An asymptotic approximation of the ISI channel capacity

Giorgio Taricco
Politecnico di Torino
Dept. of Electronics and Telecommunications
Corso Duca degli Abruzzi, 24, 10129, Torino, Italy
Email: taricco@polito.it

Joseph J. Boutros
Texas A& M University
Electrical Engineering Dept.
23874 Doha, Qatar
Email: boutros@tamu.edu

***Abstract*—An asymptotic method to calculate the information rate of an ISI channel is presented in this work. The method is based on an integral representation of the mutual information, which is then calculated by using a saddlepoint approximation along with an asymptotic expansion stemming from the Hubbard-Stratonovich transform. This asymptotic result is evaluated repeatedly to generate a large number of samples required for the Monte-Carlo approximation of the final result. The proposed method has the advantage of being manageable even when the channel memory becomes very large since the complexity grows with polynomial order in the memory length.**

***Index Terms*—Intersymbol interference channels. Information rates. Channel capacity.**

I. Introduction

Intersymbol interference (ISI) channels have drawn attention of the research community for many years. Several methods have been proposed to derive their capacity, starting from the seminal works by Hirt and Massey [1], [2]. The goal of this work is investigating analytic and simulation techniques for the calculation of the information rate of ISI channels by using asymptotic techniques based on the saddlepoint approximation. Sparse ISI channels are used to model a wide range of communication systems, such as underwater acoustic, aeronautical and satellite systems. First, we provide a brief survey of the relevant literature results, and then we highlight the contributions of this work.

A. Known results about ISI channels

The calculation of the capacity of an ISI channel has been addressed many times in the literature. Hirt and Massey were among the first researchers to study this problem in [1], [2]. They considered the ISI channel model

$$y_n = \sum_{k=0}^{K-1} h_k x_{n-k} + z_n, \quad n \in \mathbb{Z}, \tag{1}$$

where x_n is a stationary sequence of identically distributed channel input symbols, y_n is the corresponding channel output observation sequence, h_k is the gain sequence characterizing the ISI channel, and z_n is a sequence of zero-mean independent Gaussian noise sample with given variance. Then, they showed that the mutual information rate of this ISI channel model is equivalent asymptotically (as $N \to \infty$) to the mutual information rate of a corresponding vector channel model:

$$\boldsymbol{y}_N = \boldsymbol{H}_N \boldsymbol{x}_N + \boldsymbol{z}_N, \tag{2}$$

where $\boldsymbol{x}_N = (x_1 \ldots, x_N)^\mathsf{T}$, $\boldsymbol{y}_N = (y_1 \ldots, y_N)^\mathsf{T}$, $\boldsymbol{H}_N = (h_{i-j}\mathbf{1}_{0\le i-j<K})_{i,j=1}^N$, and $\boldsymbol{z}_N = (z_1 \ldots, z_N)^\mathsf{T}$. In particular, we have:

$$I = \lim_{N\to\infty} \frac{1}{N} I(\boldsymbol{x}_N; \boldsymbol{y}_N). \tag{3}$$

This problem finds application in several fields of communications, such as the transmission of pulse amplitude modulated (PAM) signals through a dispersive Gaussian channel, which is often the case of telephone lines and in magnetic recording systems, where different types of filters are used by the detector [3].

The main result obtained in [2] refers to the capacity of the ISI channel when the input symbol can be arbitrarily distributed with an average power constraint. In the case of a real ISI channel, they showed that the capacity is obtained as $C = \lim_{N\to\infty} C_N$ where

$$C_N \triangleq \frac{1}{2N} \sum_{n=0}^{N-1} \log \max(1, \mu |H_n|^2), \tag{4}$$

where $H_n = \sum_{k=0}^{K-1} h_k \mathrm{e}^{-\mathrm{j}\,2\pi kn/N}$, the discrete Fourier transform (DFT) of the channel gain sequence, and μ is obtained by *water-filling*, i.e., by solving the equation

$$\frac{P_x}{\sigma_z^2} = \frac{1}{N} \sum_{n=0}^{N-1} \max(0, \mu - |H_n|^{-2}), \tag{5}$$

where P_x is the upper limit to the average input power and σ_z^2 is the noise sample variance. The above results can also be expressed in the following asymptotic form:

$$C = \frac{1}{2\pi} \int_0^\pi \log \max[1, \mu |H(\lambda)|^2]\, d\lambda \tag{6}$$

where $H(\lambda) = \sum_{k=0}^{K-1} h_k \mathrm{e}^{-\mathrm{j}\,k\lambda}$ and μ is derived by solving the equation

$$\frac{P_x}{\sigma_z^2} = \frac{1}{\pi} \int_0^\pi \max[0, \mu - |H(\lambda)|^{-2}]\, d\lambda. \tag{7}$$

*The work of Giorgio Taricco and Joseph Boutros was supported by QNRF, a member of Qatar Foundation, project NPRP 5-600-2-242.

