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 Institue of Science and Technology
Daejeon, 305-701, Republic of Korea
Email: jaemoon@ee.kaist.ac.kr

Abstract

Provable lower bounds are presented for the information rate $I(X;X+S+N)$ where X is the symbol drawn from a fixed, finite-size alphabet, S a discrete-valued random variable (RV) and N a Gaussian RV. The information rate $I(X;X+S+N)$ serves as a tight lower bound for capacity of intersymbol interference (ISI) channels corrupted by Gaussian noise. The new bounds can be calculated with a reasonable computational load and provide a similar level of tightness as the well-known conjectured lower bound by Shamai and Laroia for a good range of finite-ISI channels of practical interest. The computation of the presented bounds requires the evaluation of the magnitude sum of the precursor ISI terms as well as the identification of dominant terms among them seen at the output of the minimum mean-squared error (MMSE) decision feedback equalizer (DFE).

Index Terms

This work was supported in part by the NSF under Theoretical Foundation grant 0728676.

I. INTRODUCTION

Consider the random variable $Y = X + \sum_{k=1}^{L} d_{-k} X_k + N = X + S + N$, where X and X_k's are all independent identically distributed (i.i.d.) RVs each taking values from a finite alphabet characterized by equally likely symbols, d_{-k}'s are dimensionless coefficients related to the channel response and N is a zero-mean Gaussian RV with variance σ_N^2. The input alphabet is assumed to be symmetrically positioned around the origin with an average symbol power P_X. Finding the information rate $I(X;Y)$, which itself acts as a close lower bound for the symmetric information rate (SIR) of the intersymbol interference (ISI) channel [1], [2], is of great interest in digital communication. However, when L becomes large, finding an analytical expression for the probability density function (pdf) of $S+N$ for evaluating $I(X;Y)$ or finding easily computable and tight bounds for $I(X;Y)$ is a long-standing problem [2], [3], [4]. Perhaps the most well-known bound for the purpose of quickly estimating $I(X;Y)$ is the Shamai-Laroia conjectured bound [2]: $I(X;X+S+N) \geq I(X;X+G)$, where G is a Gaussian RV with variance matching that of $S+N$. Although reasonably close to $I(X;X+S+N)$ in most cases, $I(X;X+G)$ remains as a conjectured bound with no proof available to date.

In this paper, we are concerned with provable lower bounds for the information rate $I(X;Y)$ that can be easily computed. The bounds we develop here are fairly tight, with their tightness generally enhanced with increasing computational load. Our approach is to first find an information-rate-like function that depends on the probability densities of the underlying RVs and is always smaller than the information rate $I(X;Y)$. We then try to further bound this function from below so that the final bound can be evaluated using numerical integration. The bound computation requires the evaluation of sum of the absolute values of d_{-k}'s as well as the identification of dominant d_{-k} values, if they exist. At a reasonable computational load, the developed bounds are shown to be as tight as the Shamai-Loria conjecture (SLC) for many practical ISI channels.

II. A PROVABLE LOWER BOUND

We first present a provable lower bound to $I(X;Y)$. Let $V = S+N$ so we can write $Y = X+V$. Note 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., $\sigma_G^2 = \sigma_V^2$.

Now define an information-rate-like function:

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

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 pdf's of the RVs, Y, V, Z, and G, respectively.

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

$$\begin{aligned} I(Y;X) - I'(Y;X) &= (H(Y) - H'(Y)) - (H(V) - H'(V)) \\ &= -\int_{-\infty}^{\infty} f_Y(t) \log \left(\frac{f_Y(t)}{f_Z(t)} \right) dt + \int_{-\infty}^{\infty} f_V(t) \log \left(\frac{f_V(t)}{f_G(t)} \right) dt \\ &= -D\Big(f_Y(t) || f_Z(t)\Big) + D\Big(f_V(t) || f_G(t)\Big) \end{aligned} \tag{2}$$

where $D\Big(p(t)||q(t)\Big)$ is 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 $(p(t)||q(t))$, [5], 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

$$D\Big(\lambda p_1(t) + (1-\lambda) p_2(t) || \lambda q_1(t) + (1-\lambda) q_2(t)\Big) \leq \lambda D\Big(p_1(t)||q_1(t)\Big) + (1-\lambda) D\Big(p_2(t)||q_2(t)\Big). \tag{3}$$

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 (3), 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, (2) 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 (3) for any i.i.d. input leads to the same conclusion.

Let us now take a close look at this information-rate-like 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 continue to assume a binary input alphabet.

These assumptions are not necessary for our development but make the presentation less cluttered and clearer. We will simply state the results in Section III.C for a non-binary/complex-valued example.

Denoting $m_i = \sum_{k=1}^{L} d_{-k} X_k$ for $i = 1, 2, \ldots, 2^L$, the pdf's can be written as

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

Accordingly, we can write

$$\begin{aligned} H'(V) &= -\int_{-\infty}^{\infty} f_V(t) \log f_G(t) dt \\ &= \frac{1}{2} \log\left(2\pi\sigma_V^2\right) + \int_{-\infty}^{\infty} \frac{t^2}{2\sigma_V^2} f_V(t) dt. \end{aligned} \tag{4}$$

Moreover, we have

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

Therefore, we can represent $-\log f_Z(t)$ 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,

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

Accordingly,

$$\begin{aligned}
H'(Y) &= -\int_{-\infty}^{\infty} f_Y(t)\log f_Z(t)dt\\
&= -\frac{1}{2}\int_{-\infty}^{\infty} f_V(t-\sqrt{P_X})\log f_Z(t)dt-\frac{1}{2}\int_{-\infty}^{\infty} f_V(t+\sqrt{P_X})\log f_Z(t)dt\\
&= \log 2+\frac{1}{2}\log\left(2\pi\sigma_V^2\right)+\int_{-\infty}^{\infty}\frac{t^2}{2\sigma_V^2}f_V(t)dt\\
&\quad -\int_{-\infty}^{\infty}\frac{1}{2}\left[\log\left\{1+\exp\left(\frac{-2\sqrt{P_X}t-2P_X}{\sigma_V^2}\right)\right\}+\log\left\{1+\exp\left(\frac{2\sqrt{P_X}t-2P_X}{\sigma_V^2}\right)\right\}\right]f_V(t)dt\\
&= \log 2+\frac{1}{2}\log\left(2\pi\sigma_V^2\right)+\int_{-\infty}^{\infty}\frac{t^2}{2\sigma_V^2}f_V(t)dt-\int_{-\infty}^{\infty}\log\left\{1+\exp\left(\frac{-2\sqrt{P_X}t-2P_X}{\sigma_V^2}\right)\right\}f_V(t)dt.
\end{aligned} \tag{5}$$

The last equality in (5) holds because $f_V(t)$ is an even function. Finally, we arrive at

$$\begin{aligned}
I'(X;Y) &= H'(Y)-H'(V)\\
&= \log 2-\int_{-\infty}^{\infty} f_V(t)\log\left\{1+\exp\left(\frac{-2\sqrt{P_X}t-2P_X}{\sigma_V^2}\right)\right\}dt.
\end{aligned} \tag{6}$$

