Oblivious Fronthaul-Constrained Relay for a Gaussian Channel

Adi Homri, Michael Peleg, *Senior Member, IEEE*, and Shlomo Shamai (Shitz), *Fellow, IEEE*

***Abstract*— We consider systems in which the transmitter conveys messages to the receiver through a capacity-limited relay station. The channel between the transmitter and the relay station is assumed to be a frequency-selective additive Gaussian noise channel. It is assumed that the transmitter can shape the spectrum and adapt the coding technique so as to optimize performance. The relay operation is oblivious (nomadic transmitters), that is, the specific codebooks used are unknown. We find the reliable information rate that can be achieved with Gaussian signaling in this setting, and to that end, employ Gaussian bottleneck results combined with Shannon's incremental frequency approach. We also prove that, unlike classical water pouring, the allocated spectrum (power and bit rate) of the optimal solution could frequently be discontinuous. These results can be applied also to a MIMO transmission scheme. We also investigate the case of an entropy-limited relay. We show that the optimal relay function is always deterministic, present lower and upper bounds on the optimal performance (in terms of mutual information), and derive an analytical approximation.**

***Index Terms*— Oblivious processing, Gaussian information bottleneck, quantization, finite entropy, relay, water-pouring.**

I. Introduction

RELAYING exploits intermediate nodes to achieve communication between two distant nodes. Elementary relaying can be coarsely divided into compress-and-forward (of which amplify-and-forward is viewed as a special case) and decode-and-forward, depending on whether the relays decode the transmitted message or just forward the received signal to the destination. In this paper we examine the 'oblivious' relay system. The oblivious approach constructs universal relaying components serving many diverse users and operators and is not dependent on a priori knowledge of the modulation method and coding. Radio Frequency (RF) heads using the Common Public Radio Interface (CPRI) [1] is a prominent example of the oblivious relay. This approach might also benefit systems used in 'cloud' communication [2] and Cloud Radio Access Networks (CRAN) architecture [3]. Consider

Manuscript received October 17, 2017; revised February 23, 2018 and May 27, 2018; accepted May 31, 2018. Date of publication June 11, 2018; date of current version November 16, 2018. This research has received funding from the European Union's Horizon 2020 Research And Innovation Programme under grant agreement no. 694630. The associate editor coordinating the review of this paper and approving it for publication was C. Tian. *(Corresponding author: Adi Homri.)*

The authors are with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa 3200003, Israel (e-mail: adich1984@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2018.2846248

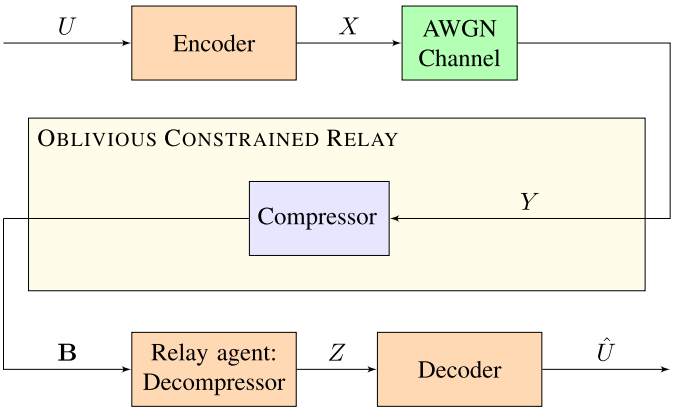


Fig. 1. The oblivious relay (blue) serving a user communicating via a Gaussian scalar channel (green).

the system in Fig. 1. The information source U, $H(U) \leq R$[nats/sec] is encoded into Gaussian symbols X and transmitted via a Gaussian scalar channel; the relay compresses the received symbols Y, encodes them into a bit-stream $\mathbf{B}$ and forwards it (without errors) to the final user's destination by a finite rate link with $H(\mathbf{B}) \leq C$ [nats/sec]. At the destination, a relay agent at the receiver side, decodes the bit-stream into symbols Z which are now an input for the receiver for estimation of U i.e. $\hat{U}$. For the user, the relay operation is hidden as it transmits the symbol X and receives symbol Z, while the effective channel is governed by the transition probability $P_{Z|X}(z|x)$. This setting provides the user a memoryless communication channel which forwards symbols from the transmitter to the receiver. We choose X to be Gaussian because of its optimality subject to a large bit-rate constraint C and because of its ubiquitous applications. In this setting, the user faces the familiar memoryless communication channel and can choose freely how to utilize it, e.g. the user can select a good error correcting code and change the codes after the oblivious system was already implemented. The relay is oblivious of the channel code used, that is, it holds no information about the transmitted codewords (randomized encoding/ nomadic transmitters, for each transmission the error correcting code is selected randomly by the transmitter and is unknown to the relay) and treats the channel output to be Gaussian and i.i.d (see [4] and [5] for a more detailed presentation of obliviousness). The relay performs lossy compression of the output of the Gaussian channel and is implemented by source coding. The trade-off between compression rate and mutual information between channel input and compressed channel output has closed-form expressions for the scalar and

vector case using the Gaussian Information Bottleneck (GIB) theorem [6]–[8]. This deviates from the classical remote rate-distortion approach [9]–[12] (rate distortion for sub-Nyquist sampling scheme) and [13] (sampling stationary signals subject to bit-rate constraints), since the distortion is measured by the equivocation $h(X|Z)$ instead of by the Minimum Mean Square Error (MMSE) $MMSE = E(X-Z)^2$. Since the distribution of X is fixed, minimizing $h(X|Z)$ means maximizing $I(X;Z) = h(X) - h(X|Z)$.

We further discuss the oblivious relay and focus our attention on the quantization process. We examine also simpler quantizers (bounded by entropy instead of by mutual-information) which can be implemented by the standard Lempel-Ziv algorithm instead of source coding. The performance of such quantizers that are optimal relative to an entropy constraint was studied for a wide class of memoryless sources (e.g. [14]–[16]). Notwithstanding, it is interesting to investigate the effect of such a constraint on the relay operation.

Our Contribution: The central contribution of this paper is operation over the frequency-selective channel, specifically deriving the frequency-domain allocation of the power and of the bit-rate which maximizes the information throughput of the oblivious relay when operating over the frequency-selective AWGN channel, see Section IV. To this end, we provide joint original treatment of the GIB and the classical water-pouring (some preliminary results were presented in [17]) and evaluate the optimal power allocation, which is very special and cannot be deduced from the classical water-pouring solution (which is a special case when $C \to \infty$). We employ Gaussian bottleneck results [6] combined with Shannon's incremental frequency approach [18]. The incremental approach leads to a clear solution for the frequency-selective channel setting. Our problem can be casted as a remote source coding as analyzed in [19] and [20] and references within but with logarithmic loss (distortion, i.e. conditional entropy) replacing the mean square error in [19] and [20]. Analysis of frequency-flat channels and MMSE optimization was reported in [21] and [22].

Our further contribution, presented in Section V, is analysis of lower complexity practical schemes, namely, lower and upper bounds on the mutual information between the transmitter and the receiver when the mutual information constraint of the full GIB is replaced by the entropy constraint as in the "deterministic GIB" approach of [23].

The remainder of this paper is organized as follows: Section II provides the system model. Section III outlines preliminaries in which we summarize quantization alternatives (III-A), demonstrate the advantages of stochastic quantizers (III-B), shows that the optimal transmitting scheme dictates independent Z_i (III-C), provide the required background and definitions for the GIB (III-D) and review the classical water-pouring method(III-E). In Section IV, we review the main results relevant to frequency-flat channels from [7] and [18] and present the new derivation for frequency selective channels and infinite-processing-time. Section V is dedicated to the finite entropy quantizer. Conclusions and proposals for future work are found in Section VI.

Notation: X is a random variable. x is a realization of a random variable. We use boldface letters for column vectors and sequences. X_i and x_i are the $i-th$ element in vectors $\mathbf{X}$ and $\mathbf{x}$, receptively. The expectation operator is denoted by $E[\cdot]$ and we follow the notation of [24] for entropy $H(\cdot)$, differential entropy $h(\cdot)$, and mutual information $I(\cdot;\cdot)$. A probability mass/distribution function is denoted by $P(\cdot)$ or $p(\cdot)$, respectively. All logarithms are natural and the unit of information is nats unless stated otherwise.

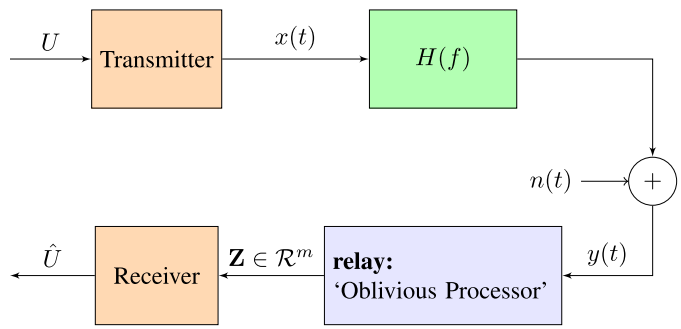


Fig. 2. A finite rate relaying operation over a fronthaul AWGN frequency selective channel.