Hirt and Massey also provide the capacity achieving distribution of the input, which corresponds to a correlated zero-mean stationary random Gaussian sequence characterized by the covariances:

$$\mathbb{E}[x_{n+k}x_n] = \frac{\sigma_z^2}{\pi}\int_0^\pi \max[0, \mu - |H(\lambda)|^{-2}]\cos(k\lambda)\,d\lambda. \quad (8)$$

Thus, the problem of determining the channel capacity is completely solved in the case of unrestricted input distribution but remains open for distributions over a discrete input alphabet.

Obviously, the previous results represent an upper limit to the achievable rate with discrete input. Hirt proposed in his Ph.D. Thesis [1] the use of a Monte-Carlo technique to calculate the multi-dimensional integral involved in the calculation of the information rate with independent equiprobable binary PAM input symbols. The technique has the disadvantage of getting more and more computationally intense as the length of the channel gain vector $\boldsymbol{h} = (h_0, \ldots, h_{K-1})$ increases.

Motivated by previous studies, several lower and upper bound to the achievable rate with discrete input symbols have been obtained in the literature assuming the channel vector has finite ℓ_2 norm (i.e., $\|\boldsymbol{h}\| < \infty$) [3], [4]. The main ones are reported as follows, where I_{iid} denotes the achievable rate for iid input symbols.

- The following lower bound holds:

$$I_{\mathsf{iid}} \geq I(X; \rho X + Z), \quad (9)$$

  where X is a random variable taking the same values with the same discrete probability distribution as the channel input sequence x_n, Z is a Gaussian random variable with zero mean and variance σ_z^2 as the noise samples in the ISI channel model, and

$$\rho \triangleq \exp\left\{\frac{1}{\pi}\int_0^\pi \ln|H(\lambda)|\,d\lambda\right\} \leq \|\boldsymbol{h}\|. \quad (10)$$

  The parameter ρ is defined as the *degradation factor* because it represents how the channel memory introduced by the ISI reduces the total received power gain represented by the norm $\|\boldsymbol{h}\|$.
- The *matched filter* upper bound holds:

$$I_{\mathsf{iid}} \leq I(X; \|\boldsymbol{h}\|X + Z), \quad (11)$$

  where, again, X is a random variable taking the same values with the same discrete probability distribution as the channel input sequence x_n and Z is a Gaussian random variable with zero mean and variance σ_z^2.
- Another upper bound can be derived by considering the capacity corresponding to iid Gaussian input symbols with average power P_x:

$$I_{\mathsf{iid}} \leq \frac{1}{2\pi}\int_0^\pi \log[1 + (P_x/\sigma_z^2)|H(\lambda)|^2]\,d\lambda. \quad (12)$$

- In another work [4], Shamai and Laroia conjectured the lower bound

$$I_{\mathsf{iid}} \geq I\left(x_0; x_0 + \sum_{k\geq 1}\alpha_k x_k + \tilde{z}\right), \quad (13)$$

  where

$$\alpha_k = \frac{\sum_\ell a_\ell h_{-\ell-k}}{\sum_\ell a_\ell h_{-\ell}}, \quad (14)$$

  the coefficients a_ℓ are arbitrarily chosen, and $\tilde{z} \sim \mathcal{N}(0, \sigma_z^2 \sum_\ell a_\ell^2/(\sum_\ell a_\ell)^2)$. The authors of [4] also conjectured that the lower bound could be further lower bounded by assuming the noise term $\sum_{k\geq 1}\alpha_k x_k + \tilde{z}$ to be Gaussian distributed. Recently, however, this conjecture has been disproved in [5].

The binary-input additive Gaussian noise channel mutual information can be calculated as illustrated in Appendix A.

B. BCJR algorithm approach

A simulation-based approach applicable in the case of independent uniformly distributed binary inputs has been proposed in [6]–[8]. This approach relies on the Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm [9] and is based on a trellis whose length corresponds to the memory of the ISI channel. As a result, complexity increases exponentially with the channel memory and its application becomes quickly unfeasible in the case of sparse ISI channels where the memory is long while the meaningful tap number is small.