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

$$\begin{aligned}
F &\triangleq \int_{-\infty}^{\infty} f_V(t) \log\left\{1 + \exp\left(\frac{-2\sqrt{P_X}t - 2P_X}{\sigma_V^2}\right)\right\} dt \\
&= 2^{-L}\sum_{i=1}^{2^L}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma_N^2}}\exp\left(-\frac{(t-m_i)^2}{2\sigma_N^2}\right)\log\left\{1+\exp\left(\frac{-2\sqrt{P_X}t-2P_X}{\sigma_V^2}\right)\right\}dt \\
&= 2^{-L}\sum_{i=1}^{2^L}\int_{-\infty}^{\infty}\frac{e^{-\tau^2/2}}{\sqrt{2\pi}}\log\left\{1+\exp\left(\frac{-2\sqrt{P_X}\left(\tau\sigma_N+m_i\right)-2P_X}{\sigma_V^2}\right)\right\}d\tau \\
&= 2^{-L}\sum_{i=1}^{2^L}\int_{-\infty}^{\infty}\frac{e^{-\tau^2/2}}{\sqrt{2\pi}}\log\left\{1+e^{-2R\rho_i}e^{-2\phi\sqrt{R}\tau-2R}\right\}d\tau \\
&= 2^{-L}\sum_{i=1}^{2^L}\mathrm{E}\left[\log\left\{1+e^{-2R\rho_i}e^{-2\phi\sqrt{R}\tau-2R}\right\}\right] \\
&= \mathrm{E}\left[\log\left\{1+e^{-2R\rho_i}e^{-2\phi\sqrt{R}\tau-2R}\right\}\right] \qquad (7a)
\end{aligned}$$

where the third equality is obtained with a variable change $(t - m_i)/\sigma_N = \tau$ and the fourth equality with $R \triangleq P_X/\sigma_V^2$, $\rho_i \triangleq m_i/\sqrt{P_X}$, and $\phi \triangleq \sigma_N/\sigma_V$. 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

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

In the final expression (7*a*), however, ρ is also treated as a RV and the expectation is over both τ and ρ. Given the pdfs of τ and ρ, the computation of the expectation now involves numerical evaluation of a double integral. Note that in (7*a*) ρ is a discrete-valued random variable distributed according to $f_\rho(t)$, which denotes the probability distribution of $\rho = (1/\sqrt{P_X})\sum_{k=1}^{L} d_{-k}X_k$.

The expression (7*a*) can also be written as

$$F = 2^{-L}\sum_{k=1}^{2^{L-1}} \mathrm{E}\left[\log\left\{1+e^{-2R\rho_k^+}e^{-2\phi\sqrt{R}\tau-2R}\right\}+\log\left\{1+e^{2R\rho_k^+}e^{-2\phi\sqrt{R}\tau-2R}\right\}\right]$$
$$= 2^{-L}\sum_{k=1}^{2^{L-1}} \mathrm{E}\left[\log\left\{1+\left(e^{-2R\rho_k^+}+e^{2R\rho_k^+}\right)e^{-2\phi\sqrt{R}\tau-2R}+e^{-4\phi\sqrt{R}\tau-4R}\right\}\right]$$
$$= 2^{-(L-1)}\sum_{k=1}^{2^{L-1}} \mathrm{E}\left[\frac{1}{2}\log\left\{1+2\cosh\left(2R\rho_k^+\right)e^{-2\phi\sqrt{R}\tau-2R}+e^{-4\phi\sqrt{R}\tau-4R}\right\}\right]$$
$$= \mathrm{E}\left[\frac{1}{2}\log\left\{1+2\cosh\left(2R\rho_k^+\right)e^{-2\phi\sqrt{R}\tau-2R}+e^{-4\phi\sqrt{R}\tau-4R}\right\}\right] \tag{7b}$$

where ρ_k^+'s form the positive-half subset of ρ_i's, and in the last equality ρ^+ is also treated as a RV and the expectation is over both τ and ρ^+. Note that ρ^+ is a discrete-valued random variable distributed according to $2f_\rho(t)u(t)$. Notice that $\cosh(2R\rho_k^+) \geq 1$ and $\phi \leq 1$.

It is insightful to compare F with

$$F_b \triangleq \log 2 - C_b(R)$$
$$= \int_{-\infty}^{\infty} \frac{e^{-\tau^2/2}}{\sqrt{2\pi}}\log\left\{1+e^{-2\sqrt{R}\tau-2R}\right\}d\tau = \mathrm{E}\left[\log\left\{1+e^{-2\sqrt{R}\tau-2R}\right\}\right]$$
$$= \mathrm{E}\left[\frac{1}{2}\log\left\{1+e^{-2\sqrt{R}\tau-2R}\right\}^2\right] \tag{8a}$$
$$= \mathrm{E}\left[\frac{1}{2}\log\left\{1+2e^{-2\sqrt{R}\tau-2R}+e^{-4\sqrt{R}\tau-4R}\right\}\right] \tag{8b}$$

where $C_b(R)$ 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 Shamai-Laroia conjectured lower bound to $I(X;Y)$. The function F_b 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 (7*b*) and $F_b(R)$ in (8*b*), we see that if $\rho^+ = 0$ so that $\phi = 1$, then $F = F_b$, and $I'(X;Y)$ and the SLC all 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 $F = F_b$ and $I'(X;Y) = C_b(R) = I(X;Y)$.

Furthermore, that $\rho^+ \geq 0$ in (7*b*) 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_b, over all interested range of SNR?. Since it is already proved that $I'(X;Y) = \log 2 - F$, if $F \leq F_b$ for some R values, then clearly $C_b(R) = \log 2 - F_b \leq I(X;Y)$ at those SNRs, i.e., the SLC holds true at least at these SNRs.

While exact computation of (7*b*) 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

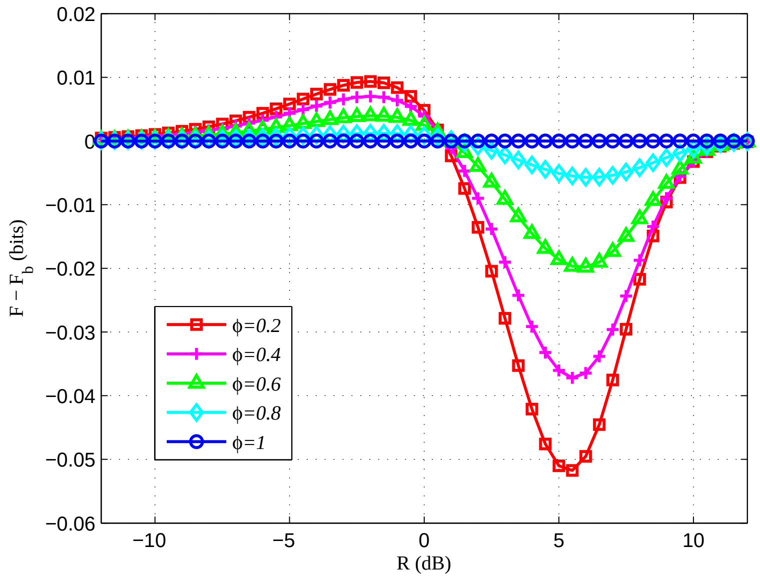


Fig. 1: $F - F_b$ as a function of R for a uniform ρ.

