the lower bound F^{u1} can be tightened as

$$F_M^{u1} \triangleq 2^{-M} \sum_{j=1}^{2^M} T_j \left(|\mu|_{\max}, \theta\right)\Big|_{\theta = \sigma_\mu} \tag{8}$$

where, for a given $|\mu|_{\max}$, the function $T_j(|\mu|_{\max}, \theta)$ is a straight line that passes through the two points of the convex function $\mathrm{E}_\tau \left[\frac{1}{2} \log \left\{1 + 2\cosh(2R\theta) e^{-2R\lambda_j} e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4R\lambda_j} e^{-4\phi\sqrt{R}\tau - 4R}\right\}\right]$ at $\theta = 0$ and $\theta = |\mu|_{\max}$, σ_μ is the standard deviation of RV μ defined as $\sigma_\mu = \sqrt{\sigma_\rho^2 - \sigma_\lambda^2}$, and $|\mu|_{\max} = |\rho|_{\max} - |\lambda|_{\max}$.

Another form of tightened upper bound based on F^{u2} is obtained as

$$F_M^{u2} \triangleq 2^{-M} \sum_{j=1}^{2^M} \mathrm{E}_\tau \left[\frac{1}{2} \log \left\{1 + 2\left(s_M \sigma_\mu + 1\right) e^{-2R\lambda_j} e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4R\lambda_j} e^{-4\phi\sqrt{R}\tau - 4R}\right\}\right] \tag{9}$$

where $s_M = \left(\cosh(2R|\mu|_{\max}) - 1\right)/|\mu|_{\max}$.

The lower bound F^l can also be tightened similarly based on the cluster identification:

$$F_M^l \triangleq 2^{-(M-1)} \sum_{k=1}^{2^{M-1}} \mathrm{E}_\tau \left[\frac{1}{2} \log \left\{1 + 2\cosh\left(2R\lambda_k^+\right) e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4\phi\sqrt{R}\tau - 4R}\right\}\right] \tag{10}$$

where λ_k^+'s form the positive-half subset of λ_j's. The detail derivations of (8), (9) and (10) can be found in Appendix D.

C. Bounds for Complex Channels with the Quaternary Alphabet Inputs

In the previous subsections, ISI coefficients and noise samples are assumed to be real-valued with the channel inputs being the binary phase shift keying (BPSK) signal. In this subsection, we provide a complex-valued example along with the channel inputs taken from a quadrature phase shift keying (QPSK) quaternary alphabet, i.e., $X_k = \sqrt{\frac{P_X}{2}}(\pm 1 \pm j)$. The extension to larger alphabets should be straightforward.

Denoting the real and imaginary parts of complex number a are denoted by $a^{(r)}$ and $a^{(i)}$ respectively, i.e., $a = a^{(r)} + ja^{(i)}$ and $m_i = \sum_{k=1}^L d_{-k} X_k$ for $i = 1, 2, \ldots, 4^L$, the pdf's of complex random variables V and G are given as

$$f_V(t) = 4^{-L} \sum_{i=1}^{4^L} \frac{1}{\pi\sigma_N^2} \exp\left(-\frac{|t - m_i|^2}{\sigma_N^2}\right)$$

$$= 4^{-L} \sum_{i=1}^{4^L} \left\{\frac{1}{\sqrt{\pi\sigma_N^2}} \exp\left(-\frac{\left(t^{(r)} - m_i^{(r)}\right)^2}{\sigma_N^2}\right) \frac{1}{\sqrt{\pi\sigma_N^2}} \exp\left(-\frac{\left(t^{(i)} - m_i^{(i)}\right)^2}{\sigma_N^2}\right)\right\}$$

$$f_G(t) = \frac{1}{\pi\sigma_V^2} \exp\left(-\frac{|t|^2}{\sigma_V^2}\right) = \frac{1}{\sqrt{\pi\sigma_V^2}} \exp\left(-\frac{\left(t^{(r)}\right)^2}{\sigma_V^2}\right) \frac{1}{\sqrt{\pi\sigma_V^2}} \exp\left(-\frac{\left(t^{(i)}\right)^2}{\sigma_V^2}\right).$$

Then, for the SLC, we write

$$
\begin{aligned}
F_{SLC} &\triangleq \log 4 - C_{SLC}(R) \\
&= 2 \int_{-\infty}^{\infty} \frac{e^{-\tau^2}}{\sqrt{\pi}} \log\left\{1 + e^{-2\sqrt{2R}\tau - 2R}\right\} d\tau \\
&= 2\mathrm{E}_{\tau}\left[\log\left\{1 + e^{-2\sqrt{2R}\tau - 2R}\right\}\right].
\end{aligned}
\tag{11}
$$

The proposed lower bounds can be derived in a similar way as

$$
\begin{aligned}
F_M^{u1} &= 4^{-M} \sum_{j=1}^{4^M} \left\{ T_j^{(r)}\left(|\mu^{(r)}|_{\max}, \theta\right)\Big|_{\theta=\frac{\sigma_\mu}{\sqrt{2}}} + T_j^{(i)}\left(|\mu^{(i)}|_{\max}, \theta\right)\Big|_{\theta=\frac{\sigma_\mu}{\sqrt{2}}} \right\} \\
&= 4^{-M} \sum_{j=1}^{4^M} 2 T_j^{(r)}\left(|\mu^{(r)}|_{\max}, \theta\right)\Big|_{\theta=\frac{\sigma_\mu}{\sqrt{2}}}
\end{aligned}
\tag{12}
$$

where, for a given $|\mu^{(r)}|_{\max}$ and $|\mu^{(i)}|_{\max}$, $T_j^{(r)}(|\mu^{(r)}|_{\max}, \theta)$ denotes a straight line that passes through two points of $\mathrm{E}_{\tau}\left[\frac{1}{2}\log\left\{1 + 2\cosh\left(2\sqrt{2}R\theta\right)e^{-2\sqrt{2}R\lambda_j^{(r)}}e^{-2\phi\sqrt{2R}\tau - 2R} + e^{-4\sqrt{2}R\lambda_j^{(r)}}e^{-4\phi\sqrt{2R}\tau - 4R}\right\}\right]$ at $\theta = 0$ and at $\theta = |\mu^{(r)}|_{\max} = |\mu|_{\max}/\sqrt{2}$ while $T_j^{(i)}(|\mu^{(i)}|_{\max}, \theta)$ is a similar straight line passing through $\mathrm{E}_{\tau}\left[\frac{1}{2}\log\left\{1 + 2\cosh\left(2\sqrt{2}R\theta\right)e^{-2\sqrt{2}R\lambda_j^{(i)}}e^{-2\phi\sqrt{2R}\tau - 2R} + e^{-4\sqrt{2}R\lambda_j^{(i)}}e^{-4\phi\sqrt{2R}\tau - 4R}\right\}\right]$ at $\theta = 0$ and at $\theta = |\mu^{(i)}|_{\max} = |\mu|_{\max}/\sqrt{2}$. The equality in (12) holds because the pdfs of $\lambda^{(r)}$ and $\lambda^{(i)}$ (or $T_j^{(r)}(|\mu^{(r)}|_{\max}, \theta)$ and $T_j^{(i)}(|\mu^{(i)}|_{\max}, \theta)$) are identical. Also, note that the variance of $\mu^{(r)}$ and $\mu^{(i)}$ are equal to $\sigma_\mu^2/2$.

Another form of the bound is given as

$$
\begin{aligned}
F_M^{u2} &\triangleq 4^{-M} \sum_{j=1}^{4^M} \Bigg(\mathrm{E}_{\tau}\left[\frac{1}{2}\log\left\{1 + 2\left(\frac{s_M^{(r)}\sigma_\mu}{\sqrt{2}} + 1\right)e^{-2\sqrt{2}R\lambda_j^{(r)}}e^{-2\phi\sqrt{2R}\tau - 2R} + e^{-4\sqrt{2}R\lambda_j^{(r)}}e^{-4\phi\sqrt{2R}\tau - 4R}\right\}\right] \\
&\quad + \mathrm{E}_{\tau}\left[\frac{1}{2}\log\left\{1 + 2\left(\frac{s_M^{(i)}\sigma_\mu}{\sqrt{2}} + 1\right)e^{-2\sqrt{2}R\lambda_j^{(i)}}e^{-2\phi\sqrt{2R}\tau - 2R} + e^{-4\sqrt{2}R\lambda_j^{(i)}}e^{-4\phi\sqrt{2R}\tau - 4R}\right\}\right] \Bigg) \\
&= 4^{-M} \sum_{j=1}^{4^M} 2\mathrm{E}_{\tau}\left[\frac{1}{2}\log\left\{1 + 2\left(\frac{s_M^{(r)}\sigma_\mu}{\sqrt{2}} + 1\right)e^{-2\sqrt{2}R\lambda_j^{(r)}}e^{-2\phi\sqrt{2R}\tau - 2R} + e^{-4\sqrt{2}R\lambda_j^{(r)}}e^{-4\phi\sqrt{2R}\tau - 4R}\right\}\right]
\end{aligned}
\tag{13}
$$

where $s_M^{(r)} = \left(\cosh(2\sqrt{2}R|\mu^{(r)}|_{\max}) - 1\right)/|\mu^{(r)}|_{\max}$ and $s_M^{(i)} = \left(\cosh(2\sqrt{2}R|\mu^{(i)}|_{\max}) - 1\right)/|\mu^{(i)}|_{\max}$. The equality of (13) holds because $s_M^{(r)} = s_M^{(i)}$ from $|\mu^{(r)}|_{\max} = |\mu^{(i)}|_{\max} = |\mu|_{\max}/\sqrt{2}$ and $\lambda^{(r)} = \lambda^{(i)}$.