C. Contributions of this work

The main contribution of this paper is an approximation of the mutual information of an ISI channel based on an asymptotic approach. The approximation is based on the following steps. First, a suitable integral representation of the mutual information is developed. Next, the integrals are interpreted as expectations and the inner expectation is evaluated according to an asymptotic approach based on the saddlepoint approximation and on the Hubbard-Stratonovich transform. The results are then averaged by using Monte-Carlo simulation.

II. Integral representation of the mutual information

In this section we develop an integral representation for the mutual information of the ISI channel which is suitable to being approximated by an asymptotic approach. This representation is interpretable as the concatenation of two averages: the inner average is approximated asymptotically; the outer average is approximated by Monte-Carlo simulation. We start by introducing the main definition for the ISI channel in order to carry out the further developments.

We consider a real ISI channel described by the following equation:

$$y_n = \sum_{k=0}^{L-1} h_k x_{n-k} + z_n. \quad (15)$$

This sequential representation can be cast into a block matrix representation of the ISI channel which is used in the further developments. To this purpose we assume that the input symbols satisfy a circular condition over a block of length N. This condition is expressed by

$$x_{-n} = x_{N-n} \quad (16)$$

for $n = 1, \ldots, L$. In this way, the input symbols $(x_{-L}, \ldots, x_0, \ldots, x_{N-1})$ are transmitted but only the symbols $(x_0, \ldots, x_{N-1})$ are information-bearing.

Setting

$$\boldsymbol{x} \triangleq \begin{pmatrix} x_{N-1} \\ x_{N-2} \\ \vdots \\ x_0 \end{pmatrix}, \ \boldsymbol{y} \triangleq \begin{pmatrix} y_{N-1} \\ y_{N-2} \\ \vdots \\ y_0 \end{pmatrix}, \ \boldsymbol{z} \triangleq \begin{pmatrix} z_{N-1} \\ z_{N-2} \\ \vdots \\ z_0 \end{pmatrix}, \tag{17}$$

and defining the matrix

$$\boldsymbol{H} \triangleq \begin{pmatrix} h_0 & h_1 & \ldots & h_{L-1} & 0 & 0 & \ldots & 0 \\ 0 & h_0 & h_1 & \ldots & h_{L-1} & 0 & \ldots & 0 \\ \vdots & & & \ddots & \ddots & & & \vdots \\ h_1 & \ldots & h_{L-1} & 0 & \ldots & \ldots & 0 & h_0 \end{pmatrix}, \tag{18}$$

we can write the block-wise ISI channel equation as follows:

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{z}. \tag{19}$$

Assuming the elements of the noise vector $\boldsymbol{z}$ to be independent and Gaussian distributed as $\mathcal{N}(0, \sigma^2)$, we can see that the conditional channel output vector distribution is given by

$$p(\boldsymbol{y}|\boldsymbol{x}) = (2\pi\sigma^2)^{-N/2} \mathrm{e}^{-\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}\|^2/(2\sigma^2)}. \tag{20}$$

Then, the marginal output distribution turns out to be

$$p(\boldsymbol{y}) = (2\pi\sigma^2)^{-N/2} \mathbb{E}_{\boldsymbol{x}} \left[\mathrm{e}^{-\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}\|^2/(2\sigma^2)} \right] \tag{21}$$

and, correspondingly, the output entropy can be represented by

$$h(\boldsymbol{y}) = (N/2)\ln(2\pi\sigma^2) - \mathbb{E}_{\boldsymbol{y}} \left[\ln \mathbb{E}_{\boldsymbol{x}} \left[\mathrm{e}^{-\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}\|^2/(2\sigma^2)} \right] \right]. \tag{22}$$

Since $h(\boldsymbol{y}|\boldsymbol{x}) = h(\boldsymbol{z}) = (N/2)\ln(2\pi\mathrm{e}\sigma^2)$, we obtain

$$I(\boldsymbol{x}; \boldsymbol{y}) = -(N/2) - \mathbb{E}_{\boldsymbol{y}} \left[\ln \mathbb{E}_{\boldsymbol{x}} \left[\mathrm{e}^{-\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}\|^2/(2\sigma^2)} \right] \right]. \tag{23}$$

The key problem is therefore calculating the double expectation reported in the above equation.