II. System Model

Consider the system depicted in Fig. 2. $x(t)$ is the transmitted signal, assumed to be Gaussian, $H(f)$ is the frequency response of the channel linear filter, and the impulse response is $\mathcal{F}^{-1}[H(f)] = h(t)$ (here, $\mathcal{F}, \mathcal{F}^{-1}$ designate the Fourier transform and its inverse)

$$y(t) = x(t) * h(t) + n(t),$$

where $n(t)$ is normalized additive white Gaussian noise with one-sided power spectral density $N_0 = 1[Watt/Hz]$, and $*$ designates convolution. We are interested in the normalized mutual information when standard coding theorems [25] guarantee that the associated rate can be reliably transmitted through the system

$$\lim_{T\to\infty} \frac{1}{2T} I\left(X_{-T}^{T}; \mathbf{Z}\right) \triangleq I^C(X;Z). \tag{1}$$

Denote X_{-T}^{T} as $(X(t), -T \le t \le T)$, $\mathbf{Z} \in \mathcal{R}^m$ is the output vector (containing the compressed channel outputs Z_i), m denotes the number of symbols in a transmitted block and is of the dimension of $\mathbf{Z}$. Also, $\mathbf{Z}$ is entropy constrained by $H(\mathbf{Z}) \le C$[nats/sec], The information in (1) is also measured in terms of [nats/sec]. Again, we seek the (one-sided) power spectral density of the input Gaussian process $S_x(f)$ which maximizes $I^C(X;Z)$ under an average power constraint in some bandwidth W: $\int_0^W S_x(f)df \le P$. The continuous signal $x(t)$ in Fig. 2 is represented by discrete symbols X in Fig. 1. The symbols X may be either standard sampling of $x(t)$ or the Fourier transform of $x(t)$, the later preserves the scalar relation between X, Y and Z in Fig. 1 in the presence of the channel response $H(f)$.

III. Preliminaries

In this Section, we outline preliminaries in which we summarize quantization alternatives (III-A), demonstrate the advantages of stochastic quantizers (III-B), show that the optimal transmitting scheme dictates independent Z_i (III-C),

provide the required background and definitions for the GIB (III-D) and review the classical water-pouring method (III-E).

A. Quantization Alternatives

Denote by X, Y, Z the channel input, the received signal and the quantized output, respectively. Our system will try to maximize $I(X;Z)$ which clearly determines the maximal information rate R of the whole system if the user utilizes good error correcting codes, while minimizing the bit-rate of the sequence $\mathbf{B}$. Here we list some possible approaches to quantization:

1) Using the Channel Code in the Relay: If the relay would not be oblivious, it would decode the original information ($\hat{U}$) and send it as the sequence $\mathbf{B}$. In this case $R = min(C, \text{radio channel capacity})$ would be achieved. But in this work the relay is oblivious.

2) Mutual Information Constrained - Stochastic Quantizer (MUIC-SQ): A class of oblivious quantizers is stochastic, as mentioned for example in [6]. For each channel output Y_i, a compressed representation Z_i is obtained by a stochastic quantizer characterized by the probability mass function $P_{Z|Y}(z_i|y_i)$ chosen to maximize $I(X_i;Z_i)$; then Z_i is compressed and sent to the user's decoder using the bit rate $C = I(Y_i;Z_i)$. The practical implementation is by means of source coding on sequences. The received sequence $\mathbf{Y}$ is encoded into the sequence of bits $\mathbf{B}$ and the destination recovers the sequence $\mathbf{Z}$ from $\mathbf{B}$. The relay bit rate can be limited to $C = I(Y_i;Z_i)$. A proof following the steps of the source coding theorem [24] can be constructed. The probability mass function $P_{Z|Y}(z_i|y_i)$ will set $I(X_i;Z_i)$ and, thus, enable a system communication rate of $R = I(X_i;Z_i)$ by the classic channel coding theorem.

Letting the quantizer be stochastic improves performance, similarly to a corresponding advantage of source coding over memoryless deterministic quantization [6]. Koch [26] treats a stochastic quantizer, where the randomness is limited to dither known to the quantizer; this is a special case and may be considered a deterministic time-varying quantizer.

The optimal stochastic quantizer for Gaussian signals is the GIB, and was thoroughly analyzed in [7] and [6]. The GIB is a corner stone in this paper and its attributes will be specified in Section III-D.

3) Entropy Constrained Stochastic Quantizer (EC-SQ): The entropy constrained stochastic quantizer (EC-SQ), works in the same way as MUIC-SQ, except that the entropy of the compressed channel output $Z_i, H(Z_i)$ (instead of $I(X_i, Z_i)$) is bounded to be less than C. Entropy compression schemes such as Huffman or Lempel-Ziv are added after the quantizer, as suggested in the literature.

It is evident that in terms of mutual information $I(X_i;Z_i)$, the EC-SQ is inferior to MUIC-SQ since

$$I(Y_i;Z_i) = H(Z_i) - H(Z_i|Y_i) \leq H(Z_i), \tag{2}$$

thus enforcing a tight constraint on $P_{Z|Y}(z_i|y_i)$. For the Gaussian case the upper bound is the I_{GIB}.

4) Entropy Constrained Deterministic Quantizer (EC-DQ): This quantizer assumes deterministic mapping $Z_i = f(Y_i)$ (i.e. $H(Z_i|Y_i) = 0$), where $f(\cdot)$ is some function on the channel output Y_i. It is clear that for general channels it is inferior to the EC-SQ, as the deterministic domain is a subset of the stochastic domain. In the AWGN channel, there is a deterministic quantizer with identical performance. This can be proven using the following steps:

- Split the range of the channel output Y_i into small segments.
- Perform a resource allocation operation on each segment in order to have *deterministic* mapping that would yield the desired transfer function from Y_i to Z_i.

See rigorous proof in Appendix D.

5) Memoryless Deterministic Quantizer: The received signal Y is mapped to a discrete valued variable Z by a deterministic function. The function is optimized for mutual information $I(X;Z)$ per symbol with a constraint on the number of bits, or alphabet size of Z, required to represent the quantizer output symbol. The optimization can be done by the Lloyd algorithm. This is well covered by published papers, e.g. [27], which also show that the optimal probability distribution function of the transmitted signal X is discrete in many cases.

6) Vector Quantizers: Assume a vector compression scheme in which we group a few variables Z into small n -length vectors $\mathbf{Z_k} = [Z_{k+1}, Z_{k+2}, \ldots, Z_{k+n}]$, each being a deterministic function of the vector $\mathbf{Y_k} = [Y_{k+1}, Y_{k+2}, \ldots, Y_{k+n}]$ under the constraint $H(\mathbf{Z_k}) \leq nC$. Entropy coding will still be possible, now over $\mathbf{Z_k}$ instead of the scalars Z. This possibility leads to the following observations:

- With large n we can implement the full GIB by compressing the sequence $\mathbf{Y}$ into sequence $\mathbf{Z}$ by the MMSE criterion under the constraint of bit-rate C.
- Vector quantizers provide many intermediate performance levels starting at the deterministic entropy-constrained quantizer $(n = 1)$ and up to the GIB quantizer.

The advantage of the stochastic quantizer over the entropy constrained quantizer is the advantage of source coding over a scalar quantizer. Next, we shall present some attributes of the stochastic quantizer.

B. Demonstrating the Advantages of the Stochastic Quantizer

The advantage of the stochastic quantizer is demonstrated by a numerical example, see Fig. 3. We examine the case of a Gaussian X over an AWGN channel with a quantization rate $C = 1$[bits/symbol]. In the memoryless deterministic quantizer case, the quantizer is the sign of the received signal. Using Kindler *et al.* [28], the sign of the received signal is the optimal one bit memoryless deterministic quantizer, and not necessarily the optimal entropy constrained deterministic quantizer. The curve in Fig. 3 is the numerical evaluation of $E_{X,Z} log \frac{P(z|x)}{P(z)}$. In the stochastic quantizer case we have, from [6], the GIB $I(X;Z) = \frac{1}{2} log\left(\frac{1+SNR}{1+SNR \cdot e^{-2C}}\right)$. The results in Fig. 3 show the clear superiority of the stochastic quantization over the deterministic one. Modifying the distribution

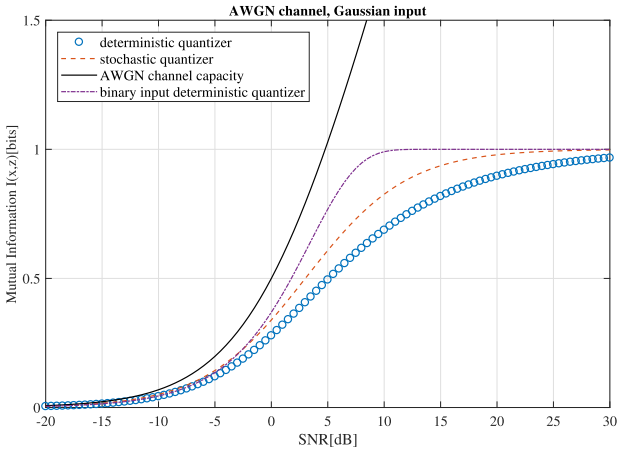


Fig. 3. Mutual information and system information rate R with Gaussian signal and two quantizers over an AWGN channel and with a quantization rate of $C = 1$[bit/channel use] as a function of SNR. Binary input is also presented for comparison.

of X would improve the rate [27] (see the improved performance with a binary input in Fig. 3).