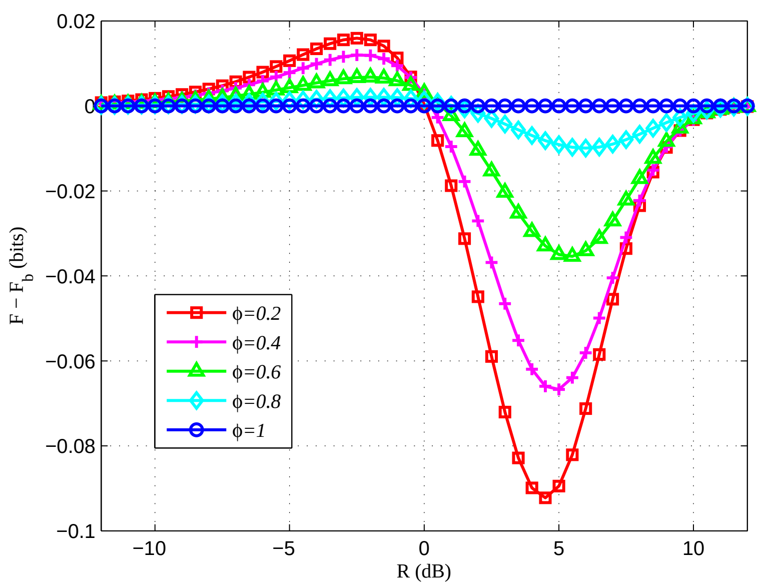


Fig. 2: $F - F_b$ as a function of R for a two-valued ρ.

know the functional form of the distribution for ρ, evaluation of (7a) or (7b) 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

$$\sigma_S^2 = P_X \sigma_\rho^2 = P_X \sum_{k=1}^{K} 2k^2\Delta^2 \frac{1}{2K+1} = \frac{P_X \Delta^2 K(K+1)}{3} = \frac{P_X \rho_{\max}(\rho_{\max}+\Delta)}{3}$$

which makes

$$\phi^2 = \frac{\sigma_N^2}{\sigma_N^2+\sigma_S^2} = \frac{\sigma_V^2-\sigma_S^2}{\sigma_V^2} = 1-\frac{\sigma_S^2}{\sigma_V^2} = 1-\frac{3P_x\rho_{\max}(\rho_{\max}+\Delta)}{\sigma_V^2} = 1-3R\rho_{\max}(\rho_{\max}+\Delta).$$

Fig. 1 shows F and F_b 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_b 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_b. Namely, F becomes smaller than F_b 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

A. Simple Bounds

Exact computation of F in general is not easy, especially when L goes to infinity. We thus resort to bounding F. 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 $\mathrm{E}\left[\frac{1}{2}\log\left(1+2\cosh(2R\rho^+)e^{-2\phi\sqrt{R}\tau}+e^{-4\phi\sqrt{R}\tau-4R}\right)\right]$ is also convex in ρ^+. Moreover, $\mathrm{E}\left[\frac{1}{2}\log\left(1+\cosh(2R\rho^+)e^{-2\phi\sqrt{R}\tau}+e^{-4\phi\sqrt{R}\tau-4R}\right)\right]$ increases as ρ^+ increases. Accordingly, we can develop bounds on F. For example, a simple upper bound is given as

$$\begin{aligned} F &= 2^{-(L-1)}\sum_{k=1}^{2^{L-1}} E\left[\frac{1}{2}\log\left\{1+2\cosh\left(2R\rho_k^+\right)e^{-2\phi\sqrt{R}\tau-2R}+e^{-4\phi\sqrt{R}\tau-4R}\right\}\right] \\ &\le 2^{-(L-1)}\sum_{k=1}^{2^{L-1}} H(\rho_k^+) = H\left(2^{-(L-1)}\sum_{k=1}^{2^{L-1}}\rho_k^+\right) = H(|\rho|_{\mathrm{avg}}) \\ &\le H(\sigma_\rho) \triangleq F^{u1} \end{aligned} \tag{9}$$

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

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

$$\begin{aligned} F &= 2^{-(L-1)}\sum_{k=1}^{2^{L-1}} \mathrm{E}\left[\frac{1}{2}\log\left\{1+2\alpha_k e^{-2\phi\sqrt{R}\tau-2R}+e^{-4\phi\sqrt{R}\tau-4R}\right\}\right] \\ &\le \mathrm{E}\left[\frac{1}{2}\log\left\{1+2\left(2^{-(L-1)}\sum_{k=1}^{2^{L-1}}\alpha_k\right)e^{-2\phi\sqrt{R}\tau-2R}+e^{-4\phi\sqrt{R}\tau-4R}\right\}\right] \\ &= \mathrm{E}\left[\frac{1}{2}\log\left\{1+2\alpha e^{-2\phi\sqrt{R}\tau-2R}+e^{-4\phi\sqrt{R}\tau-4R}\right\}\right] \end{aligned} \tag{10}$$

where $\alpha \triangleq 2^{-(L-1)}\sum_{k=1}^{2^{L-1}}\alpha_k = 2^{-(L-1)}\sum_{k=1}^{2^{L-1}}\cosh(2R\rho_k^+)$. The last expression of (10) can be further upper-bounded by replacing with α' with $\alpha' \ge \alpha$. For example, note

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

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

$$F \le \mathrm{E}\left[\frac{1}{2}\log\left\{1+2(s\sigma_\rho+1)e^{-2\phi\sqrt{R}\tau-2R}+e^{-4\phi\sqrt{R}\tau-4R}\right\}\right] \triangleq F^{u2}. \tag{11}$$

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 $\mathrm{E}\left[\log\left(1+e^{-2R\rho}e^{-2\phi\sqrt{R}\tau-2R}\right)\right]$ in ρ, write

$$\begin{aligned} F &= 2^{-L}\sum_{i=1}^{2^L}\mathrm{E}\left[\log\left\{1+e^{-2R\rho_i}e^{-2\phi\sqrt{R}\tau-2R}\right\}\right] \\ &\ge \mathrm{E}\left[\log\left\{1+\exp\left(-2R\left(2^{-L}\sum_{i=1}^{2^L}\rho_i\right)\right)e^{-2\phi\sqrt{R}\tau-2R}\right\}\right] = \mathrm{E}\left[\log\left\{1+e^{-2\phi\sqrt{R}\tau-2R}\right\}\right] \\ &= E\left[\frac{1}{2}\log\left\{1+2e^{-2\phi\sqrt{R}\tau-2R}+e^{-4\phi\sqrt{R}\tau-4R}\right\}\right] \triangleq F^l. \end{aligned} \tag{12}$$

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 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 2^M clusters of Gaussian pdfs, write

$$
\begin{aligned}
F &= 2^{-M} \sum_{j=1}^{2^M} \left(2^{-(L-M)} \sum_{i=1}^{2^{L-M}} \mathrm{E}\left[\log\left\{ 1 + e^{-2R(\mu_i+\lambda_j)} e^{-2\phi\sqrt{R}\tau - 2R} \right\} \right] \right) \\
&= 2^{-M} \sum_{j=1}^{2^M} \left(2^{-(L-M-1)} \sum_{l=1}^{2^{L-M-1}} \mathrm{E}\left[\frac{1}{2} \log\left\{ 1 + 2\cosh\left(2R\mu_l^+\right) e^{-2R\lambda_j} e^{-2\phi\sqrt{R}\tau-2R} + e^{-4R\lambda_j} e^{-4\phi\sqrt{R}\tau-4R} \right\} \right] \right) \\
&\leq 2^{-M} \sum_{j=1}^{2^M} \left(2^{-(L-M-1)} \sum_{l=1}^{2^{L-M-1}} H_j(\mu_l^+) \right) \\
&= 2^{-M} \sum_{j=1}^{2^M} H_j \left(2^{-(L-M-1)} \sum_{l=1}^{2^{L-M-1}} \mu_l^+ \right) = 2^{-M} \sum_{j=1}^{2^M} H_j(|\mu|_{\mathrm{avg}}) \\
&\leq 2^{-M} \sum_{j=1}^{2^M} H_j(\sigma_\mu) \triangleq F_M^{u1}
\end{aligned} \tag{13}
$$