A. Asymptotic calculation of the inner expectation

To calculate the inner expectation present in eq. (23) we apply Theorem B.1 with $m = N, n = 1$. We get the results in eq. (24) on page 4, where we split the channel matrix as $\boldsymbol{H} = (\boldsymbol{h}_1, \ldots, \boldsymbol{h}_N)$, and defined the auxiliary function

$$\phi(\boldsymbol{w}, \boldsymbol{z}) \triangleq \boldsymbol{w}^{\mathsf{T}}\boldsymbol{z} - \frac{(\boldsymbol{w}+\boldsymbol{z})^{\mathsf{T}}\boldsymbol{y}}{\sqrt{2}\sigma} + \sum_{i=1}^{N} \ln\cosh\left[\frac{(\boldsymbol{w}+\boldsymbol{z})^{\mathsf{T}}\boldsymbol{h}_i}{\sqrt{2}\sigma}\right]. \tag{25}$$

In order to calculate the saddlepoint approximation we derive the first-order variation of the function $\phi(\boldsymbol{w}, \boldsymbol{z})$. We obtain:

$$\begin{aligned} \delta\phi &\triangleq \frac{\partial}{\partial t}\phi(\boldsymbol{w}+t\delta\boldsymbol{w}, \boldsymbol{z}+t\delta\boldsymbol{z})\Big|_{t=0} \\ &= \delta\boldsymbol{w}^{\mathsf{T}}\left\{\boldsymbol{z} - \frac{\boldsymbol{y}}{\sqrt{2}\sigma} + \sum_{i=1}^{N}\tanh\left[\frac{(\boldsymbol{w}+\boldsymbol{z})^{\mathsf{T}}\boldsymbol{h}_i}{\sqrt{2}\sigma}\right]\frac{\boldsymbol{h}_i}{\sqrt{2}\sigma}\right\} \\ &+ \delta\boldsymbol{z}^{\mathsf{T}}\left\{\boldsymbol{w} - \frac{\boldsymbol{y}}{\sqrt{2}\sigma} + \sum_{i=1}^{N}\tanh\left[\frac{(\boldsymbol{w}+\boldsymbol{z})^{\mathsf{T}}\boldsymbol{h}_i}{\sqrt{2}\sigma}\right]\frac{\boldsymbol{h}_i}{\sqrt{2}\sigma}\right\} \end{aligned} \tag{26}$$

The solution of the saddlepoint equation $\delta\phi = 0$ leads to $\boldsymbol{z} = \boldsymbol{w} = \check{\boldsymbol{w}}$, where $\check{\boldsymbol{w}}$ is the solution of

$$\boldsymbol{w} = \frac{\boldsymbol{y}}{\sqrt{2}\sigma} - \sum_{i=1}^{N}\tanh\left(\frac{\sqrt{2}\boldsymbol{w}^{\mathsf{T}}\boldsymbol{h}_i}{\sigma}\right)\frac{\boldsymbol{h}_i}{\sqrt{2}\sigma}. \tag{27}$$

Additionally, we calculate the second-order variation of the function $\phi(\boldsymbol{w}, \boldsymbol{z})$. We obtain:

$$\begin{aligned} \delta^2\phi &\triangleq \frac{\partial^2}{\partial t^2}\phi(\boldsymbol{w}+t\delta\boldsymbol{w}, \boldsymbol{z}+t\delta\boldsymbol{z})\Big|_{t=0} \\ &= 2\delta\boldsymbol{w}^{\mathsf{T}}\delta\boldsymbol{z} + \frac{1}{2\sigma^2}\sum_{i=1}^{N}\frac{[(\delta\boldsymbol{w}+\delta\boldsymbol{z})^{\mathsf{T}}\boldsymbol{h}_i]^2}{\cosh^2[(\boldsymbol{w}+\boldsymbol{z})^{\mathsf{T}}\boldsymbol{h}_i/(\sqrt{2}\sigma)]}. \end{aligned} \tag{28}$$

This result will be used to expand the second-order approximation of the function $\phi(\boldsymbol{w}, \boldsymbol{z})$ around the saddlepoint and calculate the corresponding saddlepoint integral.

B. Saddlepoint integral

In correspondence of the saddlepoint we can write the second-order approximation:

$$\phi(\boldsymbol{w}, \boldsymbol{z}) \approx \phi(\check{\boldsymbol{w}}, \check{\boldsymbol{w}}) + \frac{1}{2}\delta^2\phi. \tag{29}$$

The steepest descent contour from the saddlepoint can be obtained by setting the imaginary part of $\delta^2\phi$ equal to 0. This is achieved when $\delta\boldsymbol{z} = -\delta\boldsymbol{w}^*$ and leads to

$$\frac{1}{2}\delta^2\phi = -\|\delta\boldsymbol{w}\|^2 - \frac{1}{\sigma^2}\sum_{i=1}^{N}\frac{[(\mathsf{Im}\delta\boldsymbol{w})^{\mathsf{T}}\boldsymbol{h}_i]^2}{\cosh^2(\sqrt{2}\check{\boldsymbol{w}}^{\mathsf{T}}\boldsymbol{h}_i/\sigma)}. \tag{30}$$