C. Independent Z_i Achieve the Optimal Performance

We might still wonder if the stochastic quantizer, while evidently better than the deterministic one, is optimal. That is, the Z_i in our scheme are statistically independent. Could this scheme be outperformed if dependence between the Z_i was permitted? For example, the channel from X to Y could be a BSC and the relay could convey information to the destination by setting Z_i to be parity bits obtained by the XOR operation on pairs of Y_i. We shall show next that the independent Z_i, each statistically dependent on a single Y_i only, achieve the best performance possible. To show this, we consider the scheme as in Fig. 1, but instead of producing $\mathbf{Z}$, the bit sequence $\mathbf{B} = \mathbf{Z}$ is derived directly from the sequence $\mathbf{Y}$ and passed to the decoder together with the compression scheme. Thus, we want to maximize $I(\mathbf{X};\mathbf{Z})$, that is, the mutual information of whole sequences, with a constraint on $I(\mathbf{Y};\mathbf{Z})$. The first term is the information rate of the whole system and the second term is an achievable lower bound on the backhaul bit-rate C. We can restate this question as an equivalent bottleneck problem: Let $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ be sequences, each comprising n elements X_i, Y_i and Z_i. Also, let the elements of $\mathbf{X}, \mathbf{Y}$ be i.i.d. and the channel $\mathbf{X} - \mathbf{Y}$ be memoryless. In this case, the bottleneck problem is finding $P_{\mathbf{Z}|\mathbf{Y}}(\mathbf{z}|\mathbf{y})$ which maximizes $I(\mathbf{X};\mathbf{Z})$ with a constrained $I(\mathbf{X};\mathbf{Y})$, and the question on hand is: *Is* $P_{\mathbf{Z}|\mathbf{Y}}(\mathbf{z}|\mathbf{y}) = \Pi_i P_{Z|Y}(z_i|y_i)$ *?* The answer was already proved positive by Witsenhausen and Wyner [29] for discrete alphabets of X, Y and also for a Gaussian X over the AWGN channel in Chechik *et al.* [6] (An alternative proof for continuous alphabets is available in [30]).

D. Gaussian Information Bottleneck (GIB)

1) Information Rate - Scalar Channel: The GIB and its derivation for the discrete-time signaling case was thoroughly studied in [6], [6]–[8], and [31]. We will now give a brief overview of the GIB. The interested reader is referred to [7] and [8] for a full treatment. A complete derivation of the information rate function for the vector case, as well as the difference between the information rate function and the rate-distortion function, namely, $I(R) \geq I^{RD}(R)$, is presented in [8].

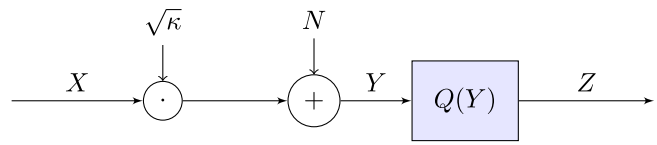


Fig. 4. Gaussian information bottleneck.

Consider the system in Fig. 4. The GIB addresses the following variational problem [6]: $\min_{P(z|y)} I(Y;Z) - \beta I(X;Z)$. In the context of the information bottleneck method, X is called the *relevance variable* and $I(X;Z)$ is termed *relevant information*. The trade-off between compression rate and relevant information is determined by the positive parameter β. It has been shown that the optimal z is jointly Gaussian with y and can be written as $Z = \alpha Y + \xi$, where $\alpha \in \mathcal{R}$ is scalar and $\xi \sim \mathcal{N}(0, \sigma_\xi)$ is independent of Y.

Definition 1: Let $X - Y - Z$ be a Markov chain. The information rate function $I : \mathcal{R}_+ \to [0, I(X;Y)]$ is defined by [7]: $I(C) \triangleq \max_{P(z|y)} I(X;Z)$ subject to $I(Y;Z) \leq C$.

$I(C)$ quantifies the maximum amount of the relevant information that can be preserved when the compression rate is at most C. Let us present $I(C)$ for the channel depicted in Fig. 4. Since X and Y are real zero-mean jointly Gaussian random variables, they obey $Y = \sqrt{\kappa}X + N$, where $h\kappa \in \mathcal{R}_+$ and $N \sim \mathcal{N}(0, \sigma^2)$ is independent of X. Setting $X \sim \mathcal{N}(0, P)$ yields $Y \sim \mathcal{N}(0, \kappa P + \sigma^2)$. The compressed representation of Y is denoted $Z = Q(Y)$. By the Markovity of $X - Y - Z$ we have $P_{Z|X}(z|x) = \int_{\mathcal{R}} P_{Z|Y}(z|y) p_{Y|X}(y|x) dy$, where $p_{Y|X}(y|x)$ is the transition probability distribution function of the Gaussian channel and $P_{Z|Y}(z|y)$ describes the compression mapping Q. The capacity of the Gaussian channel $p_{Y|X}(y|x)$ with average power constraint P and no channel compression equals [24] (units are [nats/channel use]): $\mathcal{C}(\rho) \triangleq \frac{1}{2} log(1+\rho)$ with ρ denoting the signal-to-noise ratio (SNR) $\rho \triangleq \frac{\kappa P}{\sigma^2}$. The following corollary states a closed-form expression for the information rate function and its properties [7, Th. 2].

Corollary 1: The information rate function of a Gaussian channel with SNR ρ is given by

$$I(C) = \frac{1}{2} log\left(\frac{1+\rho}{1+\rho e^{-2C}}\right). \tag{3}$$

The proof is in [7]. It should be noted that it can also be proved using the I-MMSE relation [32, Ch. 5, Sec. 7.1.3]. Fig. 5 illustrates the effect of limited-rate processing. It is clear that the total mutual information is upper bounded by the capacity for AWGN channels derived by Shannon [18].

E. Water-Pouring

We recall the classical water-pouring approach which yields the maximum $I^{\infty}(x;z)$ for $C \to \infty$. The idea of splitting the channel into incremental bands appears in [18] and [24], where each incremental band of bandwidth df is treated as an ideal (independent due to Gaussianity) band-limited channel with

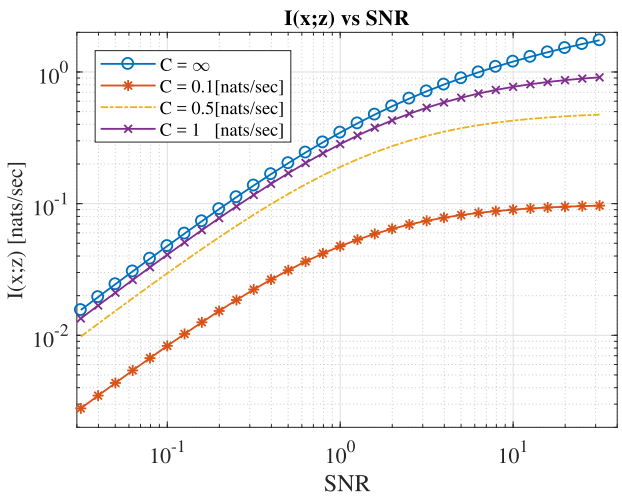


Fig. 5. GIB: Mutual information vs. rate(C) and ρ(SNR).

response $H(f)df$, and the result yields $\lim_{C\to\infty} I^C(x;z) \triangleq I^\infty(x;z) = \int_0^W log\left[1+S_x(f)|H(f)|^2\right]df$. Optimizing this over $S_x(f)$ under the power constraint yields (using the standard Euler-Lagrange method [33]) $\frac{|H(f)|^2}{1+S_x(f)|H(f)|^2} = \frac{1}{b}$.

Thus, the result is (see [18, Ch. 8]) $I_{water-pouring} \triangleq I^\infty(x;z) = \int_{\mathcal{B}} log\left[b|H(f)|^2\right]df$ and the frequency region $\mathcal{B}$ is given by $\mathcal{B} = \{f : b - \frac{1}{|H(f)|^2} \geq 0\}$.

IV. Water-Pouring With the Optimal Quantizer

In this Section, we present the new derivation for frequency selective channels and infinite-processing-time.