Finally, a lower bound to F can be shown to be

$$F_M^l \triangleq 4^{-M}/2 \sum_{j=1}^{4^M/2} \left(\mathrm{E}_\tau \left[\frac{1}{2} \log \left\{ 1 + 2 \cosh \left(2\sqrt{2} R \lambda_j^{(r)+} \right) e^{-2\phi\sqrt{2R}\tau - 2R} + e^{-4\phi\sqrt{2R}\tau - 4R} \right\} \right] \right.$$

$$\left. + \mathrm{E}_\tau \left[\frac{1}{2} \log \left\{ 1 + 2 \cosh \left(2\sqrt{2} R \lambda_j^{(i)+} \right) e^{-2\phi\sqrt{2R}\tau - 2R} + e^{-4\phi\sqrt{2R}\tau - 4R} \right\} \right] \right)$$

$$= 4^{-M} \sum_{j=1}^{4^M/2} \mathrm{E}_\tau \left[\frac{1}{2} \log \left\{ 1 + 2 \cosh \left(2\sqrt{2} R \lambda_j^{(r)+} \right) e^{-2\phi\sqrt{2R}\tau - 2R} + e^{-4\phi\sqrt{2R}\tau - 4R} \right\} \right] \quad (14)$$

where $\lambda_k^{(r)+}$'s and $\lambda_k^{(i)+}$'s form the positive-half subset of $\lambda_j^{(r)}$'s and $\lambda_j^{(i)}$'s correspondingly.

IV. APPLICATION TO ISI CHANNELS AND NUMERICAL EXAMPLES

A. The ISI Channel and MMSE-DFE

Fig. 3 shows the discrete-time equivalent system model of the finite-ISI channel with the infinite-length feedforward filter of the unbiased MMSE-DFE preceded by the matched filter (MF) for the channel. The discrete-time MF output of Fig. 3 is identical to the baud-rate sampled output of the continuous-time MF applied to the continuous-time channel, under the assumption that the channel is strictly limited to the Nyquist band.

We also assume assumed that the receiver knows the D-transform of the finite-ISI channel response, $h(D)$, x_k is an i.u.d. input sequence, and w_k is additive white Gaussian noise (AWGN) with variance $\sigma_W^2 = N_0$. Furthermore, r_k is the channel output sequence and z_k is the output sequence of the infinite-length MMSE-DFE feedforward filter.

Denoting $X = x_0$, $X_k = x_k$, and $Y = y_0$, the output of the the unbiased MMSE-DFE with ideal feedback [15] is given by

$$Y = X + \sum_{k=1}^{\infty} d_{-k} X_k + N = X + S + N = X + V$$

where N is the Gaussian noise sample observed at the DFE forward filter output and $d_{-k}X_k$ is the precursor ISI sequence. Note we are assuming stationary random processes. It is well-known that the D-transform of the precursor ISI taps d_{-k} is given by [15]

$$d(D) = \frac{N_0}{P_0 - N_0} \left(1 - \frac{1}{g^*(D^{-*})} \right) \quad (15)$$

where P_0 is such that $\log P_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log R_{ss}(e^{-j\theta}) d\theta$ and $g^*(D^{-*})$ is obtained from spectral factorization: $R_{ss}(D) = P_X R_{hh}(D) + N_0 = P_0 g(D) g^*(D^{-*})$ with $R_{hh}(D) = h(D) h^*(D^{-*})$. Note that a

convenient numerical spectral factorization algorithm exists for recursively computing the coefficients of $g^*(D^{-*})$ [16], [17].

Accordingly, the variances of V, N, and S are given as

$$\sigma_V^2 = \frac{P_X N_0}{P_0 - N_0}$$

$$\sigma_N^2 = \frac{P_X P_0 N_0}{2\pi (P_0 - N_0)^2} \int_{-\pi}^{\pi} \frac{R_{hh}(e^{-j\theta})}{R_{hh}(e^{-j\theta}) + N_0/P_X} d\theta$$

$$\sigma_S^2 = \sigma_V^2 - \sigma_N^2.$$

We can obtain $|\rho|_{\max}$ by the absolute summation of the inverse D-transform of $d(D)$ if the feedforward filter of MMSE-DFE is stable, i.e., $\sum_{k=1}^{\infty} |d_{-k}| < \infty$. Let us first consider the case when $d(D)$ has P multiple first-order poles, p_j for $j = 1, 2, \ldots, P$, then, $|\rho|_{\max}$ can be obtained by the partial fraction method since $d(D)$ is a rational function. In other words, the inverse D-transform of individual fraction terms can be found and then added together to form d_{-k}. Denoting $a(D) = \frac{1}{g^*(D^{-*})} = \sum_{j=1}^{P} \frac{c_j}{1 - p_j D^{-1}}$, the sequence a_{-k} is given as $a_{-k} = \sum_{j=1}^{P} c_j p_j^k$. Therefore,

$$
\begin{aligned}
|\rho|_{\max} &= \frac{1}{\sqrt{P_X}} \sum_{k=1}^{\infty} |d_{-k} X_k| = \sum_{k=1}^{\infty} |d_{-k}| = \frac{N_0}{(P_0 - N_0)} \left(\sum_{k=1}^{\infty} |a_{-k}| \right) \\
&= \frac{N_0}{(P_0 - N_0)} \left(\sum_{k=1}^{\infty} \left| \sum_{j=1}^{P} c_j p_j^k \right| \right) \\
&\leq \frac{N_0}{(P_0 - N_0)} \left(\sum_{j=1}^{P} \sum_{k=1}^{\infty} \left| c_j p_j^k \right| \right) \\
&= \frac{N_0}{(P_0 - N_0)} \left(\sum_{j=1}^{P} \frac{|c_j p_j|}{1 - |p_j|} \right).
\end{aligned}
\tag{16}
$$

The upper bound of $|\rho|_{\max}$ can be also tightened by identifying the first K dominant taps:

$$
\begin{aligned}
|\rho|_{\max} &= \frac{N_0}{(P_0 - N_0)} \left(\sum_{k=1}^{\infty} \left| \sum_{j=1}^{P} c_j p_j^k \right| \right) \\
&= \frac{N_0}{(P_0 - N_0)} \left(\sum_{k=1}^{K} \left| \sum_{j=1}^{P} c_j p_j^k \right| + \sum_{k=K+1}^{\infty} \left| \sum_{j=1}^{P} c_j p_j^k \right| \right) \\
&\leq \frac{N_0}{(P_0 - N_0)} \left(\sum_{k=1}^{K} \left| \sum_{j=1}^{P} c_j p_j^k \right| + \sum_{j=1}^{P} \sum_{k=K+1}^{\infty} \left| c_j p_j^k \right| \right) \\
&= \sum_{k=1}^{K} |d_{-k}| + \frac{N_0}{(P_0 - N_0)} \left(\sum_{j=1}^{P} \frac{\left| c_j p_j^{K+1} \right|}{1 - |p_j|} \right).
\end{aligned}
\tag{17}
$$

For the case of the multiple-order poles of $d(D)$, the upper bound of $|\rho|_{\max}$ can be also obtained in a similar way using the triangle inequality $|a + b| \leq |a| + |b|$.

The channel capacity C (bits/channel use) for any finite-ISI channel corrupted by Gaussian noise is given [18] as

$$C \triangleq \lim_{N \to \infty} \frac{1}{2N + 1} I\left(\{x_k\}_{-N}^N; \{r_k\}_{-N}^N\right)$$

$$\geq \lim_{N \to \infty} \frac{1}{2N + 1} I\left(\{x_k\}_{-N}^N; \{z_k\}_{-N}^N\right) \tag{18}$$

$$\geq I(X; Y) \tag{19}$$

where $\{u_k\}_{N_1}^{N_2} = \{u_k, k = N_1, N_1 + 1, \ldots, N_2\}$. The inequality in (18) holds due to the data processing theorem (equality holds if the MMSE-DFE feedforward filter is invertible [11]). The inequality of (19) can be obtained by applying the chain rule of mutual information and assuming stationarity [11].

B. Numerical Results

Now, let us examine the particular ISI channels, $h(D) = 2^{-1/2}(1 + D)$, $h(D) = 2^{-1}(1 + D - D^2 - D^3)$, and $h(D) = 0.19 + 0.35D + 0.46D^2 + 0.5D^3 + 0.46D^4 + 0.35D^5 + 0.19D^6$, which are well-known and previously investigated in e.g. [2], [10], [11], and $h(D) = 1.6099^{-1/2} \sum_{i=0}^{10} D^i / (1 + (i - 5)^2)$, which was considered in [6]. The first 20 precursor ISI tap values are computed and shown in Fig. 4 for these example channels. In addition we consider a complex-valued partial response channel: $h(D) = 2^{-1}\{(1 + j) + (1 - j)D\}$. The channel inputs are binary, except the complex-valued channel for which the inputs are assumed quaternary.

Since the infinite-length MMSE-DFE is used, i.e., $L = \infty$ in general, the probability distribution of ρ is not available. Hence the lower bounds $C_{L1,M} = \log 2 - F_M^{u1}$ and $C_{L2,M} = \log 2 - F_M^{u2}$ along with $C_{SLC} = \log 2 - F_{SLC}$ are considered as functions of $\text{SNR} = P_X / N_0$ for different values of M. When no clustering is used, we set $M = 0$. In computing $|\rho|_{\max}$ (and thus $|\mu|_{\max}$) needed to calculate F_M^{u1} or F_M^{u2}, we were able to run numerical recursive spectral factorization to find all non-negligible d_{-k} coefficients relatively quickly for all channels considered, without resorting to the bounds of (16) or (17). We observed that the lower bounds, $C_{L1,M}$ and $C_{L2,M}$, produced similar results, so only $C_{L1,M}$ were chosen and plotted as $C_{L,M}$ through Fig. 5 - Fig. 9. The SIR of each channel is also obtained using the simulation-based approach [2], [3], [4].