Treating the variation vectors $\delta\boldsymbol{w}$ and $\delta\boldsymbol{z}$ as integration variables and using the condition $\delta\boldsymbol{z} = -\delta\boldsymbol{w}^*$ we can apply the standard rules on differential form to obtain:

$$\begin{aligned} & d(\delta w_i) \wedge d(\delta z_i) \\ &= [d(\mathsf{Re}\delta w_i) + \mathrm{j}\, d(\mathsf{Im}\delta w_i)] \wedge [-d(\mathsf{Re}\delta w_i) + \mathrm{j}\, d(\mathsf{Im}\delta w_i)] \\ &= 2\mathrm{j}\, d(\mathsf{Re}\delta w_i) \wedge d(\mathsf{Im}\delta w_i). \end{aligned} \tag{31}$$

This leads to

$$\begin{aligned} & \int \mathrm{e}^{\frac{1}{2}\delta^2\phi} d\mu(\delta\boldsymbol{w}, \delta\boldsymbol{z}) \\ &= \det\left\{\boldsymbol{I} + \frac{1}{\sigma^2}\sum_{i=1}^{N}\frac{\boldsymbol{h}_i\boldsymbol{h}_i^{\mathsf{T}}}{\cosh^2(\sqrt{2}\check{\boldsymbol{w}}^{\mathsf{T}}\boldsymbol{h}_i/\sigma)}\right\}^{-1/2} \end{aligned} \tag{32}$$

and yields the asymptotic approximation

$$\begin{aligned} & \ln \mathbb{E}_{\boldsymbol{x}}\left[\mathrm{e}^{-\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}\|^2/(2\sigma^2)}\right] \\ &\simeq \|\check{\boldsymbol{w}}\|^2 - \frac{\sqrt{2}\check{\boldsymbol{w}}^{\mathsf{T}}\boldsymbol{y}}{\sigma} + \sum_{i=1}^{N}\ln\cosh\left(\frac{\sqrt{2}\check{\boldsymbol{w}}^{\mathsf{T}}\boldsymbol{h}_i}{\sigma}\right) \\ &\quad -\frac{1}{2}\ln\det\left\{\boldsymbol{I} + \frac{1}{\sigma^2}\sum_{i=1}^{N}\frac{\boldsymbol{h}_i\boldsymbol{h}_i^{\mathsf{T}}}{\cosh^2(\sqrt{2}\check{\boldsymbol{w}}^{\mathsf{T}}\boldsymbol{h}_i/\sigma)}\right\}. \end{aligned} \tag{33}$$

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{x}}\left[\mathrm{e}^{-\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}\|^2/(2\sigma^2)}\right] &= \mathbb{E}_{\boldsymbol{x}}\left[\int_{\mathcal{D}_0}\exp\{\boldsymbol{w}^{\mathsf{T}}\boldsymbol{z}-(\boldsymbol{w}+\boldsymbol{z})^{\mathsf{T}}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})/(\sqrt{2}\sigma)\}d\mu(\boldsymbol{w},\boldsymbol{z})\right] \\
&= \int_{\mathcal{D}_0}\exp[\boldsymbol{w}^{\mathsf{T}}\boldsymbol{z}-(\boldsymbol{w}+\boldsymbol{z})^{\mathsf{T}}\boldsymbol{y}/(\sqrt{2}\sigma)]\mathbb{E}_{\boldsymbol{x}}[\exp[(\boldsymbol{w}+\boldsymbol{z})^{\mathsf{T}}\boldsymbol{H}\boldsymbol{x}/(\sqrt{2}\sigma)]d\mu(\boldsymbol{w},\boldsymbol{z}) \\
&= \int_{\mathcal{D}_0}\exp[\boldsymbol{w}^{\mathsf{T}}\boldsymbol{z}-(\boldsymbol{w}+\boldsymbol{z})^{\mathsf{T}}\boldsymbol{y}/(\sqrt{2}\sigma)]\prod_{i=1}^{N}\cosh[(\boldsymbol{w}+\boldsymbol{z})^{\mathsf{T}}\boldsymbol{h}_i/(\sqrt{2}\sigma)]d\mu(\boldsymbol{w},\boldsymbol{z}) \\
&= \int_{\mathcal{D}_0}\mathrm{e}^{\phi(\boldsymbol{w},\boldsymbol{z})}d\mu(\boldsymbol{w},\boldsymbol{z}), \qquad (24)
\end{aligned}
$$