A. Processing Under Limited Bit-Rate C

As before, we adopt Shannon's incremental view, taking advantage of the fact that disjoint frequency bands are independent under the Gaussian law and stationarity. Let $\frac{1}{2}C(f)$ designate the number of [nats/channel use] assigned for delivering (processing) the band $(f, f+df)$. Since we have $2\cdot df$ independent channel uses (Nyquist) per second, the total rate per second in each band is $\frac{1}{2}C(f)2df = C(f)df$ and, hence, $\int_0^W C(f)df = C$. Culminating this view and incorporating (3), we reach the equation (for simplicity we denote $S_x(f)$ as $S(f)$) $I[f, S(f), C(f)] = \int_0^W \log\left[\frac{1+S(f)|H(f)|^2}{1+S(f)|H(f)|^2 e^{-C(f)}}\right]df$ leading to the following optimization problem:

$$\max_{S(f),C(f)} \int_0^W I[f, S(f), C(f)]df$$
$$s.t. \int_0^W S(f)df = P, \quad \int_0^W C(f)df = C. \tag{4}$$

The solution of Eq. (4) follows the standard Euler-Lagrange [33] reasoning. To that end, we follow the notation presented in [33]. $I[f,\hat{S},\hat{C}] \triangleq log\left(\frac{1+S_x(f)|H(f)|^2}{1+S_x(f)|H(f)|^2 e^{-C(f)}}\right)$ is the mutual information spectral density [nats/sec/Hz], where, $\hat{S} \triangleq S(f)$, $\hat{C} \triangleq C(f)$. The Lagrangian is

$$L\left[f,\hat{S},\hat{C}\right] = I[f,\hat{S},\hat{C}] - \lambda_c\cdot\hat{C} - \lambda_s\cdot\hat{S}, \tag{5}$$

where $\{\lambda_c, \lambda_s\} \in \Re$ are Lagrange coefficient multipliers. Differentiating Eq. (5) with respect to $\hat{C}, \hat{S}$ and defining $\hat{Q} \triangleq \exp(-\hat{C})$ and $X_f \triangleq H(f)^2 - \lambda_s - \lambda_c H(f)^2$, will lead to the following equation (see complete derivation in VI-A):

$$0 = -H(f)^2(1-\lambda_c)\hat{Q}^2 + X_f\hat{Q} - \frac{\lambda_s\lambda_c}{1-\lambda_c}. \tag{6}$$

The quadratic equation (6) produces two curve sets $\{S_i(f), Q_i(f)\}, i\in\{1,2\}$,

$$S_i(f) = \frac{X_f + \psi_i\sqrt{X_f^2 - 4H(f)^2\lambda_c\lambda_s}}{2H(f)^2\lambda_s}, \tag{7a}$$

$$Q_i(f) = \frac{X_f - \psi_i\sqrt{X_f^2 - 4H(f)^2\lambda_c\lambda_s}}{2H(f)^2(1-\lambda_c)}, \tag{7b}$$

where $\psi_i = (-1)^{i-1}$.

Proposition 1: We can discard the $\{S_2(f), Q_2(f)\}$ solution, since for each frequency, regardless of $H(f)$, $I[f,\hat{S},\hat{Q}]$ is not concave in the pair $\{S_2(f), Q_2(f)\}$.

A rigorous proof can be found within Sec. VI-E, where we derived that for each frequency f, the optimal values for $S(f), Q(f)$ are

$$S(f) = \begin{cases} \dfrac{X_f + \sqrt{X_f^2 - 4H(f)^2\lambda_c\lambda_s}}{2H(f)^2\lambda_s} & f\in\mathcal{B}_l \\ 0 & f\notin\mathcal{B}_l, \end{cases} \tag{8a}$$

$$Q(f) = \begin{cases} \dfrac{X_f - \sqrt{X_f^2 - 4H(f)^2\lambda_c\lambda_s}}{2H(f)^2(1-\lambda_c)} & f\in\mathcal{B}_l \\ 1 & f\notin\mathcal{B}_l, \end{cases} \tag{8b}$$

where $\mathcal{B}_l$ is the set of frequencies that have non-zero resource allocation (bit-rate and power). In general, $\mathcal{B}_l$ is unique unless the channel has a flat sub-band response. The algorithm for constructing $\mathcal{B}_l$ can be found in Sec. VI-C. In order to find the appropriate values for $\{\lambda_c, \lambda_s\}$ we had to use a grid search and the following proposition was used:

Proposition 2: $\{\lambda_c, \lambda_s\}$ are bounded by $0 \leq \lambda_s \leq \max H(f)^2$, $0 \leq \lambda_c \leq 1$.

Proof: See Sec. VI-B. ■

In stark contrast to classical water-pouring [13], [24], the optimal solution will frequently be discontinuous. As shown in Fig. 11, zero resources is a singular point inside the non-concave region. Since $C(f)$ and $S(f)$ can never drop gradually down to zero, the transition will always have an abrupt part. A simple example is the case where H(f) is constant over f, the SNR is sufficient and C is rather low. In this case an attempt to use frequency-constant $S(f)$ and $C(f)$ will place us in the non-concave region; a better solution will use only part of the available spectrum and utilize the available nats better by transmitting less information about the channel noise (see similar behavior in [21]). Fig. 6 demonstrates this idea, assuming a flat channel (i.e. $H(f) = 1$). For a given total power $P = 2[Watt]$, capacity $C = 0.5[nats/sec]$ and allocated user's bandwidth $W = 100[Hz]$, we calculated the mutual information rate when distributing the power and bit-rate uniformly over the bandwidth B used:$I\left[S(f), C(f)\right] = B\log\left[\frac{1+P/B}{1+P/Be^{\frac{-C}{B}}}\right]$. It is clear that the best course would be to use only part of the spectrum, namely $B/W \approx 0.3[\%]$, which is the maximum of the oblivious curve (blue).

TABLE I
PERFORMANCE OVER DIFFERENT CHANNELS (ALL UNITS ARE [nats/sec])

Case	α_1	f_1	α_2	f_2	Remark	$IC(x,z)$ - Our Optimal Scheme	$I^C(x,z)$ - Uniform Allocation	$I^C(x,z)$ - Limited-Rate Water-Pouring	$I^\infty(x,z)$
1	0.25	$0.25W$	0.75	$0.75W$	Channel A	2.94	2.83	0.77	2.944
2	0.25	$0.25W$	0.75	$0.75W$	Channel B , Fig. 7(a)	3.40	1.98	3.03	4.53
3	0	0	1	0.5W	Channel A	3.73	0.98	3.68	3.98
4	0	0	1	0.5W	Channel B	4.98	3.28	4.62	7.85
5	-	-	-	-	Allpass , Fig. 7(b)	7.92	7.75	7.75	23.98
6	-	-	-	-	Allpass, $W = 100[Hz]$	7.92	4.39	4.39	69.31

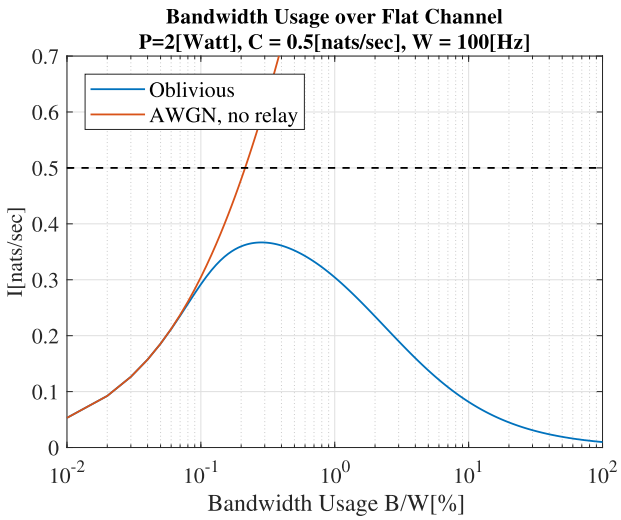


Fig. 6. Information rate as a function of allocated bandwidth.

B. Numerical Analysis

The proposed method has been applied on different types of channels (denoted as "Channel A") of the form $H_A(f) \equiv \alpha_1 N(f_1, 1) + \alpha_2 N(f_2, 1)$, where $N(\mu, \sigma^2)$ is the Gaussian curve. We used $P = 100[Watt]$ and $C = 9[nats/sec]$, while $W = 10[Hz]$ is the allocated user bandwidth. We also tested the "reciprocal" channel - denoted as "Channel B" (i.e, $H_B(f) = max[H_A(f)] - H_A(f)$). In each scenario, we compared the overall information rate using the following methods: the proposed method; uniform allocation of rate and power; classical water-pouring, as presented in [18], for the case of $C \to \infty$; "limited-Rate Water-Pouring", which is:

1) Calculate $S(f)$ using the classical water-pouring approach.
2) The allocated rate is: $C(f) = \frac{C}{P} S(f)$.

The results are summarized in Table I. Fig. 7 contains curves of $S(f), C(f), H(f)$ normalized to a unity average, and also a (normalized) classical water-pouring power allocation curve, $S(f)_{Water-Pouring}$, for comparison to the proposed approach. It should be noted that the curves $S(f), C(f)$ of Fig. 7(b) are not unique (algorithm dependent) since the channel has a flat response; however, the total mutual information is maximized nonetheless. It is clear from the results that:

- The proposed approach for allocating the power $S(f)$ and rate $C(f)$ is indeed optimal and superior to the other methods that were presented. Evidently, the rate is upper bounded by the classical water-pouring result ($C \to \infty$). It is evident that

$$\begin{aligned} I^\infty(X;Z) &\geq I^C(X;Z)|_{\text{Our Optimal Scheme}} \\ &\geq I^C(X;Z)|_{\text{Limited-Rate Water-Pouring}} \\ &\geq I^C(X;Z)|_{\text{Uniform Allocation}} \end{aligned}$$

- The price of obliviousness is demonstrated; as for a cognitive relay the reliable rate is $min(I^\infty(X;Z), C)$, achieved by a relay that decodes the signal and then transmits the decoded information at the maximum allowable rate (C).