For each capacity figure, we first plotted the SIR and C_{SLC}. We then plotted C_L for $M = 0$ and then another C_L by choosing an M value for which the C_L bound is almost as tight as the C_{SLC} conjecture

(this is why the C_{SLC} curve is almost overwritten and indistinguishable in some figures). We also show for each channel how the upper and lower bound of F close on each other as M increases. The bounds on F are shown with F_{SLC} subtracted from them. In this way, it should be clear that for those SNR values where $F^u - F_{SLC}$ becomes less than zero eventually, F is less than F_{SLC}, guaranteeing that $I'(X : Y) = \log 2 - F$ is tighter than C_{SLC}. In fact, it can be seen from the figures that this is true for the high SNR range corresponding to all rates higher than roughly 0.6 in all channel considered. An obvious by-product of this observation is the assurance that the SLC surely holds true at this SNR range. The curves of $F^l - F_{SLC}$ for different M values also provide a detailed picture of how large M should be in order for C_L to get close enough to the C_{SLC}.

Note that the computational load for evaluating the integral of (8) and (9) to obtain the bound depends exponentially on M, the number of clusters in the pdf $f_V(t)$. The computational load in computing the dominant precursor ISI taps and their magnitude sum is minimal. The results summarized in the figures indicate that in each channel considered, a relatively small value of M (and thus a reasonably low computational load) yields a bound as tight as the SLC. As a case in point, comparison of Fig. 8 with the results of [6] (Figure 6 of [6], specifically) gives a good idea on the usefulness of an easily computable bound such as the one presented here. At a rate 0.9, for example, one can observe from a close examination of Fig. 6 of [6] that the lower bound of [6] approaches the SIR within about 0.88 dB with 2 iterations, which would require basically running the BCJR algorithm twice on a reduced channel trellis of 64 states. In contrast, our bound based on just two clusters is about 0.84 dB away from the SIR at the same rate, as estimated from Fig. 8. This bound requires computation of $2^2 = 4$ single-dimensional integrals, the required complexity of which amounts to virtually nothing relative to that required running two BCJR simulation runs in the method of [6]. The simulation-based bound of [6] does narrow the gap to about 0.65 dB with five iterations, but at the expense of much more computational time.

We finally remark that the value of the simulation-based SIR estimation methods is not in their ability to provide easily obtained bounds; rather they play a critical role in estimating the SIR (or capacity) with a very high accuracy, given the ample computational resources. As for providing convenient and easily computed SIR estimates or bounds, the need for analytically evaluated bounds such as the one developed in this paper remains high.

V. CONCLUSION

In this paper, we derived a lower bound to the SIR of the ISI channel driven by discrete and finite-amplitude inputs. The approach taken was to introduce a "mismatched" mutual information function that

acts as a lower bound to the symmetric information rate between the channel input and the ideal-feedback MMSE-DFE filter output. This function turns out to be tighter than the Shamai-Laroria conjecture for a practically significant range of SNR values for some example channels. We then further lower-bounded this function by another function that can be evaluated via numerical integration with a small computational load. The final computation also requires finding a few large precursor ISI tap values as well as the absolute sum of the remaining ISI terms, which can be done easily. The final lower bounds are demonstrated for a number of well-known finite-ISI channels, and the results indicate that the new bounds computed at a fairly low computational load are as tight as the SLC.

APPENDIX A

PROOF OF LEMMA 1

We show below that $I'(X;Y) \leq I(X;Y)$. Start by writing

$$
\begin{aligned}
I(X;Y) - I'(X;Y) &= (H(Y) - H'(Y)) - (H(V) - H'(V)) \\
&= -\int_{-\infty}^{\infty} f_Y(t) \log\left(\frac{f_Y(t)}{f_Z(t)}\right) dt + \int_{-\infty}^{\infty} f_V(t) \log\left(\frac{f_V(t)}{f_G(t)}\right) dt \\
&= -D\Big(f_Y(t)\|f_Z(t)\Big) + D\Big(f_V(t)\|f_G(t)\Big)
\end{aligned}
$$

(20)

where $D\Big(p(t)\|q(t)\Big)$ is the Kullback-Leibler (K-L) divergence defined as

$$
D\Big(p(t)\|q(t)\Big) \triangleq \int_{-\infty}^{\infty} p(t) \log\left(\frac{p(t)}{q(t)}\right) dt.
$$

The K-L divergence is always greater than or equal to zero and convex in pair $(p(t)\|q(t))$, [19], i.e, assuming $p_1(t), q_1(t), p_2(t),$ and $q_2(t)$ are all pdfs, for $0 \leq \lambda \leq 1$, we have

$$
D\Big(\lambda p_1(t) + (1-\lambda)p_2(t)\|\lambda q_1(t) + (1-\lambda)q_2(t)\Big) \leq \lambda D\Big(p_1(t)\|q_1(t)\Big) + (1-\lambda)D\Big(p_2(t)\|q_2(t)\Big). \quad (21)
$$

For the sake of clarity, we assume that X is from the binary phase shift keying (BPSK) alphabet, i.e., $X \in \{\pm\sqrt{P_X}\}$. Then,

$$
\begin{aligned}
f_Y(t) &= \frac{1}{2}\left\{f_V(t - \sqrt{P_X}) + f_V(t + \sqrt{P_X})\right\} \\
f_Z(t) &= \frac{1}{2}\left\{f_G(t - \sqrt{P_X}) + f_G(t + \sqrt{P_X})\right\}.
\end{aligned}
$$

Substituting $p_1(t) = f_V(t - \sqrt{P_X})$, $p_2(t) = f_V(t + \sqrt{P_X})$, $q_1(t) = f_G(t - \sqrt{P_X})$, $q_2(t) = f_G(t + \sqrt{P_X})$, and $\lambda = 0.5$ in (21), we get

$$
\begin{aligned}
D\Big(f_Y(t)\|f_Z(t)\Big) &\leq \frac{1}{2}\left\{D\Big(f_V(t - \sqrt{P_X})\|f_G(t - \sqrt{P_X})\Big) + D\Big(f_V(t + \sqrt{P_X})\|f_G(t + \sqrt{P_X})\Big)\right\} \\
&= D\Big(f_V(t)\|f_G(t)\Big).
\end{aligned}
$$

Accordingly, (20) is always greater than or equal to zero or $I'(X;Y) \leq I(X;Y)$. While this proof is for the binary alphabet, it is easy to see that the application of the pair wise convexity of (21) for any independent and identically distributed (i.i.d.) input leads to the same conclusion.

APPENDIX B

DERIVATION OF THE PROPOSITION 1

From the pdfs of RVs V and G, we can write

$$
\begin{aligned}
H'(V) &= -\int_{-\infty}^{\infty} f_V(t) \log f_G(t) dt \\
&= \frac{1}{2} \log\left(2\pi\sigma_V^2\right) + \int_{-\infty}^{\infty} \frac{t^2}{2\sigma_V^2} f_V(t) dt.
\end{aligned} \tag{22}
$$

Moreover, we have

$$
f_Y(t) = \frac{1}{2} \left\{ f_V(t - \sqrt{P_X}) + f_V(t + \sqrt{P_X}) \right\}
$$

$$
\begin{aligned}
f_Z(t) &= \frac{1}{2} \left\{ f_G(t - \sqrt{P_X}) + f_G(t + \sqrt{P_X}) \right\} \\
&= \frac{1}{2} \left\{ \frac{1}{\sqrt{2\pi\sigma_V^2}} \exp\left(-\frac{(t - \sqrt{P_X})^2}{2\sigma_V^2}\right) + \frac{1}{\sqrt{2\pi\sigma_V^2}} \exp\left(-\frac{(t + \sqrt{P_X})^2}{2\sigma_V^2}\right) \right\} \\
&= \frac{1}{2\sqrt{2\pi\sigma_V^2}} \exp\left(-\frac{(t - \sqrt{P_X})^2}{2\sigma_V^2}\right) \left\{ 1 + \exp\left(\frac{-2\sqrt{P_X}t}{\sigma_V^2}\right) \right\} \\
&= \frac{1}{2\sqrt{2\pi\sigma_V^2}} \exp\left(-\frac{(t + \sqrt{P_X})^2}{2\sigma_V^2}\right) \left\{ 1 + \exp\left(\frac{2\sqrt{P_X}t}{\sigma_V^2}\right) \right\}.
\end{aligned}
$$

We can write $-\log f_Z(t)$ in two different ways:

$$
\begin{aligned}
-\log f_Z(t) &= \log 2 + \frac{1}{2} \log\left(2\pi\sigma_V^2\right) + \frac{(t - \sqrt{P_X})^2}{2\sigma_V^2} - \log\left\{ 1 + \exp\left(\frac{-2\sqrt{P_X}t}{\sigma_V^2}\right) \right\} \\
&= \log 2 + \frac{1}{2} \log\left(2\pi\sigma_V^2\right) + \frac{(t + \sqrt{P_X})^2}{2\sigma_V^2} - \log\left\{ 1 + \exp\left(\frac{2\sqrt{P_X}t}{\sigma_V^2}\right) \right\}.
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
-\frac{1}{2} \int_{-\infty}^{\infty} f_V(t - \sqrt{P_X}) \log f_Z(t) dt &= \frac{1}{2} \left\{ \log 2 + \frac{1}{2} \log\left(2\pi\sigma_V^2\right) \right\} + \frac{1}{2} \int_{-\infty}^{\infty} \frac{(t - \sqrt{P_X})^2}{2\sigma_V^2} f_V(t - \sqrt{P_X}) dt \\
&\quad - \frac{1}{2} \int_{-\infty}^{\infty} \log\left\{ 1 + \exp\left(\frac{-2\sqrt{P_X}t}{\sigma_V^2}\right) \right\} f_V(t - \sqrt{P_X}) dt \\
&= \frac{1}{2} \left\{ \log 2 + \frac{1}{2} \log\left(2\pi\sigma_V^2\right) \right\} + \frac{1}{2} \int_{-\infty}^{\infty} \frac{t^2}{2\sigma_V^2} f_V(t) dt \\
&\quad - \frac{1}{2} \int_{-\infty}^{\infty} \log\left\{ 1 + \exp\left(\frac{-2\sqrt{P_X}t - 2P_X}{\sigma_V^2}\right) \right\} f_V(t) dt.
\end{aligned}
$$