The previous results can be summarized by the following asymptotic approximation of the mutual information:

$$
\begin{aligned}
I(\boldsymbol{x};\boldsymbol{y}) &= -(N/2)-\mathbb{E}_{\boldsymbol{y}}\left[\ln\mathbb{E}_{\boldsymbol{x}}\left[\mathrm{e}^{-\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}\|^2/(2\sigma^2)}\right]\right] \\
&\simeq -\frac{N}{2}-\mathbb{E}_{\boldsymbol{y}}\left[\|\check{\boldsymbol{w}}\|^2-\frac{\sqrt{2}\check{\boldsymbol{w}}^{\mathsf{T}}\boldsymbol{y}}{\sigma}\right. \\
&\quad +\sum_{i=1}^{N}\ln\cosh\left(\frac{\sqrt{2}\check{\boldsymbol{w}}^{\mathsf{T}}\boldsymbol{h}_i}{\sigma}\right) \\
&\quad \left.-\frac{1}{2}\ln\det\left\{\boldsymbol{I}+\frac{1}{\sigma^2}\sum_{i=1}^{N}\frac{\boldsymbol{h}_i\boldsymbol{h}_i^{\mathsf{T}}}{\cosh^2(\sqrt{2}\check{\boldsymbol{w}}^{\mathsf{T}}\boldsymbol{h}_i/\sigma)}\right\}\right], \qquad (34)
\end{aligned}
$$

where $\check{\boldsymbol{w}}$ is the vector satisfying the equation

$$
\check{\boldsymbol{w}}=\frac{\boldsymbol{y}}{\sqrt{2}\sigma}-\sum_{i=1}^{N}\tanh\left(\frac{\sqrt{2}\check{\boldsymbol{w}}^{\mathsf{T}}\boldsymbol{h}_i}{\sigma}\right)\frac{\boldsymbol{h}_i}{\sqrt{2}\sigma}. \qquad (35)
$$

1) Existence of the solution: It can be shown that the solution of eq. (35) always exists by using Brouwer's Fixed-Point Theorem [10].

$$
\boldsymbol{\varphi}(\boldsymbol{w})\triangleq\frac{\boldsymbol{y}}{\sqrt{2}\sigma}-\sum_{i=1}^{N}\tanh\left(\frac{\sqrt{2}\boldsymbol{w}^{\mathsf{T}}\boldsymbol{h}_i}{\sigma}\right)\frac{\boldsymbol{h}_i}{\sqrt{2}\sigma}. \qquad (36)
$$

It is plain to check that

$$
|(\boldsymbol{\varphi}(\boldsymbol{w}))_j|\leq\frac{|(\boldsymbol{y})_j|+\sum_{i=1}^{N}|(\boldsymbol{h}_i)_j|}{\sqrt{2}\sigma}. \qquad (37)
$$

Therefore, the codomain of $\boldsymbol{\varphi}(\cdot)$ is a subset of a closed bounded set, hence it is compact so that Brouwer's Fixed-Point Theorem applies and guarantees the existence of at least one fixed-point.

2) Newton's method: In order to find the saddlepoint vector $\check{\boldsymbol{w}}$ satisfying the saddlepoint equation (27) we resort to Newton's method. We can rewrite the saddlepoint equation (27) as

$$
\boldsymbol{f}(\boldsymbol{w})\triangleq\boldsymbol{w}-\frac{\boldsymbol{y}}{\sqrt{2}\sigma}+\sum_{i=1}^{N}\tanh\left(\frac{\sqrt{2}\boldsymbol{w}^{\mathsf{T}}\boldsymbol{h}_i}{\sigma}\right)\frac{\boldsymbol{h}_i}{\sqrt{2}\sigma}=\boldsymbol{0}. \qquad (38)
$$

Then, Newton's method consists in solving the first-order approximation

$$
\boldsymbol{f}(\boldsymbol{w}+\delta\boldsymbol{w})=\boldsymbol{f}(\boldsymbol{w})+\nabla\boldsymbol{f}(\boldsymbol{w})\delta\boldsymbol{w}=\boldsymbol{0} \qquad (39)
$$

for $\delta\boldsymbol{w}$, which yields

$$
\delta\boldsymbol{w}=-[\nabla\boldsymbol{f}(\boldsymbol{w})]^{-1}\boldsymbol{f}(\boldsymbol{w}). \qquad (40)
$$