V. FINITE OUTPUT ENTROPY $H(Z)$

In this section we analyze the performance of finite output entropy quantizers which can be implemented by a standard Lempel-Ziv algorithm, at a small cost in terms of performance. Analytic solutions for optimal information bottleneck quantizers are rarely available; here, we investigate optimization algorithms, since most practical algorithms cannot guarantee reaching a global optimum [34].

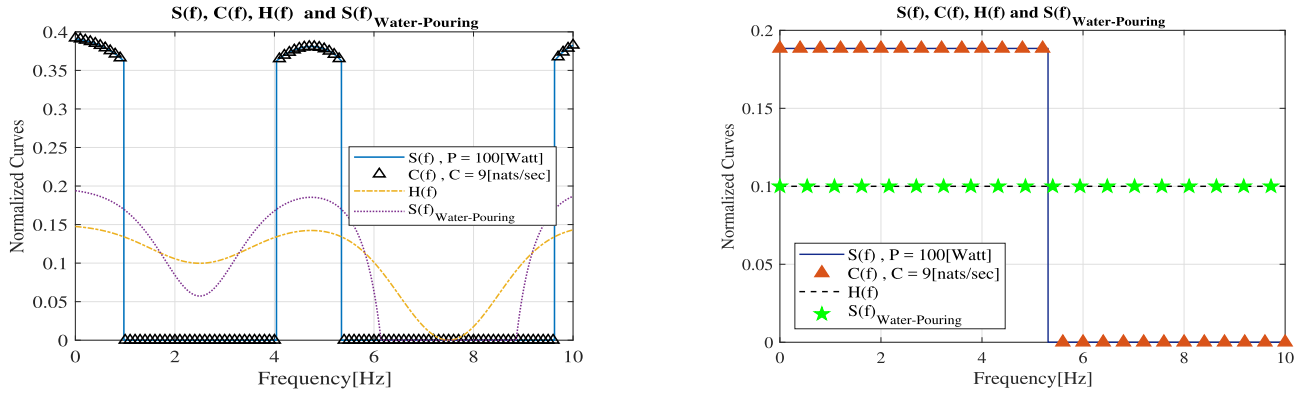
A. Deterministic Quantizer Model and Preliminaries

Reviewing the scalar bottleneck problem, we assume

$$Y = \sqrt{snr} \cdot X + N, \tag{9}$$

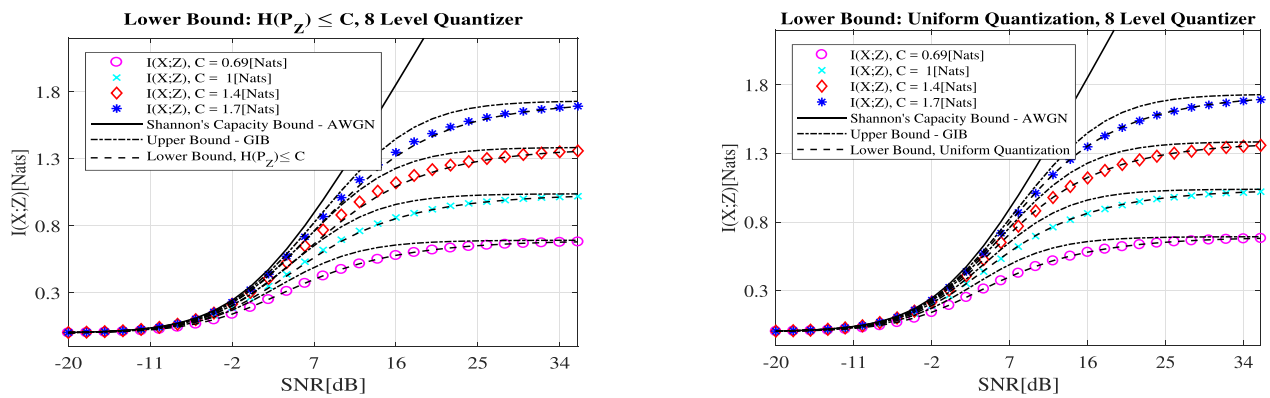
where X and N are unit variance independent Gaussian signals, and, hence, Y designates the output of a scalar Gaussian channel with a Gaussian input. The Finite-Entropy-Bottleneck Problem, reads: Find the maximum of $I(X;Z)$ under the Markov condition $X - Y - Z$, while $H(Z) \leq C$. In mathematical form: $\max_{P_{Z|Y}(Z|Y):H(Z)=C} I(X;Z)$. As mentioned in the preliminaries, the deterministic solution is optimal. In order to make computation feasible, the search was carried out for a K-bin or (K-level) deterministic quantizer $\hat{Q}$. $\hat{Q}$ maps the real input Y into one of K-bins, $Z = \hat{Q}(Y)$, producing discrete outputs with alphabet $Z \in \chi, |\chi| < \infty$. Bear in mind that $H(Z) \leq C$, but now $H(Z|Y) = 0$ since the quantizer is deterministic: hence $I(Y;Z) = H(Z)$. First we list a few definitions: Assume that $Z = z_i$ if $Y \in [q_{i-1}, q_i]$. We know that $P_{Y|X}(y|x) = N\left(\sqrt{snr} \cdot x, 1\right), \sigma_Y = \sqrt{1 + snr}$ and hence the probabilities $P_{Z|X}(z|x)$ and $P_Z(z)$ are

$$\begin{aligned} P_{Z|X}(z_i|x) &= p_{Y|X}(q_{i-1} \leq y \leq q_i|x) \\ &= Q\Big(q_{i-1} - \sqrt{snr} \cdot x\Big) - Q\Big(q_i - \sqrt{snr} \cdot x\Big), \\ P_Z(z_i) &= p_Y(q_{i-1} \leq y \leq q_i) = Q\Big(\frac{q_{i-1}}{\sigma_Y}\Big) - Q\Big(\frac{q_i}{\sigma_Y}\Big). \end{aligned}$$



(a) (b)

Fig. 7. (a) Allocated (normalized) power and bit-rate vs arbitrary channel and comparison to the allocated power resulting from water-pouring method; user bandwidth $W = 10[Hz]$. (b) An example of the abrupt nature of the optimal spectral allocation of power and bit-rate vs flat channel and comparison to the allocated power resulting from water-pouring method; user bandwidth $W = 10[Hz]$.



(a) (b)

Fig. 8. (a) Lower bound: setting the probability mass function $P_Z(z)$ s.t $H(Z) \leq C$, for each bit-rate constraint C, we present the numerical result for the optimal quantizer, the GIB upper bound, and the lower bound resulting from setting $H(P_Z) \leq C$. (b) Lower bound: Uniform quantization. For each bit-rate constraint C, we present the numerical result for the optimal quantizer, the GIB upper bound, and the lower bound resulting from uniform quantization of the channel output.

The resulting $H(Z|X), H(Z)$ and $I(X;Z)$ are

$$H(Z|X=x) = \sum_{z_i \in \chi} \Big[-P_{Z|X}(z_i|x) \cdot \log\Big(P_{Z|X}(z_i|x)\Big)\Big],$$
$$H(Z|X) = \int dF(x) H(Z|X=x),$$
$$H(Z) = \sum_{z_i \in \chi} \Big[-P_Z(z_i) \log P_Z(z_i)\Big],$$
$$I(X;Z) = H(Z) - H(Z|X).$$

$Q(x)$ denotes the complementary Gaussian distribution function $\frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt$. $F(x)$ is the cumulative distribution function (cdf) of x. Since both the information source and the noise are symmetric, we limit ourselves to the class of symmetric quantizers such as Eq. (10). The optimal quantizer problem can be stated as follows: $\max_{\{q_i\}:H(Z)\leq C} I(X;Z)$. The maximization is performed over the quantizer thresholds, $\{q_i\}$. In the following subsections we present numerical results of the problem under various conditions, and will gain some insights on the nature of the optimal quantizer and develop bounds and an analytical approximation.

B. Numerical Analysis of the Entropy Constrained Deterministic Quantizer

Our numerical optimization yields a 3-bit symmetric quantizer with thresholds

$$\{q_i\}_1^7 = \{-q_3, -q_2, -q_1, 0, q_1, q_2, q_3\}. \tag{10}$$

The thresholds were optimized to maximize the mutual information $I(X;Z)$ for various types of SNR and C. In Fig. 8(a) we see the resulting mutual information, as well as upper and lower bounds. From the results we see that:

- The mutual information $I(X;Z)$ increases with SNR and C
- The mutual information is bounded by the GIB.