Similarly,

$$-\frac{1}{2}\int_{-\infty}^{\infty} f_V(t+\sqrt{P_X})\log f_Z(t)dt = \frac{1}{2}\left\{\log 2 + \frac{1}{2}\log\left(2\pi\sigma_V^2\right)\right\} + \frac{1}{2}\int_{-\infty}^{\infty}\frac{(t+\sqrt{P_X})^2}{2\sigma_V^2}f_V(t+\sqrt{P_X})dt$$

$$-\frac{1}{2}\int_{-\infty}^{\infty}\log\left\{1+\exp\left(\frac{2\sqrt{P_X}t}{\sigma_V^2}\right)\right\}f_V(t+\sqrt{P_X})dt$$

$$= \frac{1}{2}\left\{\log 2 + \frac{1}{2}\log\left(2\pi\sigma_V^2\right)\right\} + \frac{1}{2}\int_{-\infty}^{\infty}\frac{t^2}{2\sigma_V^2}f_V(t)dt$$

$$-\frac{1}{2}\int_{-\infty}^{\infty}\log\left\{1+\exp\left(\frac{2\sqrt{P_X}t-2P_X}{\sigma_V^2}\right)\right\}f_V(t)dt.$$

Accordingly,

$$H'(Y) = -\int_{-\infty}^{\infty} f_Y(t)\log f_Z(t)dt$$

$$= -\frac{1}{2}\int_{-\infty}^{\infty} f_V(t-\sqrt{P_X})\log f_Z(t)dt - \frac{1}{2}\int_{-\infty}^{\infty} f_V(t+\sqrt{P_X})\log f_Z(t)dt$$

$$= \log 2 + \frac{1}{2}\log\left(2\pi\sigma_V^2\right) + \int_{-\infty}^{\infty}\frac{t^2}{2\sigma_V^2}f_V(t)dt$$

$$-\int_{-\infty}^{\infty}\frac{1}{2}\left[\log\left\{1+\exp\left(\frac{-2\sqrt{P_X}t-2P_X}{\sigma_V^2}\right)\right\} + \log\left\{1+\exp\left(\frac{2\sqrt{P_X}t-2P_X}{\sigma_V^2}\right)\right\}\right]f_V(t)dt$$

$$= \log 2 + \frac{1}{2}\log\left(2\pi\sigma_V^2\right) + \int_{-\infty}^{\infty}\frac{t^2}{2\sigma_V^2}f_V(t)dt - \int_{-\infty}^{\infty}\log\left\{1+\exp\left(\frac{-2\sqrt{P_X}t-2P_X}{\sigma_V^2}\right)\right\}f_V(t)dt. \tag{23}$$

The last equality in (23) holds because $f_V(t)$ is an even function. Finally, we arrive at

$$I'(X;Y) = H'(Y) - H'(V)$$

$$= \log 2 - \int_{-\infty}^{\infty} f_V(t)\log\left\{1+\exp\left(\frac{-2\sqrt{P_X}t-2P_X}{\sigma_V^2}\right)\right\}dt. \tag{24}$$

Now write $I'(X;Y) = \log 2 - F$ with the new definition

$$F \triangleq \int_{-\infty}^{\infty} f_V(t)\log\left\{1+\exp\left(\frac{-2\sqrt{P_X}t-2P_X}{\sigma_V^2}\right)\right\}dt$$

$$= 2^{-L}\sum_{i=1}^{2^L}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma_N^2}}\exp\left(-\frac{(t-m_i)^2}{2\sigma_N^2}\right)\log\left\{1+\exp\left(\frac{-2\sqrt{P_X}t-2P_X}{\sigma_V^2}\right)\right\}dt$$

$$= 2^{-L}\sum_{i=1}^{2^L}\int_{-\infty}^{\infty}\frac{e^{-\tau^2/2}}{\sqrt{2\pi}}\log\left\{1+\exp\left(\frac{-2\sqrt{P_X}\left(\tau\sigma_N+m_i\right)-2P_X}{\sigma_V^2}\right)\right\}d\tau$$

$$= 2^{-L}\sum_{i=1}^{2^L}\int_{-\infty}^{\infty}\frac{e^{-\tau^2/2}}{\sqrt{2\pi}}\log\left\{1+e^{-2R\rho_i}e^{-2\phi\sqrt{R}\tau-2R}\right\}d\tau$$

$$= 2^{-L}\sum_{i=1}^{2^L}\mathrm{E}_\tau\left[\log\left\{1+e^{-2R\rho_i}e^{-2\phi\sqrt{R}\tau-2R}\right\}\right] \tag{25a}$$

where the third equality is obtained with a variable change $(t - m_i)/\sigma_N = \tau$ and $\rho_i \triangleq m_i/\sqrt{P_X}$, $R \triangleq P_X/\sigma_V^2$, and $\phi \triangleq \sigma_N/\sigma_V$. The expression (25a) can also be written as

$$
\begin{aligned}
F &= 2^{-L} \sum_{k=1}^{2^{L-1}} \mathrm{E}_\tau \left[\log\left\{ 1 + e^{-2R\rho_k^+} e^{-2\phi\sqrt{R}\tau - 2R} \right\} + \log\left\{ 1 + e^{2R\rho_k^+} e^{-2\phi\sqrt{R}\tau - 2R} \right\} \right] \\
&= 2^{-L} \sum_{k=1}^{2^{L-1}} \mathrm{E}_\tau \left[\log\left\{ 1 + \left(e^{-2R\rho_k^+} + e^{2R\rho_k^+} \right) e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4\phi\sqrt{R}\tau - 4R} \right\} \right] \\
&= 2^{-(L-1)} \sum_{k=1}^{2^{L-1}} \mathrm{E}_\tau \left[\frac{1}{2} \log\left\{ 1 + 2\cosh\left(2R\rho_k^+ \right) e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4\phi\sqrt{R}\tau - 4R} \right\} \right]
\end{aligned}
\tag{25b}
$$

where ρ_k^+'s is the positive-half subset of ρ_i's.

APPENDIX C

DERIVATION OF THE SIMPLE BOUNDS

Due to the convexity of $\mathrm{E}_\tau \left[\frac{1}{2} \log\left(1 + 2\cosh(2R\rho^+)e^{-2\phi\sqrt{R}\tau} + e^{-4\phi\sqrt{R}\tau - 4R} \right) \right]$ in ρ^+, the upper bound of F can be found as

$$
\begin{aligned}
F &= 2^{-(L-1)} \sum_{k=1}^{2^{L-1}} \mathrm{E}_\tau \left[\frac{1}{2} \log\left\{ 1 + 2\cosh\left(2R\rho_k^+ \right) e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4\phi\sqrt{R}\tau - 4R} \right\} \right] \\
&\leq 2^{-(L-1)} \sum_{k=1}^{2^{L-1}} \left\{ T\left(|\rho|_{\max}, \theta \right) \Big|_{\theta = \rho_k^+} \right\} \\
&= T\left(|\rho|_{\max}, \theta \right) \Big|_{\theta = 2^{-(L-1)} \sum_{k=1}^{2^{L-1}} \rho_k^+} = T\left(|\rho|_{\max}, \theta \right) \Big|_{\theta = |\rho|_{\mathrm{avg}}} \\
&\leq T\left(|\rho|_{\max}, \theta \right) \Big|_{\theta = \sigma_\rho} \triangleq F^{u1}
\end{aligned}
\tag{26}
$$

where $|\rho|_{\mathrm{avg}} \triangleq 2^{-L} \sum_{i=1}^{2^L} |\rho_i| = 2^{-(L-1)} \sum_{k=1}^{2^{L-1}} \rho_k^+$ and, for a given $|\rho|_{\max}$, $T(|\rho|_{\max}, \theta)$ represents a straight line passing through the two points of $\mathrm{E}_\tau \left[\frac{1}{2} \log\left(1 + 2\cosh(2R\theta)e^{-2\phi\sqrt{R}\tau} + e^{-4\phi\sqrt{R}\tau - 4R} \right) \right]$: at $\theta = 0$ and at $\theta = |\rho|_{\max}$. The last inequality is obtained from the Cauchy-Schwarz inequality: $|\rho|_{\mathrm{avg}} \leq \sigma_\rho$.