Here, $\nabla\boldsymbol{f}(\boldsymbol{w})$ is the matrix of the first derivatives of the vector function $\boldsymbol{f}(\boldsymbol{w})$ with respect to the elements of the vector $\boldsymbol{w}$. We can calculate it by finding the variation of $\boldsymbol{f}(\boldsymbol{w})$ and obtain, after some algebra:

$$
\nabla\boldsymbol{f}(\boldsymbol{w})=\boldsymbol{I}_N+\frac{1}{\sigma^2}\sum_{i=1}^{N}\frac{\boldsymbol{h}_i\boldsymbol{h}_i^{\mathsf{T}}}{\cosh^2(\sqrt{2}\boldsymbol{w}^{\mathsf{T}}\boldsymbol{h}_i/\sigma)}. \qquad (41)
$$

Thus, the vector perturbation $\delta\boldsymbol{w}$ can be derived as follows:

$$
\begin{aligned}
\delta\boldsymbol{w} &\approx \left[\boldsymbol{I}_N+\frac{1}{\sigma^2}\sum_{i=1}^{N}\frac{\boldsymbol{h}_i\boldsymbol{h}_i^{\mathsf{T}}}{\cosh^2(\sqrt{2}\boldsymbol{w}^{\mathsf{T}}\boldsymbol{h}_i/\sigma)}\right]^{-1} \\
&\quad \left[\frac{\boldsymbol{y}}{\sqrt{2}\sigma}-\boldsymbol{w}-\sum_{i=1}^{N}\tanh\left(\frac{\sqrt{2}\boldsymbol{w}^{\mathsf{T}}\boldsymbol{h}_i}{\sigma}\right)\frac{\boldsymbol{h}_i}{\sqrt{2}\sigma}\right] \qquad (42)
\end{aligned}
$$

Summarizing the previous results we obtain the Newton iterative algorithm which can be used to solve the saddlepoint equation (27) after proper initialization of the vector $\boldsymbol{w}_0$:

$$
\begin{aligned}
\boldsymbol{w}_{n+1} &= \boldsymbol{w}_n+\left[\boldsymbol{I}_N+\frac{1}{\sigma^2}\sum_{i=1}^{N}\frac{\boldsymbol{h}_i\boldsymbol{h}_i^{\mathsf{T}}}{\cosh^2(\sqrt{2}\boldsymbol{w}_n^{\mathsf{T}}\boldsymbol{h}_i/\sigma)}\right]^{-1} \\
&\quad \left[\frac{\boldsymbol{y}}{\sqrt{2}\sigma}-\boldsymbol{w}_n-\sum_{i=1}^{N}\tanh\left(\frac{\sqrt{2}\boldsymbol{w}_n^{\mathsf{T}}\boldsymbol{h}_i}{\sigma}\right)\frac{\boldsymbol{h}_i}{\sqrt{2}\sigma}\right]. \qquad (43)
\end{aligned}
$$

III. Conclusions

We provided an asymptotic method to calculate the mutual information of an ISI channel. The method is based on the Hubbard-Stratonovich transform. It has the advantage of being applicable when the channel memory becomes very large since the complexity grows with polynomial order in the memory length.

Appendix A
Simple case: Binary input channel

This channel is modeled by the equation

$$
y=hx+z \qquad (44)
$$

where $z \sim \mathcal{N}(0, \sigma^2)$ and we assume $P(x = \pm 1) = 1/2$. From the conditional pdf

$$p(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-(y-hx)^2/(2\sigma^2)} \tag{45}$$

we obtain the marginal pdf

$$p(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-(y^2+h^2)/(2\sigma^2)} \cosh(hy/\sigma^2), \tag{46}$$

and the differential entropy

$$\begin{aligned} & h(y) \\ &= \frac{1}{2}\ln(2\pi\sigma^2) + \mathbb{E}\Big[(y^2+h^2)/(2\sigma^2) - \ln\cosh(hy/\sigma^2)\Big] \\ &= \frac{1}{2}\ln(2\pi\sigma^2) + 1/2 + h^2/\sigma^2 - \mathbb{E}\Big[\ln\cosh(hy/\sigma^2)\Big]. \end{aligned} \tag{47}$$