C. The Effect of an Entropy Constraint on the Deterministic Quantizer Operation

We examine the case of an entropy constraint deterministic quantizer ($C \leq \log_2 |\chi|$, where $Z \in \chi$, χ is the alphabet of Z). From Fig. 9, it is evident that increasing the number of levels of the quantizer above the entropy constraint has

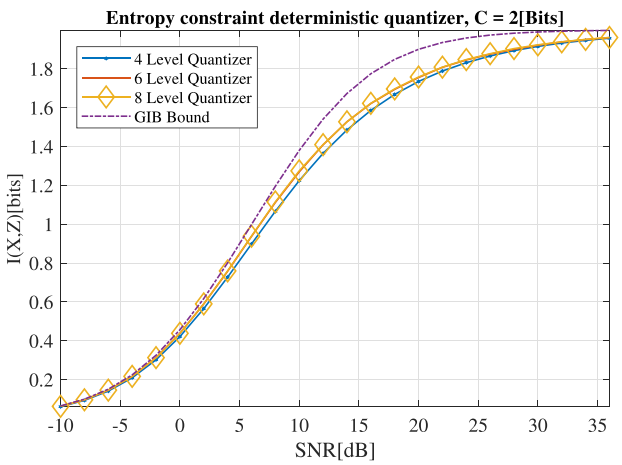


Fig. 9. Entropy constraint deterministic quantizer $C = 2[bits]$. The performance with six and eight levels is nearly identical.

almost no effect on the mutual information; thus the number of bins used was sufficient. The mutual information is bounded, as expected, by $\log_2 |\chi|$. One can see that even in a low SNR scenario the difference between the quantizers is negligible.

D. Lower and Upper Bounds on the Optimal Performance

We now try to bound the mutual information and apply an upper bound and two lower bounds. As before, the GIB can serve as an upper bound. For the lower bound (which are also interesting achievability schemes), we tested two schemes:

1) Lower Bound - Setting Output Entropy $H(Z) = C$: We chose a quantization scheme which will lead to an output entropy $H(Z) = C$. In order to assure the required entropy, we changed the cardinality of the output $|\mathcal{Z}|$ and the induced (probability mass function) $P_Z(z)$ using the method described in Sec. VI-F. Once the output probability mass function $P_Z(z)$ was set, the (symmetric) quantizer thresholds, $\{q_i\}_1^{|\mathcal{Z}|-1}$, can be found by taking an auxiliary variable ν_i: $\nu_i = \sum_{z=1}^{i} P_Z(z)$. The threshold $q_i = \sigma_Y \cdot Q^{-1}(\nu_i)$, where $Q^{-1}(x)$ denotes the inverse of $Q(x)$. Fig. 8(a) demonstrates these results.

2) Lower Bound - Uniform Quantizer: We tested a uniform quantizer, in which the quantizer step q was increased until the resulting probability mass function $P_Z(z)$ of the quantizer output had output entropy $H(P_Z) = C$. The output of the uniform quantizer has infinite cardinality since its input is unbounded. To that end, we discarded values that are higher (in their absolute value) than $M\sigma_Y$, ensuring output cardinality of $|\mathcal{Z}| \approx \frac{2M\sigma_Y}{q}$ for some large M.

Fig. 8(b) presents the results. For each bit-rate constraint C, we plot the numerically optimized quantizer, the GIB upper bound, and the lower bound resulting from uniform quantization. As one can see, the lower bound is fairly near to the curve of the numerically optimized quantizer. This method produced a tighter bound than the previous.

E. Analytic Approximation of Optimal Performance

Let $Z - Y - X$ be the inverse of the Markov chain defined in Sec. V and Eq. (9). Define $Y - X$, the inverse channel, as $X = E[X|Y] + (X - E[X|Y])$. Thus, X can also be written as $X = \frac{\sqrt{snr}}{1+snr}Y + \frac{1}{\sqrt{1+snr}}M$. Since $E[X|Y] = \alpha Y$ (where $\alpha = \frac{\sqrt{snr}}{1+snr}$), and due to the fact that the error term $X - E[X|Y]$ is independent of the measurement Y, M is a normalized Gaussian variable independent of Y. Having done so, note that

$$\begin{aligned} I(X;Y,Z) &= I(X;Z) + I(X,Y|Z) \\ &= I(X;Y) + I(X;Z|Y) \\ &= I(X;Y) + 0. \end{aligned}$$

$I(X;Z|Y) = 0$ due to Markovity, leading to $I(X;Z) = I(X;Y) - I(X,Y|Z)$. Then $I(X;Y|Z)$ is no more than a standard Gaussian channel from $Y \to X$, but Y is conditioned on Z since Gaussian inputs are optimal given the variance constraint

$$I(X;Y|Z) \leq E_Z\left\{\frac{1}{2}\log\left(\frac{snr}{1+snr}VAR(Y|Z)+1\right)\right\},$$

where $VAR(Y|Z) = E_{Y|Z}\{[Y - E(Y|Z)]^2|Z\}$. Incorporating the Jensen inequality will lead to

$$I(X;Y|Z) \leq \frac{1}{2}\log\left(\frac{snr}{1+snr}MMSE(Y|Z)+1\right), \quad (11)$$

where the $MMSE(Y|Z) = E[Y - E(Y|Z)]^2$ is the MMSE error of Y given Z. At this point we can utilize the results of Gish and Pierce [14], where $E[Y - E(Y|Z)]^2$ is minimized under the constraint of the entropy of $Z, H(Z)$, leading to the lower bound

$$\begin{aligned} I(X;Z) \geq\ & \frac{1}{2}\log(1+snr) \\ & - \frac{1}{2}\log\left(\frac{snr}{1+snr}MMSE(Y|Z)+1\right). \quad (12) \end{aligned}$$

Now, from Gish (when large output entropy is permitted or the quantization interval tends to zero), $C[bits] = H(Z) \approx 0.5\log_2\left(\frac{\sigma_Y^2}{MMSE(Y|Z)}\right) + 0.255[bits]$, where $\sigma_Y^2 = 1 + snr$; hence (converting bits to nats),

$$MMSE(Y|Z) \approx (1+snr)e^{(0.354-2C)}. \quad (13)$$

One can see that the quantization noise, $MMSE(Y|Z)$ decreases as $C \to \infty$, and increases with SNR (as the power of Y increases with it). As mentioned by Gish, the approximation is tight for low quantization noise and high output entropy (i.e. both $MMSE(Y|Z)$ and the SNR tend to $0, C \to \infty$). Incorporating (13) into (12) will lead to $I(X;Z) \geq \frac{1}{2}\log\left(\frac{1+snr}{1+snr\cdot e^{(0.354-2C)}}\right)$. Massey [35] has proved that in an AWGN channel at low SNR and with a zero-mean input, the capacity is the same function of the mean power regardless of the input's probability distribution function. It is also evident that zeroing the added component $0.354[nats]$ leads to the GIB and Gish's bounds coinciding, since

$$MMSE(Y|Z) = (1+snr)e^{-2C}. \quad (14)$$

Incorporating (14) in (11) will lead exactly to the GIB bound, which is achieved in the case where the inverse channel input $Y|Z$ is Gaussian, as GIB dictates. Fig. 10 demonstrates these results. Thus, the difference in performance at a low SNR and high C between the stochastic mutual information constrained

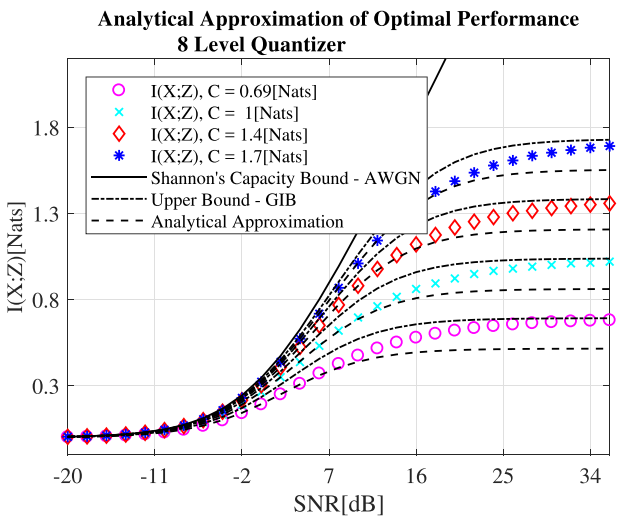


Fig. 10. Analytical approximation: For each bit-rate constraint C, we present the numerical result for the optimal quantizer, the GIB upper bound, and the analytical approximation.

quantizer and the deterministic entropy constrained quantizer is exactly the 0.255 bits per symbol in the relay bit-rate C.

VI. Conclusions and Future Directions

We presented and analyzed the rate- and power-limited oblivious relay over the frequency selective AWGN channel and derived the optimal transmit power spectral density and the optimal allocation of the relay bit-rate for Gaussian signaling. Our results relate directly to the classical water-pouring method, as well as to the Gaussian bottleneck frameworks. The advantage of this approach over other methods was demonstrated. Our problem can be viewed as a remote source coding [19] but with the logarithmic loss (distortion) replacing the mean square error and leading to basically different conclusions. We also investigated the class of finite entropy quantizers and, while it is difficult to find an analytical expression for the optimal quantizer, we devised lower and upper bounds for this case. We showed that with a Gaussian channel, a deterministic relaying is optimal. Our results on water-pouring also apply directly to the frequency dependent vector (MIMO) channels. Such channels can be transformed to a set of parallel independent channels [36]. Thus, equation (17) and the optimization algorithms can be applied on them with no need for modification by considering those independent channels as occupying independent frequency bands. A modern implementation of such a MIMO system might use the OFDM framework in which the MIMO channel diagonalization is convenient to implement (see, for example, [37]). One could extend the method presented in this paper to the setting where only partial channel state information is available to the transmitter. For example, the transmitter may assume a flat channel and a lower bound on the SNR at the relay. We expect such an approach to improve the performance with respect to other methods since the optimal scheme would use only part of the spectrum, as presented above.