Another upper bound of F can be also found as

$$
\begin{aligned}
F &= 2^{-(L-1)} \sum_{k=1}^{2^{L-1}} \mathrm{E}_\tau \left[\frac{1}{2} \log\left\{ 1 + 2\alpha_k e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4\phi\sqrt{R}\tau - 4R} \right\} \right] \\
&\leq \mathrm{E}_\tau \left[\frac{1}{2} \log\left\{ 1 + 2\left(2^{-(L-1)} \sum_{k=1}^{2^{L-1}} \alpha_k \right) e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4\phi\sqrt{R}\tau - 4R} \right\} \right] \\
&= \mathrm{E}_\tau \left[\frac{1}{2} \log\left\{ 1 + 2\alpha_{\mathrm{avg}} e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4\phi\sqrt{R}\tau - 4R} \right\} \right]
\end{aligned}
\tag{27}
$$

where $\alpha_k \triangleq \cosh(2R\rho_k^+)$ and $\alpha_{\mathrm{avg}} \triangleq 2^{-(L-1)} \sum_{k=1}^{2^{L-1}} \alpha_k = 2^{-(L-1)} \sum_{k=1}^{2^{L-1}} \cosh(2R\rho_k^+)$. The inequality comes from the concavity of $\mathrm{E}_\tau \left[\frac{1}{2} \log\left(1 + 2\alpha e^{-2\phi\sqrt{R}\tau} + e^{-4\phi\sqrt{R}\tau - 4R} \right) \right]$ in α. Moreover, since it is an

increasing function of α, the last expression of (27) can be further upper-bounded by replacing α' with $\alpha' \geq \alpha_{\mathrm{avg}}$. For example, note

$$\alpha_{\mathrm{avg}} \leq \frac{1}{2^{(L-1)}} \sum_{k=1}^{2^{L-1}} \left(s\rho_k^+ + 1 \right) = s\left|\rho\right|_{\mathrm{avg}} + 1 \leq s\sigma_\rho + 1 \triangleq \alpha'$$

where $s = \left(\cosh(2R|\rho|_{\max}) - 1\right)/|\rho|_{\max}$, the slope of a straight line connecting two points $(0, 1)$ and $(|\rho|_{\max}, \cosh(2R|\rho|_{\max}))$. This gives

$$F \leq \mathrm{E}_\tau \left[\frac{1}{2} \log \left\{ 1 + 2(s\sigma_\rho + 1)e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4\phi\sqrt{R}\tau - 4R} \right\} \right] \triangleq F^{u2}. \tag{28}$$

By using the convexity of $\mathrm{E}_\tau \left[\log \left(1 + e^{-2R\rho}e^{-2\phi\sqrt{R}\tau - 2R} \right) \right]$ in ρ, the lower bound of F is also found as

$$
\begin{aligned}
F &= 2^{-L} \sum_{i=1}^{2^L} \mathrm{E}_\tau \left[\log \left\{ 1 + e^{-2R\rho_i} e^{-2\phi\sqrt{R}\tau - 2R} \right\} \right] \\
&\geq \mathrm{E}_\tau \left[\log \left\{ 1 + \exp\left(-2R \left(2^{-L} \sum_{i=1}^{2^L} \rho_i \right) \right) e^{-2\phi\sqrt{R}\tau - 2R} \right\} \right] \\
&= \mathrm{E}_\tau \left[\log \left\{ 1 + e^{-2\phi\sqrt{R}\tau - 2R} \right\} \right] \\
&= \mathrm{E}_\tau \left[\frac{1}{2} \log \left\{ 1 + 2e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4\phi\sqrt{R}\tau - 4R} \right\} \right] \triangleq F^l. \tag{29}
\end{aligned}
$$

APPENDIX D

DERIVATION OF THE TIGHTENED BOUNDS

The tightened bounds are derived in a similar way using the convexity or concavity of the function except the cluster identification needs be incorporated. Since $\rho_k = \lambda_j + \mu_i$, we can write F as

$$
\begin{aligned}
F &= 2^{-M} \sum_{j=1}^{2^M} \left(2^{-(L-M)} \sum_{i=1}^{2^{L-M}} \mathrm{E}_\tau \left[\log \left\{ 1 + e^{-2R(\mu_i + \lambda_j)} e^{-2\phi\sqrt{R}\tau - 2R} \right\} \right] \right) \\
&= 2^{-M} \sum_{j=1}^{2^M} \left(2^{-(L-M-1)} \sum_{l=1}^{2^{L-M-1}} \mathrm{E}_\tau \left[\frac{1}{2} \log \left\{ 1 + 2\cosh\left(2R\mu_l^+\right) e^{-2R\lambda_j} e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4R\lambda_j} e^{-4\phi\sqrt{R}\tau - 4R} \right\} \right] \right) \\
&\leq 2^{-M} \sum_{j=1}^{2^M} \left(2^{-(L-M-1)} \sum_{l=1}^{2^{L-M-1}} \left\{ T_j(|\mu|_{\max}, \theta) \Big|_{\theta = \mu_l^+} \right\} \right) \\
&= 2^{-M} \sum_{j=1}^{2^M} \left\{ T_j\left(|\mu|_{\max}, \theta\right) \Big|_{\theta = 2^{-(L-M-1)} \sum_{l=1}^{2^{L-M-1}} \mu_l^+} \right\} = 2^{-M} \sum_{j=1}^{2^M} \left\{ T_j(|\mu|_{\max}, \theta) \Big|_{\theta = |\mu|_{\mathrm{avg}}} \right\} \\
&\leq 2^{-M} \sum_{j=1}^{2^M} \left\{ T_j(|\mu|_{\max}, \theta) \Big|_{\theta = \sigma_\mu} \right\} \triangleq F_M^{u1} \tag{30}
\end{aligned}
$$

where μ_l^+'s form the positive-half subset of μ_i's and, for a given $|\mu|_{\max}$, $T_j(|\mu|_{\max}, \theta)$ is a straight line that passes through the convex function $\mathrm{E}_\tau \left[\frac{1}{2} \log \left\{ 1 + 2 \cosh(2R\theta) e^{-2R\lambda_j} e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4R\lambda_j} e^{-4\phi\sqrt{R}\tau - 4R} \right\} \right]$ at $\theta = 0$ and $\theta = |\mu|_{\max}$. Moreover, $|\mu|_{\mathrm{avg}} \triangleq 2^{-(L-M)} \sum_{i=1}^{2^{L-M}} |\mu_i| = 2^{-(L-M-1)} \sum_{l=1}^{2^{L-M-1}} \mu_l^+$. The last inequality also follows from $|\mu|_{\mathrm{avg}} \le \sigma_\mu$, and note $\sigma_\mu = \sqrt{\sigma_\rho^2 - \sigma_\lambda^2}$ and $|\mu|_{\max} = |\rho|_{\max} - |\lambda|_{\max}$.

Another form of tightened upper bound of F is obtained as

$$F = 2^{-M} \sum_{j=1}^{2^M} \left(2^{-(L-M-1)} \sum_{l=1}^{2^{L-M-1}} \mathrm{E}_\tau \left[\frac{1}{2} \log \left\{ 1 + 2 \cosh\left(2R\mu_l^+\right) e^{-2R\lambda_j} e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4R\lambda_j} e^{-4\phi\sqrt{R}\tau - 4R} \right\} \right] \right)$$

$$= 2^{-M} \sum_{j=1}^{2^M} \left(2^{-(L-M-1)} \sum_{l=1}^{2^{L-M-1}} \mathrm{E}_\tau \left[\frac{1}{2} \log \left\{ 1 + 2\beta_l e^{-2R\lambda_j} e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4R\lambda_j} e^{-4\phi\sqrt{R}\tau - 4R} \right\} \right] \right)$$

$$\le 2^{-M} \sum_{j=1}^{2^M} \mathrm{E}_\tau \left[\frac{1}{2} \log \left\{ 1 + 2 \left(2^{-(L-M-1)} \sum_{l=1}^{2^{L-M-1}} \beta_l \right) e^{-2R\lambda_j} e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4R\lambda_j} e^{-4\phi\sqrt{R}\tau - 4R} \right\} \right]$$

$$= 2^{-M} \sum_{j=1}^{2^M} \mathrm{E}_\tau \left[\frac{1}{2} \log \left\{ 1 + 2\beta_{\mathrm{avg}} e^{-2R\lambda_j} e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4R\lambda_j} e^{-4\phi\sqrt{R}\tau - 4R} \right\} \right]$$

$$\le 2^{-M} \sum_{j=1}^{2^M} \mathrm{E}_\tau \left[\frac{1}{2} \log \left\{ 1 + 2\beta' e^{-2R\lambda_j} e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4R\lambda_j} e^{-4\phi\sqrt{R}\tau - 4R} \right\} \right]$$

$$= 2^{-M} \sum_{j=1}^{2^M} \mathrm{E}_\tau \left[\frac{1}{2} \log \left\{ 1 + 2 \left(s_M \sigma_\mu + 1 \right) e^{-2R\lambda_j} e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4R\lambda_j} e^{-4\phi\sqrt{R}\tau - 4R} \right\} \right] \triangleq F_M^{u2} \tag{31}$$

where $\beta_l \triangleq \cosh\left(2R\mu_l^+\right)$, $\beta_{\mathrm{avg}} \triangleq 2^{-(L-M-1)} \sum_{l=1}^{2^{L-M-1}} \beta_l = 2^{-(L-M-1)} \sum_{l=1}^{2^{L-M-1}} \cosh(2R\mu_l^+)$ and

$$\beta_{\mathrm{avg}} \le \frac{1}{2^{(L-M-1)}} \sum_{k=1}^{2^{L-M-1}} \left(s_M \mu_k^+ + 1 \right) = s_M |\mu|_{\mathrm{avg}} + 1 \le s_M \sigma_\mu + 1 \triangleq \beta'$$

which is based on a straight line connecting two points of the convex function $\cosh(2R\mu)$, $(0, 1)$ and $(|\mu|_{\max}, \cosh(2R|\mu|_{\max}))$, having a slope $s_M = \left(\cosh(2R|\mu|_{\max}) - 1 \right) / |\mu|_{\max}$.

The tightened lower bound of F based on cluster identification is obtained as

$$
\begin{aligned}
F &= 2^{-M} \sum_{j=1}^{2^M} \left(2^{-(L-M)} \sum_{i=1}^{2^{L-M}} \mathrm{E}_\tau \left[\log \left\{ 1 + e^{-2R(\mu_i + \lambda_j)} e^{-2\phi\sqrt{R}\tau - 2R} \right\} \right] \right) \\
&= 2^{-M} \sum_{j=1}^{2^M} \left(2^{-(L-M)} \sum_{i=1}^{2^{L-M}} \mathrm{E}_\tau \left[\log \left\{ 1 + e^{-2R\mu_i} e^{-2R\lambda_j} e^{-2\phi\sqrt{R}\tau - 2R} \right\} \right] \right) \\
&\geq 2^{-M} \sum_{j=1}^{2^M} \mathrm{E}_\tau \left[\log \left\{ 1 + \exp \left[-2R \left(2^{-(L-M)} \sum_{i=1}^{2^{L-M}} \mu_i \right) \right] e^{-2R\lambda_j} e^{-2\phi\sqrt{R}\tau - 2R} \right\} \right] \\
&= 2^{-M} \sum_{j=1}^{2^M} \mathrm{E}_\tau \left[\log \left\{ 1 + e^{-2R\lambda_j} e^{-2\phi\sqrt{R}\tau - 2R} \right\} \right] \\
&= 2^{-(M-1)} \sum_{k=1}^{2^{M-1}} \mathrm{E}_\tau \left[\frac{1}{2} \log \left\{ 1 + 2\cosh\left(2R\lambda_k^+ \right) e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4\phi\sqrt{R}\tau - 4R} \right\} \right] \triangleq F_M^l \quad (32)
\end{aligned}
$$

where λ_k^+'s form the positive-half subset of λ_j's.