where λ_j is 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; μ_l^+'s form the positive-half subset of μ_i's. The function $H_j(\theta)$ represents a straight line that passes through the two points of the convex function $\mathrm{E}\left[\frac{1}{2}\log\left\{1+2\cosh\left(2R\theta\right)e^{-2R\lambda_j}e^{-2\phi\sqrt{R}\tau-2R}+e^{-4R\lambda_j}e^{-4\phi\sqrt{R}\tau-4R}\right\}\right]$ at $\theta = 0$ and $\theta = |\mu|_{\max}$. The last inequality follows from $|\mu|_{\mathrm{avg}} \leq \sigma_\mu$, and note $\sigma_\mu = \sqrt{\sigma_\rho^2 - \sigma_\lambda^2}$ and $|\mu|_{\max} = |\rho|_{\max} - |\lambda|_{\max}$.

Another form of tightened upper bound is obtained as

$$
\begin{aligned}
F &= 2^{-M} \sum_{j=1}^{2^M} \left(2^{-(L-M-1)} \sum_{l=1}^{2^{L-M-1}} \mathrm{E}\left[\frac{1}{2} \log\left\{ 1 + 2\cosh\left(2R\mu_l^+\right) e^{-2R\lambda_j} e^{-2\phi\sqrt{R}\tau-2R} + e^{-4R\lambda_j} e^{-4\phi\sqrt{R}\tau-4R} \right\} \right] \right) \\
&= 2^{-M} \sum_{j=1}^{2^M} \left(2^{-(L-M-1)} \sum_{l=1}^{2^{L-M-1}} \mathrm{E}\left[\frac{1}{2} \log\left\{ 1 + 2\alpha_l e^{-2R\lambda_j} e^{-2\phi\sqrt{R}\tau-2R} + e^{-4R\lambda_j} e^{-4\phi\sqrt{R}\tau-4R} \right\} \right] \right) \\
&\leq 2^{-M} \sum_{j=1}^{2^M} \mathrm{E}\left[\frac{1}{2} \log\left\{ 1 + 2\left(2^{-(L-M-1)} \sum_{l=1}^{2^{L-M-1}} \alpha_l \right) e^{-2R\lambda_j} e^{-2\phi\sqrt{R}\tau-2R} + e^{-4R\lambda_j} e^{-4\phi\sqrt{R}\tau-4R} \right\} \right] \\
&= 2^{-M} \sum_{j=1}^{2^M} \mathrm{E}\left[\frac{1}{2} \log\left\{ 1 + 2\alpha_M e^{-2R\lambda_j} e^{-2\phi\sqrt{R}\tau-2R} + e^{-4R\lambda_j} e^{-4\phi\sqrt{R}\tau-4R} \right\} \right] \\
&\leq 2^{-M} \sum_{j=1}^{2^M} \mathrm{E}\left[\frac{1}{2} \log\left\{ 1 + 2\alpha'_M e^{-2R\lambda_j} e^{-2\phi\sqrt{R}\tau-2R} + e^{-4R\lambda_j} e^{-4\phi\sqrt{R}\tau-4R} \right\} \right] \\
&= 2^{-M} \sum_{j=1}^{2^M} \mathrm{E}\left[\frac{1}{2} \log\left\{ 1 + 2\left(s_M\sigma_\mu + 1\right) e^{-2R\lambda_j} e^{-2\phi\sqrt{R}\tau-2R} + e^{-4R\lambda_j} e^{-4\phi\sqrt{R}\tau-4R} \right\} \right] \triangleq F_M^{u2}
\end{aligned} \tag{14}
$$

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

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

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

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

$$
\begin{aligned}
F &= 2^{-M} \sum_{j=1}^{2^M} \left(2^{-(L-M)} \sum_{i=1}^{2^{L-M}} \mathrm{E}\left[\log\left\{ 1 + e^{-2R(\mu_i+\lambda_j)} e^{-2\phi\sqrt{R}\tau - 2R} \right\} \right] \right) \\
&= 2^{-M} \sum_{j=1}^{2^M} \left(2^{-(L-M)} \sum_{i=1}^{2^{L-M}} \mathrm{E}\left[\log\left\{ 1 + e^{-2R\mu_i} e^{-2R\lambda_j} e^{-2\phi\sqrt{R}\tau - 2R} \right\} \right] \right) \\
&\geq 2^{-M} \sum_{j=1}^{2^M} \mathrm{E}\left[\log\left\{ 1 + \exp\left[-2R \left(2^{-(L-M)} \sum_{i=1}^{2^{L-M}} \mu_i \right) \right] e^{-2R\lambda_j} e^{-2\phi\sqrt{R}\tau - 2R} \right\} \right] \\
&= 2^{-M} \sum_{j=1}^{2^M} \mathrm{E}\left[\log\left\{ 1 + e^{-2R\lambda_j} e^{-2\phi\sqrt{R}\tau - 2R} \right\} \right] \\
&= 2^{-(M-1)} \sum_{k=1}^{2^{M-1}} \mathrm{E}\left[\frac{1}{2} \log\left\{ 1 + 2\cosh\left(2R\lambda_k^+ \right) e^{-2\phi\sqrt{R}\tau - 2R} + e^{-4\phi\sqrt{R}\tau - 4R} \right\} \right] \triangleq F_M^l
\end{aligned}
\tag{15}
$$

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

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 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

$$
\begin{aligned}
f_V(t) &= 4^{-L} \sum_{i=1}^{4^L} \frac{1}{\pi\sigma_N^2} \exp\left(-\frac{|t - m_i|^2}{\sigma_N^2} \right) \\
&= 4^{-L} \sum_{i=1}^{4^L} \left\{ \frac{1}{\sqrt{\pi\sigma_N^2}} \exp\left(-\frac{\left(t^{(r)} - m_i^{(r)}\right)^2}{\sigma_N^2} \right) \frac{1}{\sqrt{\pi\sigma_N^2}} \exp\left(-\frac{\left(t^{(i)} - m_i^{(i)}\right)^2}{\sigma_N^2} \right) \right\} \\
f_G(t) &= \frac{1}{\pi\sigma_V^2} \exp\left(-\frac{|t|^2}{\sigma_V^2} \right) = \frac{1}{\sqrt{\pi\sigma_V^2}} \exp\left(-\frac{\left(t^{(r)}\right)^2}{\sigma_V^2} \right) \frac{1}{\sqrt{\pi\sigma_V^2}} \exp\left(-\frac{\left(t^{(i)}\right)^2}{\sigma_V^2} \right).
\end{aligned}
$$

Then, the SLC is

$$
\begin{aligned}
F_q &\triangleq \log 4 - C_q(R) \\
&= 2\int_{-\infty}^{\infty} \frac{e^{-\tau^2}}{\sqrt{\pi}} \log\left\{1 + e^{-2\sqrt{2R}\tau - 2R}\right\} d\tau \\
&= 2\mathrm{E}\left[\log\left\{1 + e^{-2\sqrt{2R}\tau - 2R}\right\}\right].
\end{aligned} \tag{16}
$$