Appendix

A. Solution of Eq. (5)

Differentiating Eq. (5) with respect to $\hat{C}, \hat{S}$ leads to $\frac{\partial I[f,\hat{S},\hat{C}]}{\partial \hat{S}} - \lambda_s = 0$, $\frac{\partial I[f,\hat{S},\hat{C}]}{\partial \hat{C}} - \lambda_c = 0$. Hence,

$$0 = \frac{H(f)^2(1-e^{-\hat{C}})}{(1+\hat{S}H(f)^2)(1+\hat{S}H(f)^2e^{-\hat{C}})} - \lambda_s, \quad (15a)$$

$$0 = \frac{\hat{S}H(f)^2e^{-\hat{C}}}{1+\hat{S}H(f)^2e^{-\hat{C}}} - \lambda_c. \quad (15b)$$

In order to simplify notation we use the following definitions: $\hat{Q} \triangleq \exp(-\hat{C})$, $X_f \triangleq H(f)^2 - \lambda_s - \lambda_c H(f)^2$. From (15b) it is clear that

$$\hat{S} = \frac{\lambda_c}{\hat{Q}H(f)^2(1-\lambda_c)}. \quad (16)$$

Substituting (16) in (15a) will lead to Eq. (6). We now have two sets of solutions for $\{\hat{S}, \hat{Q}\}$, where $\psi_i = (-1)^{i-1}$, then the solution for $\{\hat{S}, \hat{Q}\}$ is

$$\hat{S}_i = \frac{2\lambda_c}{X_f - \psi_i\sqrt{X_f^2 - 4H(f)^2\lambda_c\lambda_s}},$$

$$\hat{Q}_i = \frac{X_f - \psi_i\sqrt{X_f^2 - 4H(f)^2\lambda_c\lambda_s}}{2H(f)^2(1-\lambda_c)}.$$

Multiplying the denominator and numerator by $X_f + \psi_i\sqrt{X_f^2 - 4H(f)^2\lambda_c\lambda_s}$ will lead to

$$\hat{S}_i = \frac{X_f + \psi_i\sqrt{X_f^2 - 4H(f)^2\lambda_c\lambda_s}}{2H(f)^2\lambda_s}, \quad (17)$$

$$\hat{Q}_i = \frac{X_f - \psi_i\sqrt{X_f^2 - 4H(f)^2\lambda_c\lambda_s}}{2H(f)^2(1-\lambda_c)}. \quad (18)$$

At this point, we continue in accordance with Proposition 1 and discard the $\{\hat{S}_2, \hat{Q}_2\}$ curve since it is a non-concave solution.

B. Proof of Proposition 2

By investigating the derivatives of $I[f, \hat{S}, \hat{C}]$ w.r.t $\{\hat{S}, \hat{C}\}$ and taking into account that $\hat{S} \geq 0, \hat{C} \geq 0$ one can see that

$$\lambda_s = \frac{\partial I[f,\hat{S},\hat{C}]}{\partial \hat{S}} = \frac{H(f)^2(1-e^{-\hat{C}})}{(1+\hat{S}H(f)^2e^{-\hat{C}})(1+\hat{S}H(f)^2)} > 0, \quad (19a)$$

$$\lambda_s = \frac{\partial I[f,\hat{S},\hat{C}]}{\partial \hat{S}} \leq H(f)^2 \leq \max H(f)^2, \quad (19b)$$

$$\lambda_c = \frac{\partial I[f,\hat{S},\hat{C}]}{\partial \hat{C}} = \frac{\hat{S}H(f)^2e^{-\hat{C}}}{(1+\hat{S}H(f)^2e^{-\hat{C}})} \geq 0, \quad (19c)$$

$$\lambda_c = \frac{\partial I[f,\hat{S},\hat{C}]}{\partial \hat{C}} = \frac{\hat{S}H(f)^2e^{-\hat{C}}}{(1+\hat{S}H(f)^2e^{-\hat{C}})} \leq 1. \quad (19d)$$

The bounds for $\{\lambda_c, \lambda_s\}$ follow from (19).

C. Constructing the Set of Operating Frequencies $\mathcal{B}_l$

We perform a bounded grid search (see Proposition 2) on $\{\lambda_s, \lambda_c\}$ that will yield the maximum mutual information: $\int_0^W I[f, S(f), Q(f)]df$. For each λ_s, λ_c, the produced curves of $S(f), Q(f)$ (and hence, $S(f), C(f)$) might not meet the resource constraint ($\int_0^W S(f)df > P$ or $\int_0^W C(f)df > C$). At this point, we sort $I[f, S(f), Q(f)]$ and discard the frequencies (i.e. $S(f) = 0, Q(f) = 1$) that contribute least to the total mutual information, until compliance. The set of frequencies that were not discarded is $\mathcal{B}_l$.

D. Proof of Equivalence Between Entropy Constraint Stochastic and Deterministic Quantizers

A stochastic quantizer with limited $H(z)$ over the AWGN channel is characterized by $P_{Z|Y}(z|y)$. We construct a deterministic quantizer with the same performance as follows. Divide the range of y into segments γ_j small enough so that in each segment $p_{Y|X}(y|x)$ changes as little as desired. Denote by y_j the value of y in the center of γ_j. Divide each segment γ_j into subsegments, each mapped into a different z_k by the deterministic quantizer so that $P_{Z|Y}(z_k|\gamma_j)$ is preserved. The division is straightforward since in each segment $p_{Y|X}(y|x)$ is as constant as desired for all x and, clearly, so is $p_Y(y)$. Then, the probability of each subsegment when γ_j is given is the ratio of the length of the subsegment to the length of the segment. Thus, in each segment, each z_k is mapped to a subsegment the length of which is proportional to $P_{Z|Y}(z_k|\gamma_j)$ in the original stochastic quantizer. To prove equal performance of both the quantizers it is sufficient to establish that $P_{Z|X}(z_k|x)$ is preserved since it determines $I(x;z)$, $P_Z(z)$ and $H(z)$. For the stochastic quantizer and using the relations $p_Y(y) = p_Y(y_j)$, $p_{Y|X}(y|x) = p_{Y|X}(y_j|x)$ holding to any desired accuracy in each segment, we have

$$P_{Z|X}(z_k|x) = \int_{-\infty}^{\infty} P_{Z|Y}(z_k|y)p_{Y|X}(y|x)dy \tag{20a}$$

$$= \sum_j \int_{\gamma_j} P_{Z|Y}(z_k|y)p_{Y|X}(y|x)dy \tag{20b}$$

$$= \sum_j \int_{\gamma_j} \frac{P_{Z|Y}(z_k, y)}{p_Y(y)} p_{Y|X}(y|x)dy \tag{20c}$$

$$= \sum_j \frac{p_{Y|X}(y_j|x)}{p_Y(y_j)} \int_{\gamma_j} p_{Z,Y}(z_k, y)dy \tag{20d}$$

$$= \sum_j \frac{p_{Y|X}(y_j|x)}{p_Y(y_j)} p_{Z,Y}(z_k, \gamma_j)dy \tag{20e}$$

$$= \sum_j \frac{p_{Y|X}(\gamma_j|x)}{p_Y(\gamma_j)} p_{Z,Y}(z_k, \gamma_j)dy \tag{20f}$$

$$= \sum_j p_{Y|X}(\gamma_j|x)P_{Z|Y}(z_k|\gamma_j)dy. \tag{20g}$$

Where (20a) stems from Markovity. (20g) clearly also holds for the deterministic quantizer.

E. Proof of Proposition 1

In this subsection we prove that we can discard the solution $\{S_2(f), Q_2(f)\}$ of (17), based on the concavity of $I[f, S(f), C(f)]$ on the set $\{S(f), C(f)\}$ at each fixed frequency f (for simplicity we discard the channel dependence from now on and denote $I[S(f), C(f)]$). To this end, we shall prove that:

- Any point in the optimal solution, cannot reside in the non-concave region of $I[S(f), C(f)]$.
- The regions of concavity of $I[S(f), C(f)]$ coincide with the $\{S_1(f), C_1(f)\}$ solution.