REFERENCES

[1] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate," *IEEE Trans. Information Theory*, vol. IT-20, pp. 284-287, Mar. 1974.

[2] D. Arnold and H. -A. Loeliger, "On the Information Rate of Binary-Input Channels with Memory," *in Proc. IEEE Int. Conf. Communications*, Helsinki, Finland, June 2001, pp. 2692-2695.

[3] V. Sharma and S. K. Singh, "Entropy and Channel Capacity in the Regenerative Setup with Applications to Markov Channels," *in Proc. IEEE Int. Symp. on Information Theory*, Washington, DC, USA, Jun. 2001, pp. 283.

[4] H. D. Pfister, J. B. Soriaga, and P. H. Siegel, "On the Achievable Information Rates of Finite State ISI Channels," *In Proc. IEEE GLOBECOM*, San Antonio, Texas, USA, Nov. 2001, pp. 2992-2996.

[5] D. Arnold, H. -A. Loeliger, P. Vontobel, A. Kavčić, and W. Zeng "Simulation-Based Computation of Information Rates for Channels with Memory," *IEEE Trans. Information Theory*, vol. 52, no. 8, pp. 3498-3508, Aug. 2006.

[6] P. Sadeghi, P. O. Vontobel, and R. Shams, "Optimization of Information Rate Upper and Lower Bounds for Channels With Memory," *IEEE Trans. Information Theory*, vol. 55, no. 2, pp. 663-688, Feb. 2009.

[7] F. Rusek and D. Fertonani, "Lower Bounds on the Information Rate of Intersymbol Interference Channels based on the Ungerboeck Observation Model," *in Proc. IEEE Int. Symp. on Information Theory*, Seoul, Korea, pp. 1649-1653, June-July 2009.

[8] G. Ungerboeck, "Adaptive Maximum-Likelihood Receiver for Carrier-Modulated Data-Transmission Systems," *IEEE Trans. Communications*, vol. 22, pp. 624-636, May 1974.

[9] W. Hirt, "Capacity and Information Rates of Discrete-Time Channels with Memory," *Ph.D. thesis no. 8671*, ETH Zurich, 1988.

[10] S. Shamai, L. H. Ozarow, and A. D. Wyner, "Infomation Rate for a Discrete-Time Gaussian Channel with Intersymbol Interference and Stationary Inputs," *IEEE Trans. Information Theory*, vol. 37, no. 6, pp. 1527-1539, Nov. 1991.

[11] S. Shamai and R. Laroia, "The Intersymbol Interference Channel: Lower Bounds on Capacity and Channel Precoding Loss," *IEEE Trans. Information Theory*, vol. 42, no. 5, pp. 1388-1404, Sept. 1996.

[12] S. Shamai and S. Verdu, "Worst-case power-constrained noise for binary-input channels," *IEEE Trans. Information Theory*, vol. 38, no. 5, pp. 1494-1511, Sept. 1992.

[13] A. M. Garsia, "Entropy and singularity of infinite convolutions," *Pacific J. Math.*, vol. 13, no. 4, pp. 1159-1169, 1963.

[14] P. H. Wittke, W. S. Smith, and L. L. Campbell, "Infinite Series of Interference Variables with Cantor-Type Distributions," *IEEE Trans. Information Theory*, vol. 34, no. 6, pp. 1428-1436, Nov. 1988.

[15] J. M. Cioffi, G. P. Dudevior, M. V. Eyuboglu, and G. D. Forney, Jr., "MMSE decision-feedback equalizers and coding - Part I: Equalization Results," *IEEE Trans. Communications*, vol. 43, no. 10, pp. 2582-2593, Oct. 1995.

[16] D. G. Messerschmitt, "A Geometric Theory of Intersymbol Interference, Part I: Zero-Forcing and Decision-Feedback Equalization," *Bell Syst. Tech. J.*, vol. 52, no. 9, pp. 1483-1539, Nov. 1973.

[17] G. D. Forney, Jr. and G. Ungerboeck, "Modulation and Coding for Linear Gaussian Channels," *IEEE Trans. Information Theory*, vol. 44, no. 6, pp. 2384-2415, Oct. 1998.

[18] R. G. Gallager, *Information Theory and Reliable Communication.* New York Wiley, 1968, pp. 97-112, 176-188.

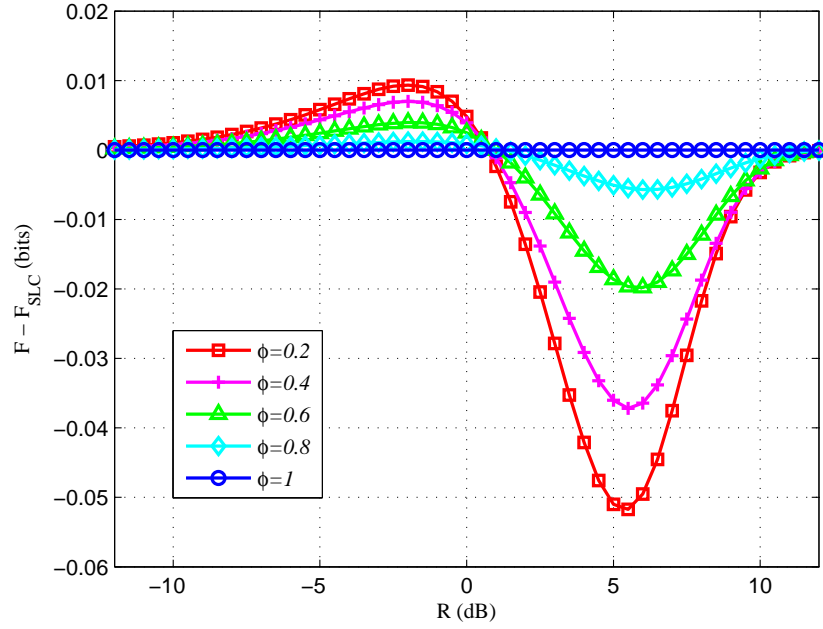[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York Wiley, 1991, pp. 29-31.

Fig. 1: $F - F_{SLC}$ as a function of R for a uniform ρ.
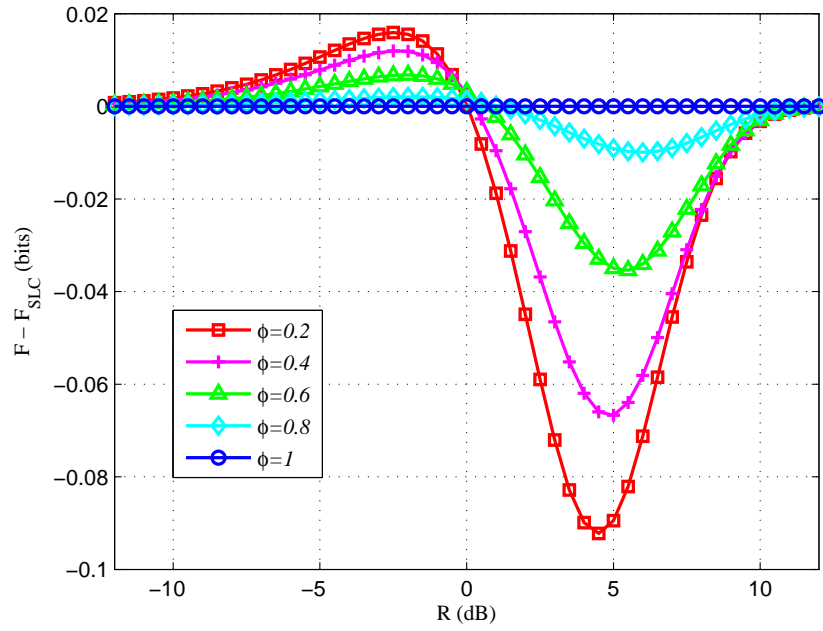

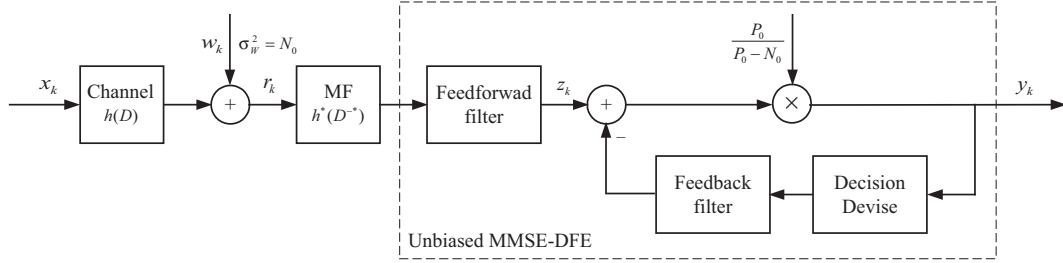Fig. 2: $F - F_{SLC}$ as a function of R for a two-valued ρ.
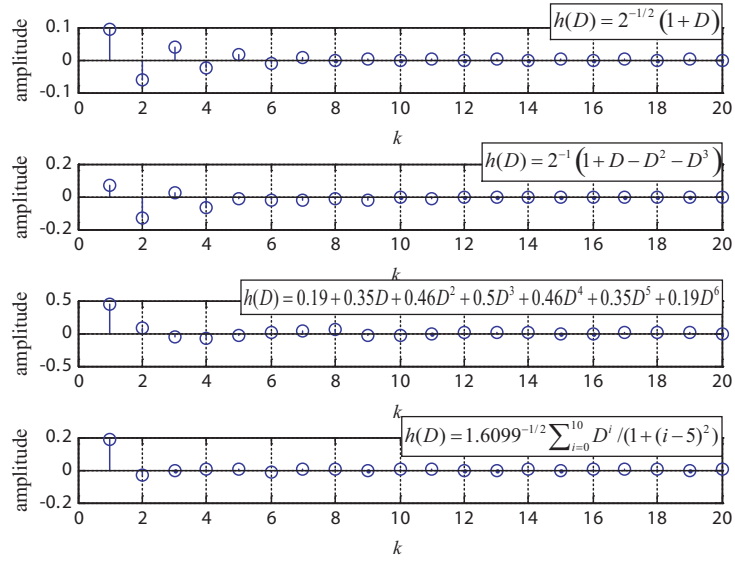
Fig. 3: System Model of ISI channels.