Moreover, the proposed lower bounds can be derived in a similar way as

$$
\begin{aligned}
F_M^{u1} &= 4^{-M} \sum_{j=1}^{4^M} \left\{ H_j^{(r)}\left(\frac{\sigma_\mu}{\sqrt{2}}\right) + H_j^{(i)}\left(\frac{\sigma_\mu}{\sqrt{2}}\right) \right\} \\
&= 4^{-M} \sum_{j=1}^{4^M} 2H_j^{(r)}\left(\frac{\sigma_\mu}{\sqrt{2}}\right).
\end{aligned} \tag{17}
$$

where the function $H_j^{(r)}(\theta)$ represents a straight line that passes through two points of the function $\mathrm{E}\left[\frac{1}{2}\log\left\{1 + 2\cosh\left(2\sqrt{2}R\theta\right) e^{-2\sqrt{2}R\lambda_j^{(r)}} e^{-2\phi\sqrt{2R}\tau - 2R} + e^{-4\sqrt{2}R\lambda_j^{(r)}} e^{-4\phi\sqrt{2R}\tau - 4R}\right\}\right]$ at $\theta = 0$ and at $\theta = |\mu^{(r)}|_{\max} = |\mu|_{\max}/\sqrt{2}$ while the function $H_j^{(i)}(\theta)$ is a similar straight line passing through $\mathrm{E}\left[\frac{1}{2}\log\left\{1 + 2\cosh\left(2\sqrt{2}R\theta\right) e^{-2\sqrt{2}R\lambda_j^{(i)}} e^{-2\phi\sqrt{2R}\tau - 2R} + e^{-4\sqrt{2}R\lambda_j^{(i)}} e^{-4\phi\sqrt{2R}\tau - 4R}\right\}\right]$ at $\theta = 0$ and at $\theta = |\mu^{(i)}|_{\max} = |\mu|_{\max}/\sqrt{2}$. The equality in (17) holds because the pdf's of $\lambda^{(r)}$ and $\lambda^{(i)}$ (or $H_j^{(r)}(\theta)$ and $H_j^{(i)}(\theta)$) are identical. Also, note that the variance of $\mu^{(r)}$ and $\mu^{(i)}$ are equal to $\sigma_\mu^2/2$.

The other form of the bound is given as

$$
\begin{aligned}
F_M^{u2} \triangleq 4^{-M} \sum_{j=1}^{4^M} \Bigg(&\mathrm{E}\left[\frac{1}{2}\log\left\{1 + 2\left(\frac{s_M^{(r)}\sigma_\mu}{\sqrt{2}} + 1\right) e^{-2\sqrt{2}R\lambda_j^{(r)}} e^{-2\phi\sqrt{2R}\tau - 2R} + e^{-4\sqrt{2}R\lambda_j^{(r)}} e^{-4\phi\sqrt{2R}\tau - 4R}\right\}\right] \\
&+ \mathrm{E}\left[\frac{1}{2}\log\left\{1 + 2\left(\frac{s_M^{(i)}\sigma_\mu}{\sqrt{2}} + 1\right) e^{-2\sqrt{2}R\lambda_j^{(i)}} e^{-2\phi\sqrt{2R}\tau - 2R} + e^{-4\sqrt{2}R\lambda_j^{(i)}} e^{-4\phi\sqrt{2R}\tau - 4R}\right\}\right] \Bigg) \\
= 4^{-M} \sum_{j=1}^{4^M} 2\mathrm{E} &\left[\frac{1}{2}\log\left\{1 + 2\left(\frac{s_M^{(r)}\sigma_\mu}{\sqrt{2}} + 1\right) e^{-2\sqrt{2}R\lambda_j^{(r)}} e^{-2\phi\sqrt{2R}\tau - 2R} + e^{-4\sqrt{2}R\lambda_j^{(r)}} e^{-4\phi\sqrt{2R}\tau - 4R}\right\}\right]
\end{aligned} \tag{18}
$$

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 (18) holds because $s_M^{(r)} = s_M^{(i)}$ from $|\mu^{(r)}|_{\max} = |\mu^{(i)}|_{\max} = |\mu|_{\max}/\sqrt{2}$ and $\lambda^{(r)} = \lambda^{(i)}$.

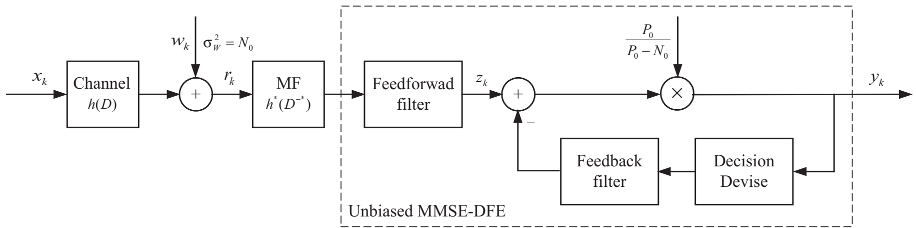


Fig. 3: System Model of ISI channels.

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

$$F_M^l \triangleq 4^{-M} \sum_{j=1}^{4^M} \left(\mathrm{E}\left[\frac{1}{2} \log \left\{ 1 + 2\cosh\left(2\sqrt{2}R\lambda_j^{(r)}\right) e^{-2\phi\sqrt{2R}\tau - 2R} + e^{-4\phi\sqrt{2R}\tau - 4R} \right\} \right] \right.$$
$$\left. + \mathrm{E}\left[\frac{1}{2} \log \left\{ 1 + 2\cosh\left(2\sqrt{2}R\lambda_j^{(i)}\right) e^{-2\phi\sqrt{2R}\tau - 2R} + e^{-4\phi\sqrt{2R}\tau - 4R} \right\} \right] \right)$$
$$= 4^{-M} \sum_{j=1}^{4^M} 2\mathrm{E}\left[\frac{1}{2} \log \left\{ 1 + 2\cosh\left(2\sqrt{2}R\lambda_j^{(r)}\right) e^{-2\phi\sqrt{2R}\tau - 2R} + e^{-4\phi\sqrt{2R}\tau - 4R} \right\} \right]. \tag{19}$$

IV. Application to ISI Channels and Numerical Examples

A. The ISI Channel and MMSE-DFE

Fig. 3 shows the system model of the finite-ISI channel with the infinite-length feedforward filter of unbiased minimum mean-squared error decision feedback equalizer (MMSE-DFE) preceded by the discrete-time matched filter (MF) for the channel. In this model, it is assumed that the receiver knows the D-transform of the finite-ISI channel response, $h(D)$, x_k is an i.i.d input sequence, and w_k is additive white Gaussian noise (AWGN) with variance $\sigma_W^2 = N_0$.

Denoting $X = x_0$, $X_k = x_k$, and $Y = y_0$, the output of the the unbiased MMSE-DFE with ideal feedback [6] 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 [6]

$$d(D) = \frac{N_0}{P_0 - N_0} \left(1 - \frac{1}{g^*(D^{-*})} \right) \tag{20}$$

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 a spectral factorization: $P_X R_{hh}(D) + N_0 = P_0 g(D) g^*(D^{-*})$ with $R_{hh}(D) = h(D)h^*(D^{-*})$. Note that a convenient numerical spectral factorization algorithm exists for recursively computing the coefficients of $g^*(D^{-*})$ [7], [8].