Since $h(y|x) = \frac{1}{2}\ln(2\pi\mathrm{e}\sigma^2)$, we get:

$$\begin{aligned} & I(x;y) \\ &= h^2/\sigma^2 - \mathbb{E}\Big[\ln\cosh(hy/\sigma^2)\Big] \\ &= h^2/\sigma^2 - \mathbb{E}[h(hx+z)/\sigma^2] - \mathbb{E}\Big[\ln((1+\mathrm{e}^{-2hy/\sigma^2})/2)\Big] \\ &= \ln 2 - \mathbb{E}\Big[\ln(1+\mathrm{e}^{-2hy/\sigma^2})\Big]. \end{aligned} \tag{48}$$

Since $y|x \sim \mathcal{N}(hx, \sigma^2)$, if $\sigma \to 0$, then $I(x;y) \to \ln 2$. If $\sigma \to \infty$, then $I(x;y) \to 0$.

Appendix B
Hubbard-Stratonovich transform

The following result is called the Hubbard-Stratonovich transform. Its proof can be found, *e.g.,* in [11].

Theorem B.1 *For any complex matrices* $\boldsymbol{A}^{\mathsf{T}}, \boldsymbol{B}, \boldsymbol{W}_0, \boldsymbol{Z}_0^{\mathsf{T}} \in \mathbb{C}^{m\times n}$*, we have:*

$$\mathrm{etr}(-\boldsymbol{B}\boldsymbol{A}) = \int_{\mathcal{D}_0} \mathrm{etr}(\boldsymbol{W}\boldsymbol{Z} - \boldsymbol{W}\boldsymbol{A} - \boldsymbol{B}\boldsymbol{Z}) d\mu(\boldsymbol{W}, \boldsymbol{Z}) \tag{49}$$

where

$$d\mu(\boldsymbol{W}, \boldsymbol{Z}) = \prod_{i=1}^{m}\prod_{j=1}^{n} \frac{d(\boldsymbol{W})_{ij}\, d(\boldsymbol{Z})_{ji}}{2\pi\mathrm{j}}, \tag{50}$$

and the integration domain $\mathcal{D}_0$ *is defined as*

$$\mathcal{D}_0 = \Big\{\boldsymbol{W}, \boldsymbol{Z} : \boldsymbol{W} = \boldsymbol{W}_0 + \boldsymbol{\Omega}, \boldsymbol{Z} = \boldsymbol{Z}_0 - \boldsymbol{\Omega}^{\mathsf{H}}, \boldsymbol{\Omega} \in \mathbb{C}^{m\times n}\Big\}, \tag{51}$$

The integral in (49) *is* ***absolutely convergent****.*

References

[1] W. Hirt, "Capacity and information rates of discrete-time channels with memory," Doctoral dissert. (Diss. ETH No. 86711, Swiss Federal Inst. of Technol. (ETH), Zurich, Switzerland, 1988.

[2] W. Hirt and J. L. Massey, "Capacity of the discrete-time Gaussian channel with intersymbol interference," *IEEE Trans. Inform. Theory*, vol. 34, no. 3, pp. 380-388, May 1988.

[3] S. Shamai, L. H. Ozarow, and A. D. Wyner, "Information rates for a discrete-time Gaussian channel with intersymbol interference and stationary inputs," *IEEE Trans. Inf. Theory*, vol. 37, pp. 1527–1539, Nov. 1991.

[4] S. Shamai (Shitz) and R. Laroia, "The intersymbol interference channel: lower bounds on capacity and channel precoding loss," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 13881404, Sep. 1996.

[5] Y. Carmon, S. Shamai and T. Weissman, "Disproof of the Shamai-Laroia Conjecture," *Proc. 27th IEEE Conf. Electrical and Electronics Engineers Eilat*, Israel, November 14 17th, 2012 (invited).

[6] D. Arnold and H.-A. Loeliger, "On the information rate of binary-input channels with memory," *Proc. 2001 IEEE Int. Conf. Commun.*, vol. 9, pp. 26922695.

[7] H.D. Pfister, J.B. Soriaga, and P.H. Siegel, "On the achievable information rates of finite state ISI channels," in *Proc. IEEE GLOBECOM 2001*, pp. 29922996.

[8] D.M. Arnold, H. Loeliger, P.O. Vontobel, A. Kavcic, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inf. Theory*, vol. 52, pp. 3498-3508, Aug. 2006.

[9] L.R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. 20, pp. 284–287, Mar. 1974.

[10] R.P. Agarwal, M. Meehan, and D. O'Regan, *Fixed Point Theory and Applications*. Cambridge University Press 2004.

[11] G. Taricco, "Further results on the asymptotic mutual information of Rician fading MIMO channels, *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 894-913, Jan. 2013.