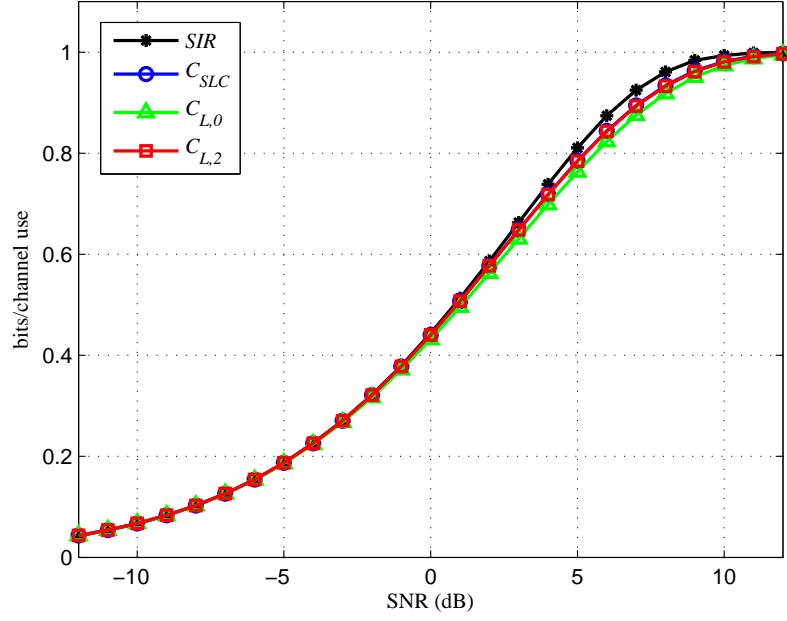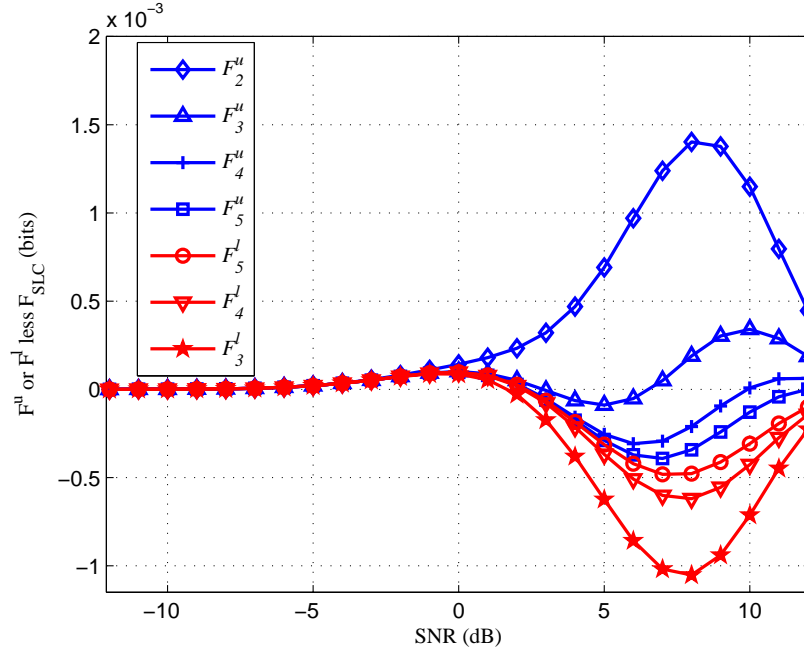Fig. 4: First 20 Precursor taps after unbiased MMSE-DFE at SNR=10 dB for four example channels.
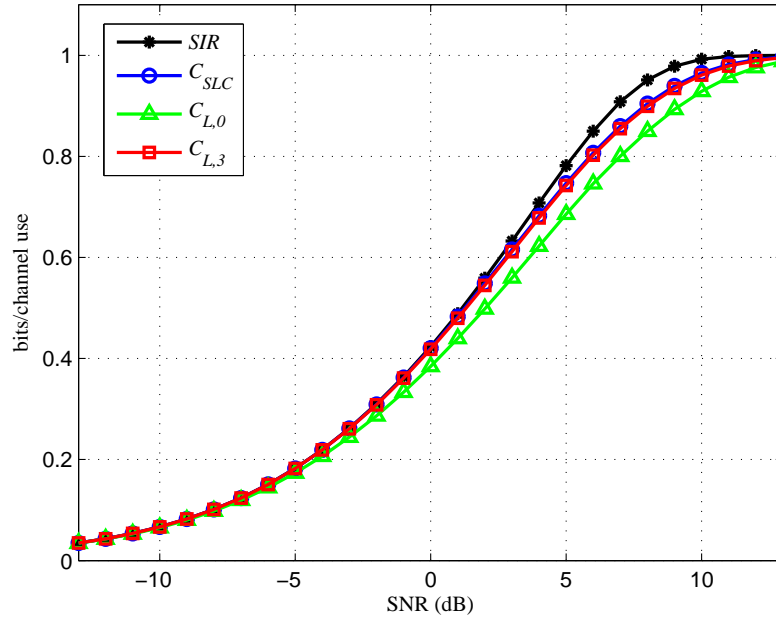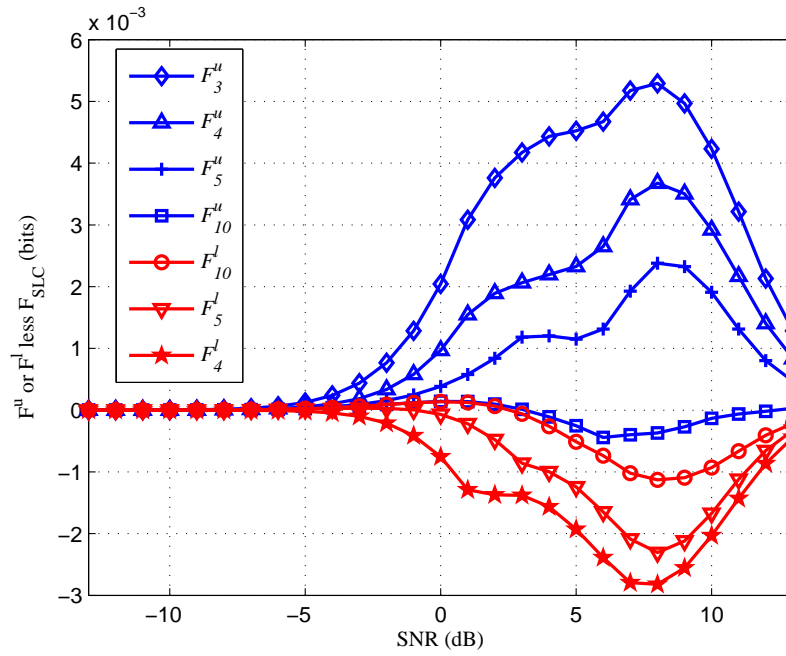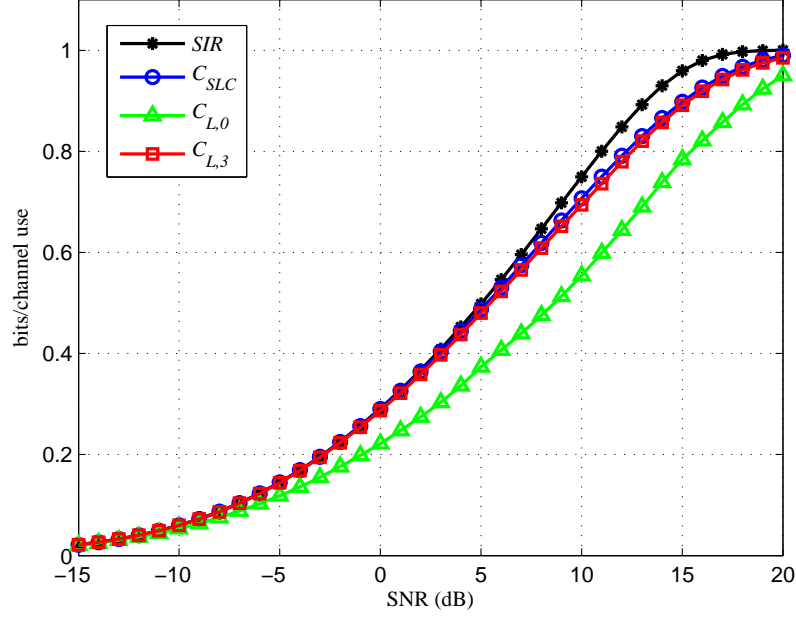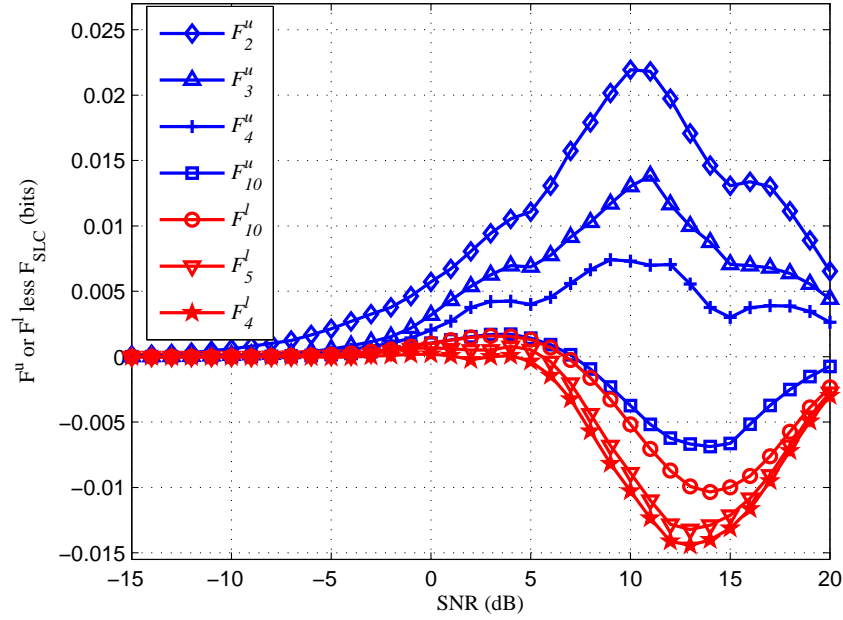
(a)