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

$$\begin{aligned}\sigma_V^2 &= \frac{P_X N_0}{P_0 - N_0}\\ \sigma_N^2 &= \frac{P_X P_0 N_0}{2\pi\left(P_0 - N_0\right)^2}\int_{-\pi}^{\pi}\frac{R_{hh}(e^{-j\theta})}{R_{hh}(e^{-j\theta}) + N_0/P_X}d\theta\\ \sigma_S^2 &= \sigma_V^2 - \sigma_N^2.\end{aligned}$$

We can obtain the $|\rho|_{\max}$ 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 in a form of 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,

$$\begin{aligned}|\rho|_{\max} &= \frac{1}{\sqrt{P_X}}\sum_{k=1}^{\infty}|d_{-k}X_k| = \sum_{k=1}^{\infty}|d_{-k}| = \frac{N_0}{(P_0 - N_0)}\left(\sum_{k=1}^{\infty}|a_{-k}|\right)\\ &= \frac{N_0}{(P_0 - N_0)}\left(\sum_{k=1}^{\infty}\left|\sum_{j=1}^{P}c_j p_j^k\right|\right)\\ &\le \frac{N_0}{(P_0 - N_0)}\left(\sum_{j=1}^{P}\sum_{k=1}^{\infty}\left|c_j p_j^k\right|\right)\\ &= \frac{N_0}{(P_0 - N_0)}\left(\sum_{j=1}^{P}\frac{|c_j p_j|}{1 - |p_j|}\right).\end{aligned}\tag{21}$$

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

$$\begin{aligned}|\rho|_{\max} &= \frac{N_0}{(P_0 - N_0)}\left(\sum_{k=1}^{\infty}\left|\sum_{j=1}^{P}c_j p_j^k\right|\right)\\ &= \frac{N_0}{(P_0 - N_0)}\left(\sum_{k=1}^{K}\left|\sum_{j=1}^{P}c_j p_j^k\right| + \sum_{k=K+1}^{\infty}\left|\sum_{j=1}^{P}c_j p_j^k\right|\right)\\ &\le \frac{N_0}{(P_0 - N_0)}\left(\sum_{k=1}^{K}\left|\sum_{j=1}^{P}c_j p_j^k\right| + \sum_{j=1}^{P}\sum_{k=K+1}^{\infty}\left|c_j p_j^k\right|\right)\\ &= \sum_{k=1}^{K}|d_{-k}| + \frac{N_0}{(P_0 - N_0)}\left(\sum_{j=1}^{P}\frac{\left|c_j p_j^{K+1}\right|}{1 - |p_j|}\right).\end{aligned}\tag{22}$$

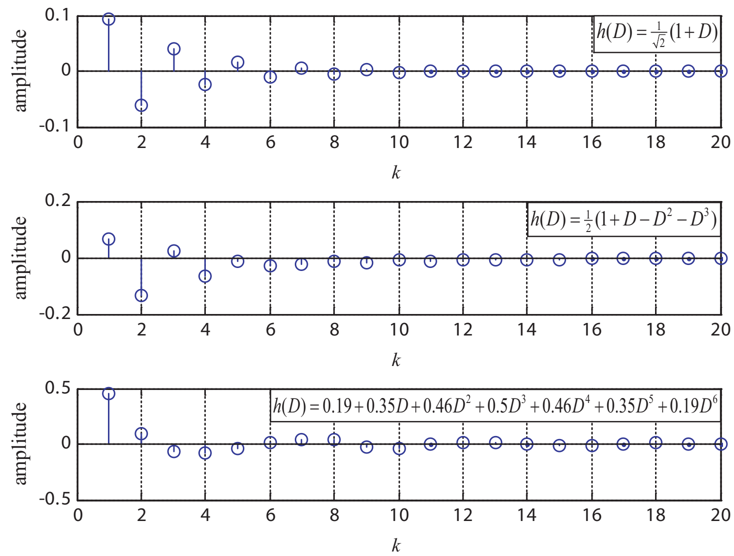


Fig. 4: Precursor taps after unbiased MMSE-DFE at SNR=10 dB for three example channels.

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 with the triangle inequality, i.e., $|a+b| \leq |a| + |b|$.

From [9], the maximum attainable capacity C (bits/channel use) for any finite-ISI channel corrupted by Gaussian noise is given as

$$
\begin{aligned}
C &\triangleq \lim_{N\to\infty} \frac{1}{2N+1} I\left(\{x_k\}_{-N}^{N}; \{r_k\}_{-N}^{N}\right) \\
&\geq \lim_{N\to\infty} \frac{1}{2N+1} I\left(\{x_k\}_{-N}^{N}; \{z_k\}_{-N}^{N}\right) \qquad (23) \\
&\geq I(X;Y) \qquad (24)
\end{aligned}
$$

where $\{u_k\}_{N_1}^{N_2} = \{u_k, k = N_1, N_1+1, \ldots, N_2\}$. The inequality in (23) holds due to the data processing theorem (equality holds if the MMSE-DFE feedforward filter is invertible [2]). The inequality of (24) can be obtained by applying the chain rule of mutual information and assuming stationarity [2].

Now, let us examine the particular ISI channels, $h(D) = \frac{1}{\sqrt{2}}(1+D)$, $h(D) = \frac{1}{2}(1+D-D^2-D^3)$, 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. [1], [2], [10]. The precursor ISI tap values are computed and shown in Fig. 4 for these example channels. In addition we include a complex-valued partial response channel: $h(D) = \frac{1}{2}\{(1+j) + (1-j)D\}$. The channel inputs are binary, except the complex-valued channel for which the inputs are assumed quaternary.

B. Numerical Results

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 $C_{SL} = \log 2 - F_b$ are considered as functions of $\mathrm{SNR} = P_X/N_0$ 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 (21) or (22). 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. 8. The SIR of each channel is also obtained using the simulation-based approach [10], [11], [12].

For each capacity figure, we first plotted the SIR and C_{SL}. 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_{SL} conjecture (this is why the C_{SL} 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_b subtracted from them. In this way, it should be clear that for those SNR values where $F^u - F_b$ becomes less than zero eventually, F is less than F_b, guaranteeing that $I'(X:Y) = \log 2 - F$ is tighter than C_{SL}. 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 that the SLC surely holds true at this SNR range. The curves of $F^l - F_b$ 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_{SL}.

Note that the computational load for evaluating the integral of (13) and (14) 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 at a reasonably low computational load) yields a bound as tight as the SLC.

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 pdf-dependent function that acts as a lower bound to the 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 reasonable 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.

REFERENCES

[1] 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. on Information Theory*, vol. 37, no. 6, pp. 1527-1539, Nov. 1991.

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

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

[4] P. H. Wittke, W. S. Smith, and L. L. Campbell, "Infinite series of interference variables with Cantor-type distributions," *IEEE Trans. Inform. Theory*, vol. 34, no. 6, pp. 1428-1436, Nov. 1988.

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

[6] J. M. Cioffi, G. P. Dudevior, M. V. Eyuboglu, and G. D. Forney, Jr., "MMSE decision-feedback equalizers and coding - Part I and II," *IEEE Trans. Commun.*, vol. 43, no. 10, pp. 2582-2604, Oct. 1995.

[7] 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.

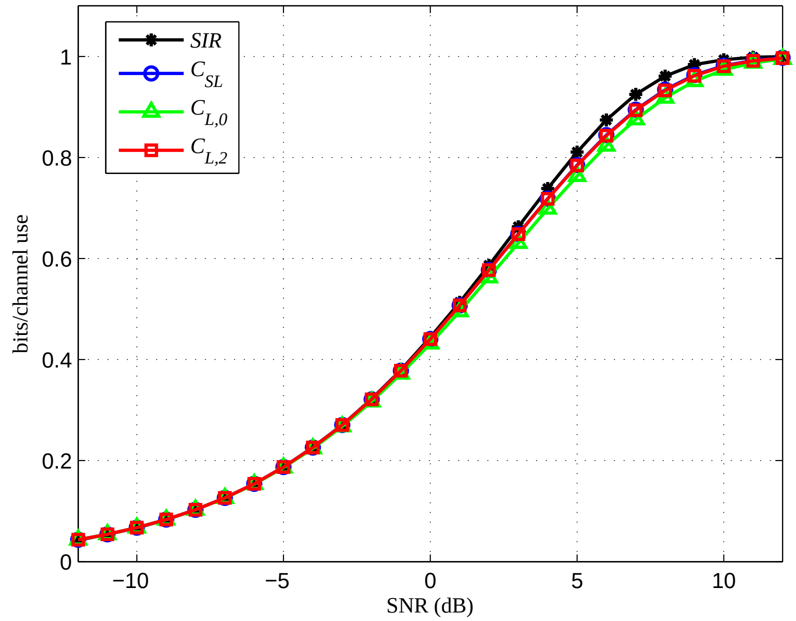
[8] G. D. Forney, Jr. and G. Ungerboeck, "Modulation and Coding for Linear Gaussian Channels," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2384-2415, Oct. 1998.

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

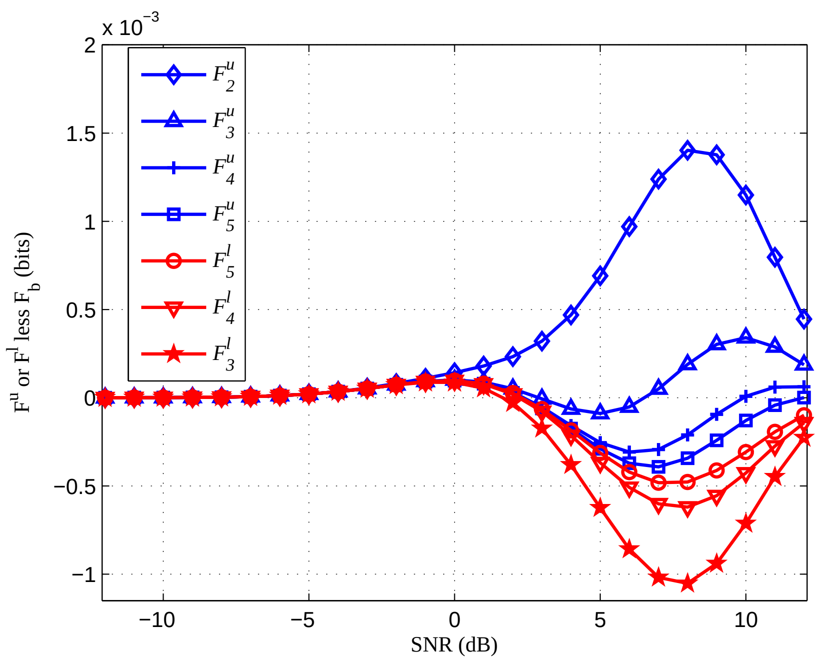
[10] D. Arnold, H. -A. Loeliger, "On the Information Rate of Binary-Input Channels with Memory," *in Proc. IEEE Int. Conf. Communications*, Helsinki, Finland, Jun. 2001, pp. 2692-2695.

[11] 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 Inform. Theory*, Washington, DC, USA, Jun. 2001, pp. 283.

[12] H. D. Pfister, J. B. Soriaga, and P. H. Siegel, "On the achievable information rates of finite state ISI channels," *In Proc. IEEE Global Telecom. Conf.*, San Antonio, Texas, USA, Nov. 2001, pp. 2992-2996.

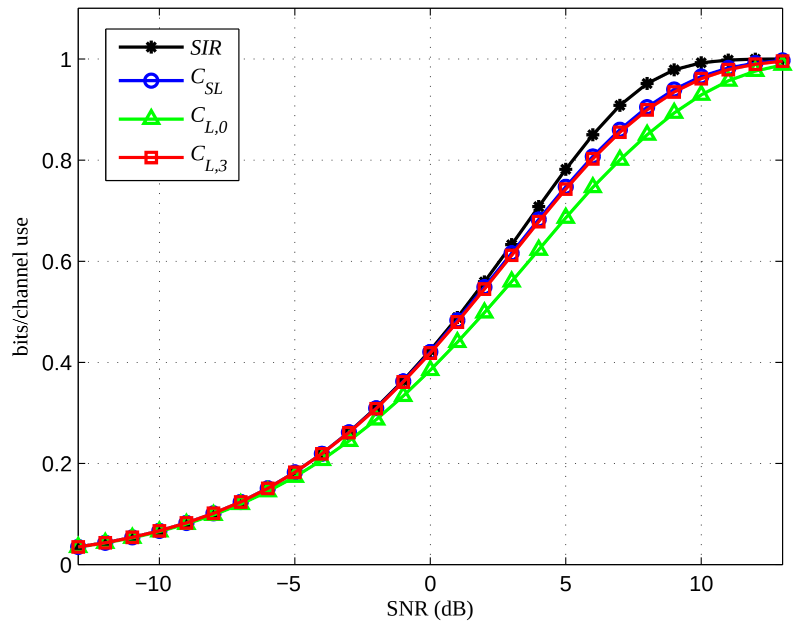


(a)

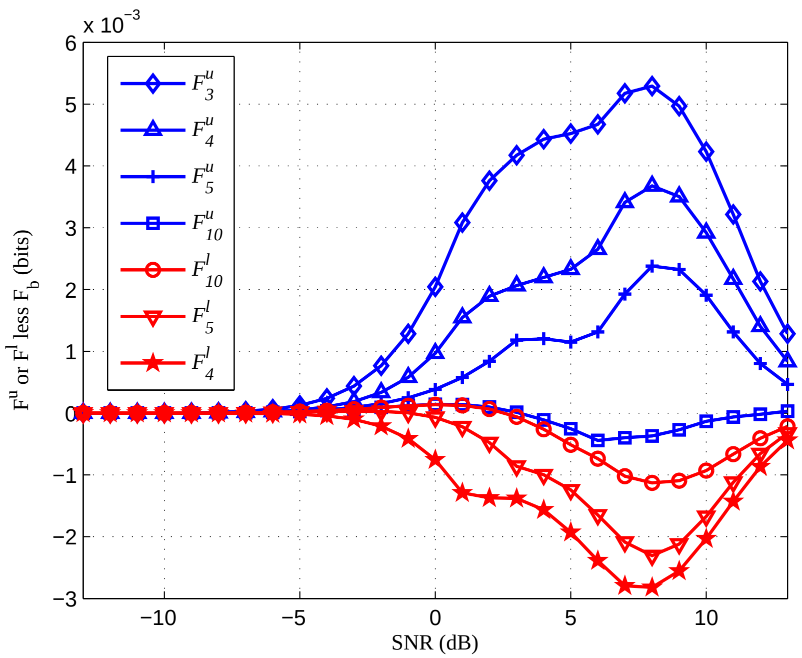


(b)

Fig. 5: Example channel: $h(D) = \frac{1}{\sqrt{2}}(1+D)$ 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_b, plotted against SNR.