(b)

Fig. 5: Example channel: $h(D) = 2^{-1/2}(1 + D)$ with BPSK inputs (a) SIR, SLC and the new lower bounds as functions of SNR (b) Upper and lower bounds of F, for different M, less F_{SLC}, plotted against SNR.
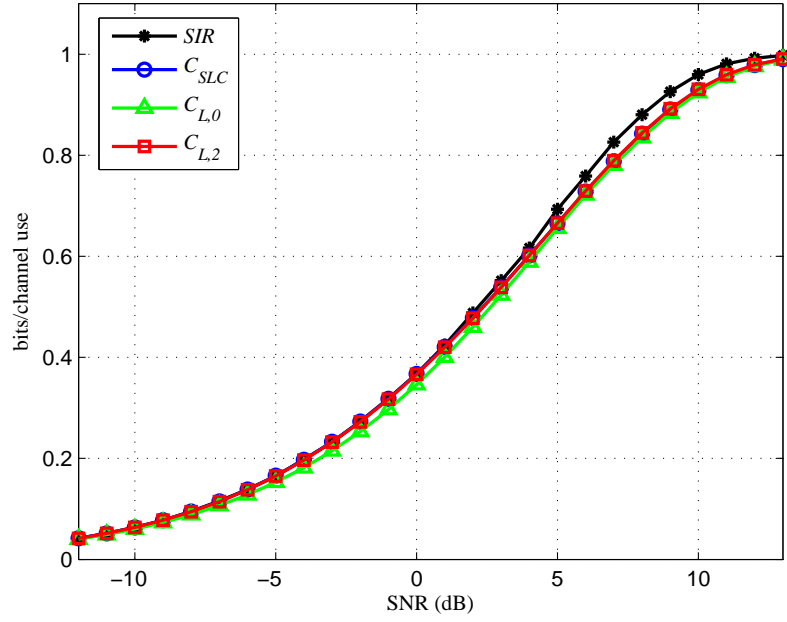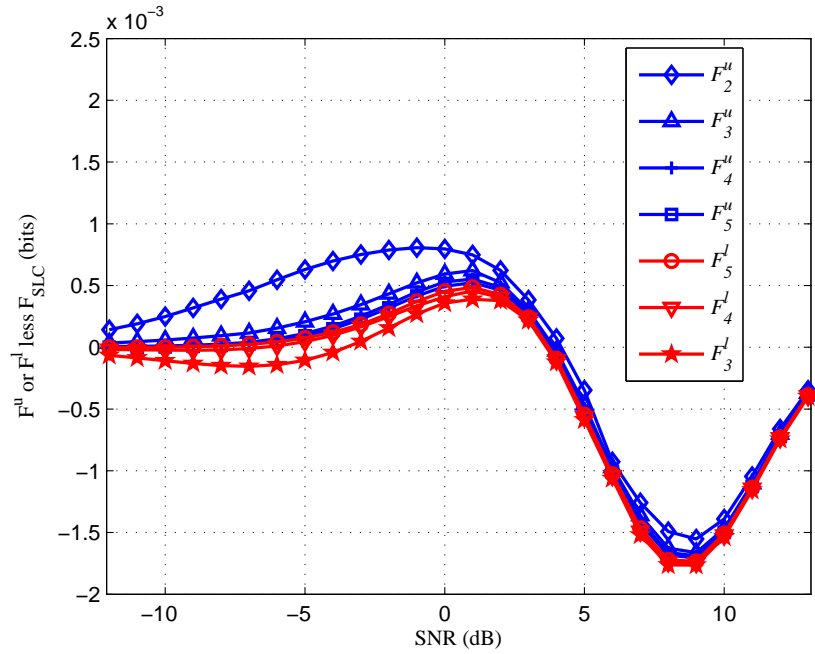
(a)

(b)

Fig. 6: Example channel: $h(D) = 2^{-1}(1 + D - D^2 - D^3)$ with BPSK inputs (a) SIR, SLC and the new lower bounds as functions of SNR (b) Upper and lower bounds of F, for different M, less F_{SLC}, plotted against SNR.

(a)

(b)

Fig. 7: Example channel: $h(D) = 0.19 + 0.35D + 0.46D^2 + 0.5D^3 + 0.46D^4 + 0.35D^5 + 0.19D^6$ with BPSK inputs (a) SIR, SLC and the new lower bounds as functions of SNR (b) Upper and lower bounds of F, for different M, less F_{SLC}, plotted against SNR.
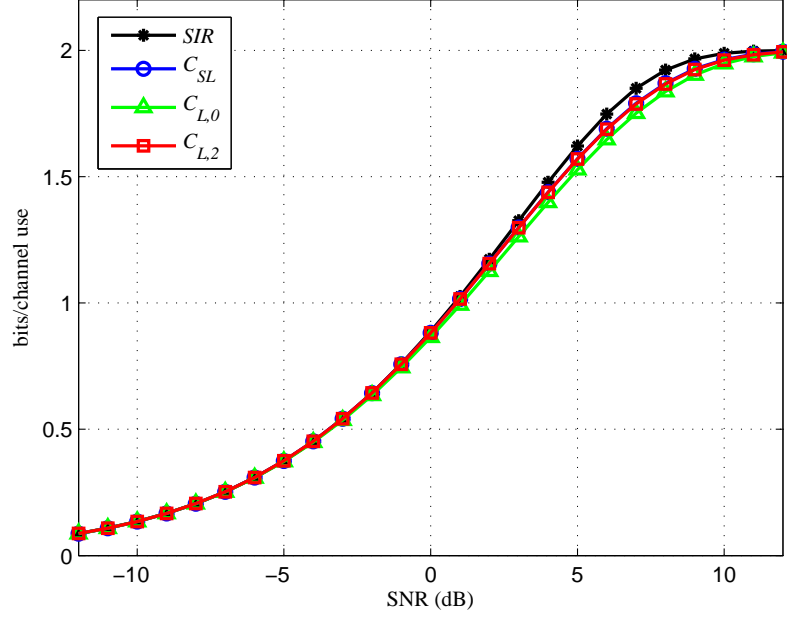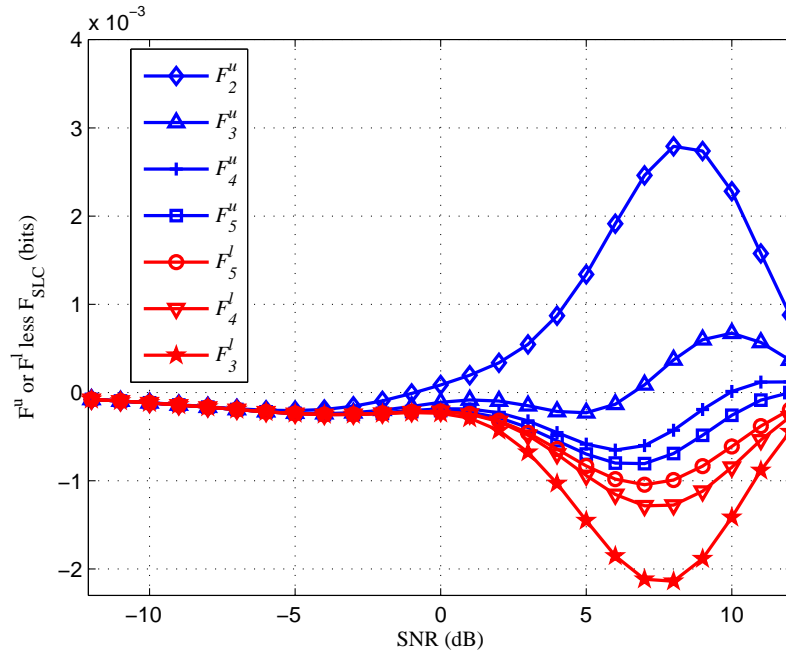
(a)

(b)

Fig. 8: Example channel: $h(D) = 1.6099^{-1/2} \sum_{i=0}^{10} D^i/(1 + (i-5)^2)$ with BPSK inputs (a) SIR, SLC and the new lower bounds as functions of SNR (b) Upper and lower bounds of F, for different M, less F_{SLC}, plotted against SNR.

(a)

(b)

Fig. 9: Example channel: $h(D) = 2^{-1}\{(1+j) + (1-j)D\}$ with QPSK inputs (a) SIR, SLC and the new lower bounds as functions of SNR (b) Upper and lower bounds of F, for different M, less F_{SLC}, plotted against SNR.