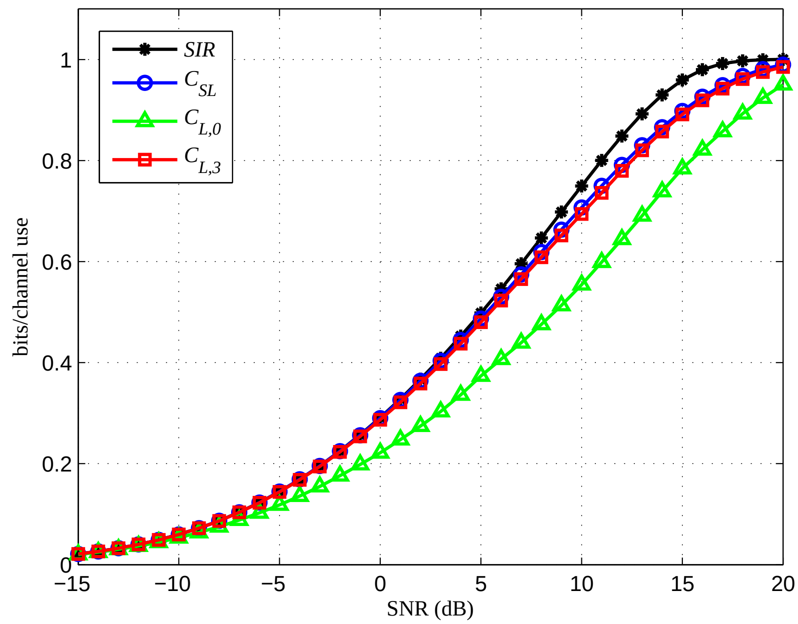


(a)

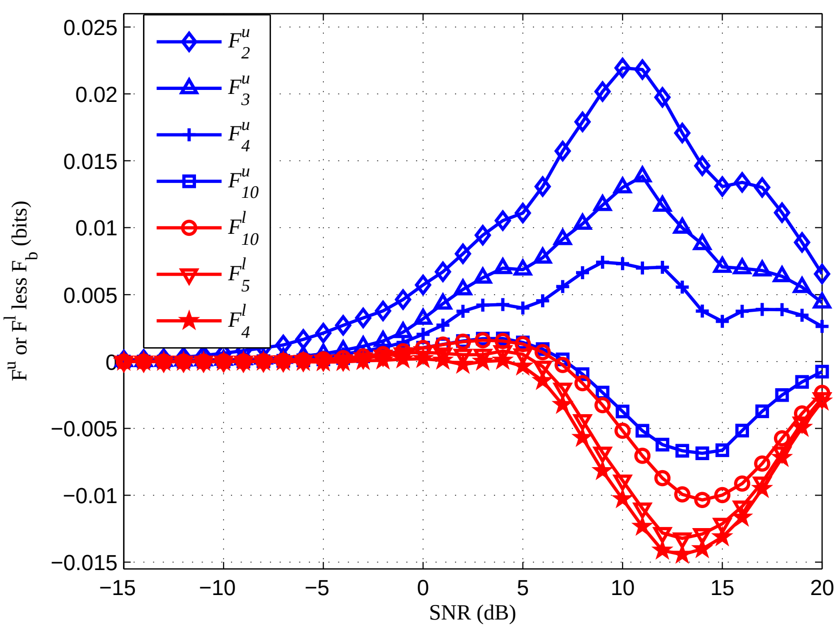


(b)

Fig. 6: Example channel: $h(D) = \frac{1}{2}(1 + D - D^2 - D^3)$ 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_b, plotted against SNR.

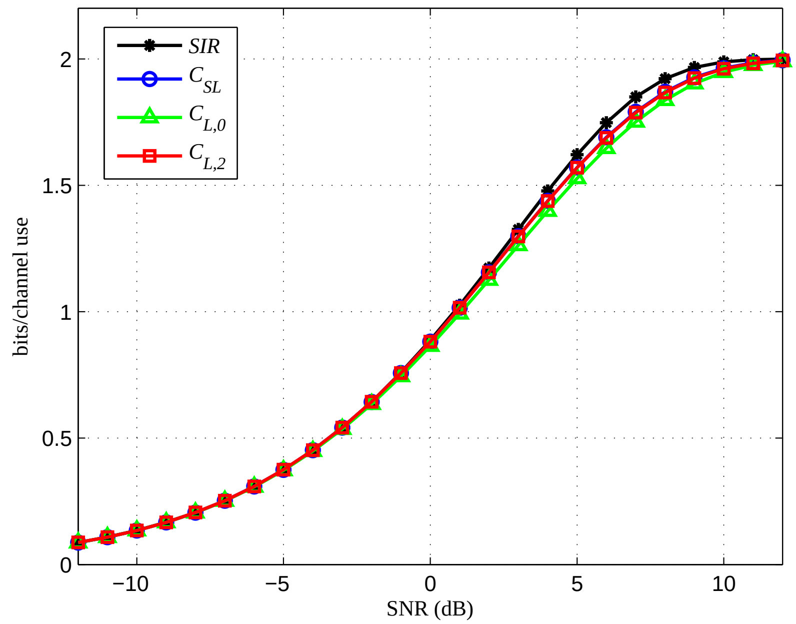


(a)

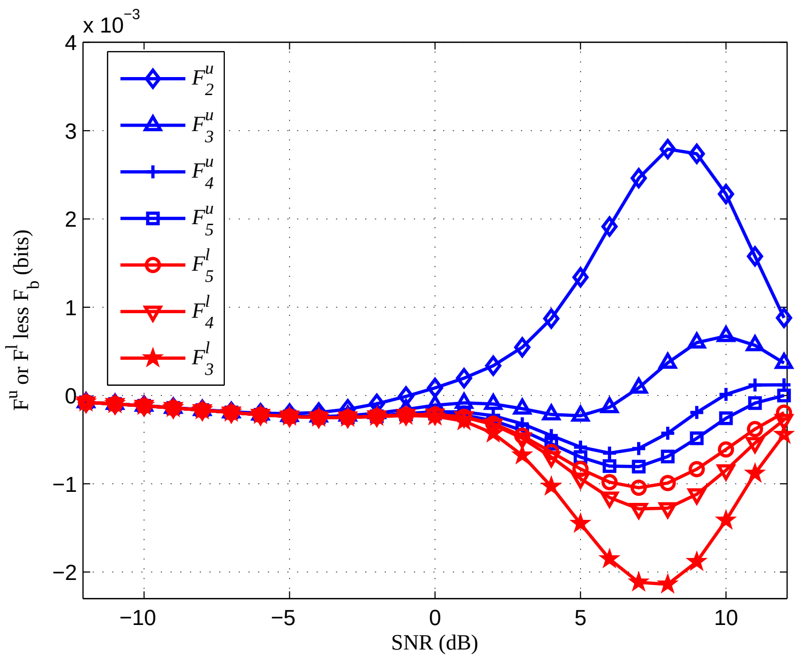


(b)

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_b, plotted against SNR.



(a)



(b)

Fig. 8: Example channel: $h(D) = \frac{1}{2}\{(1+j)+(1-j)D\}$ 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_b, plotted against SNR.