The optimal solution at any frequency f cannot be in the non-concave region of $I[S(f), C(f)]$, because such a solution can be improved as follows: Suppose the solution assigned the resources $df \cdot S$ and $df \cdot C$ in an infinitesimal frequency band df around f such that $I[S(f), C(f)]$ is not concave at this point. Then, the df band can be split into two sub-bands with the resource assignment perturbed in each, but with the sum of the resources in df unchanged while increasing the performance I in df using the non-concavity. This is to be expected since our optimization equations are necessary but not sufficient conditions of global optimality [33]. Since we are dealing with a single frequency, $H(f)$ is constant and its influence is only a scaling of $S(f)$. Let us rewrite the Lagrangian at (5): $L = log\left(\frac{1+S}{1+Se^{-C}}\right) - \lambda_s S - \lambda_c C$ and equation (17) becomes (remembering that $Q \triangleq e^{-C}, C = -log(Q)$)

$$S = \frac{1 - \lambda_c - \lambda_s + \psi_i\sqrt{(1 - \lambda_c - \lambda_s)^2 - 4\lambda_c\lambda_s}}{2\lambda_s}, \tag{21a}$$

$$Q = \frac{1 - \lambda_c - \lambda_s - \psi_i\sqrt{(1 - \lambda_c - \lambda_s)^2 - 4\lambda_c\lambda_s}}{2(1 - \lambda_c)}. \tag{21b}$$

We can also write λ_s, λ_c as a function of (S, Q):

$$\lambda_c = -\frac{\partial L}{\partial C} = \frac{SQ}{1 + SQ}, \tag{22a}$$

$$\lambda_s = -\frac{\partial L}{\partial S} = \frac{1}{1 + S} - \frac{\lambda_c}{S}. \tag{22b}$$

We would like to choose only the concave solution, that is to choose (S, C) such that

$$\frac{\partial^2 I}{\partial S^2}\frac{\partial^2 I}{\partial C^2} - \frac{\partial^2 I}{\partial S \partial C}\frac{\partial^2 I}{\partial C \partial S} \geq 0, \tag{23a}$$

$$\frac{\partial^2 I}{\partial S^2} \leq 0 \;\; and \;\; \frac{\partial^2 I}{\partial C^2} \leq 0. \tag{23b}$$

We then prove that regions with a value of $\Psi_i = 1$ and concavity are identical.

Lemma 1: Different ψ_i in (21) enforce different (S, C).

Proof: Each (S, C) pair corresponds by (22) to a unique (λ_s, λ_c). Thus, the same (S, C) cannot be the outcome of two distinct (λ_s, λ_c) with different ψ. ■

The next step will be to show that the lines $S(\lambda_s, \lambda_c) = f(C; \lambda_s, \lambda_c)$ that split the regions of concavity and the sign of ψ coincide. Let us derive the dividing line between the $\pm$ regions in (21). At the dividing line $(1+\lambda_c+\lambda_s)^2 - 4\lambda_c\lambda_s$ must be zero by the proof of Lemma 1 and the fact the functions are continuous, so at any point the dividing line must be the result of (21) regardless of the value of ψ.

Particularly: two points infinitesimally near and each on a different side of the dividing line have the same S, C in the limit and, on the other hand, also the same (λ_s, λ_c) by (22),

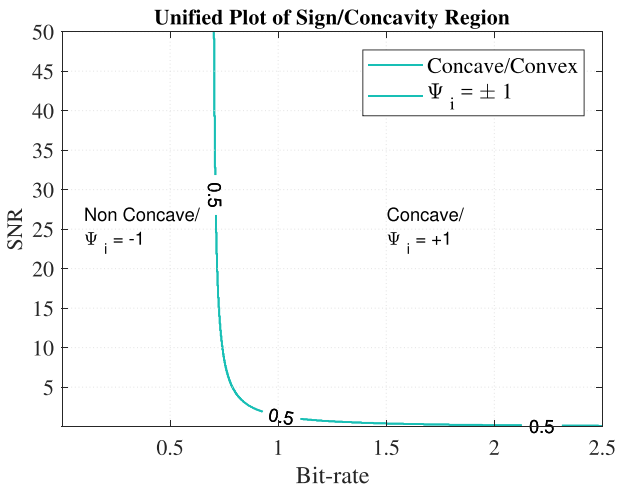


Fig. 11. Unified plot of concavity/sign regions.

so in the limit the value of ψ will not matter. $(1-\lambda_c-\lambda_s)^2 - 4\lambda_c\lambda_s = 0$, leading to

$$\lambda_{s,i} = (\lambda_c + 1) + \eta_i\sqrt{4\lambda_c}, \tag{24}$$

where $\eta_i = (-1)^{i-1}$. Substituting (24) back into (21) yields $S_i = \frac{1-\lambda_{s,i}-\lambda_c}{2\lambda_{s,i}} = \frac{\lambda_c+\eta_i\sqrt{\lambda_c}}{\lambda_c+1+2\eta_i\sqrt{\lambda_c}}$. The allocated power S_i must be non-negative; hence, by elimination, we discard η_1:

$$S_{plus/minus} = \frac{-(\lambda_c - \sqrt{\lambda_c})}{\lambda_c + 1 - 2\sqrt{\lambda_c}} = \frac{\sqrt{\lambda_c}}{1-\sqrt{\lambda_c}}, \tag{25}$$

where $S_{plus/minus}$ is S on the dividing line defined by the sign of ψ_i. Remembering that $Q = \frac{1}{S}\frac{\lambda_c}{1-\lambda_c}$ and $C = -log(Q)$ we have the curve $(S(\lambda_c), C(\lambda_c))$. We now examine the concavity regions of (23). We use the following derivatives:

$$\begin{aligned}\frac{\partial^2 I}{\partial S^2} &= -\frac{1}{(1+S)^2} + \frac{e^{-2C}}{(1+Se^{-C})^2}\\ &= -\frac{1}{(1+S)^2} + \frac{\lambda_c}{S^2},\end{aligned} \tag{26a}$$

$$\frac{\partial^2 I}{\partial C^2} = \frac{S^2e^{-2C}}{(1+Se^{-C})^2} - \frac{Se^{-C}}{1+Se^{-C}} = \lambda_c^2 - \lambda_c, \tag{26b}$$

$$\begin{aligned}\frac{\partial^2 I}{\partial S\partial C} &= \frac{\partial^2 I}{\partial C\partial S}\\ &= -\frac{Se^{-2C}}{(1+Se^{-C})^2} + \frac{e^{-C}}{1+Se^{-C}}\\ &= \frac{1}{S}(\lambda_c^2 - \lambda_c).\end{aligned} \tag{26c}$$

It is easy to see that $\frac{\partial^2 I}{\partial S^2} < 0$ and $\frac{\partial^2 I}{\partial C^2} < 0$, but we need to examine the sign regions of $\frac{\partial^2 I}{\partial S^2}\frac{\partial^2 I}{\partial C^2} - \frac{\partial^2 I}{\partial S\partial C}\frac{\partial^2 I}{\partial C\partial S}$. Substituting (26) in (23) we get

$$\begin{aligned}0 &= \frac{\partial^2 I}{\partial S^2}\frac{\partial^2 I}{\partial C^2} - \frac{\partial^2 I}{\partial S\partial C}\frac{\partial^2 I}{\partial C\partial S}\\ &= \Big[-\frac{1}{(1+S)^2} + \frac{\lambda_c}{S^2}\Big]\Big[\lambda_c^2 - \lambda_c\Big] - \Big[\frac{1}{S}(\lambda_c^2 - \lambda_c)\Big]^2.\end{aligned} \tag{27}$$

Eq. (27) leads to the following quadratic equation in S: $S^2(\lambda_c - 1) + 2S\lambda_c + \lambda_c = 0$.

Once more, we get to solution

$$S_i = \frac{\lambda_c + \eta_i\sqrt{\lambda_c}}{1-\lambda_c} = \frac{\sqrt{\lambda_c}(\sqrt{\lambda_c} + \eta_i)}{1-\lambda_c}. \tag{28}$$

We can discard $\eta_2(-)$ in order to ensure a non-negative solution for S_i, leading to $S_{concave/convex} = \frac{\sqrt{\lambda_c}}{1-\sqrt{\lambda_c}}$, which is exactly the dividing line as in (25). Thus, we have that regions with the sign of ψ and concavity/convexity regions of (S, C) are identical (since λ_c determines the same unique S in both cases and (λ_c, S) determine a unique C). This phenomenon can be easily demonstrated also numerically. We choose a square domain of (S, C); calculate (λ_s, λ_c) by (22) and choose the correct sign function Ψ_i in order to get back (S, C). Once the sign is set, we test for concavity. Fig. 11 shows that the regions of concavity and sign are identical. In this case we select the plus sign in order to get the concave solution. To conclude, let us investigate the lower limit on C using (25): $Q = \frac{1}{S}\cdot\frac{\lambda_c}{1-\lambda_c} = \frac{\sqrt{\lambda_c}}{1+\sqrt{\lambda_c}}$, and hence, $\lim_{\lambda_c\to 1} C = \lim_{\lambda_c\to 1} -log(Q) = log(2)$. This is the analytic derivation of the $1[bits/Hz]$ limit.

F. Calculating $P_Z(Z)$ s.t. $H(Z) = C$

This algorithm is designed to calculate the (entropy-limited deterministic) quantizer's output probability mass function, Z, that would meet the entropy constraint. The main idea here is to set an appropriate alphabet size $|Z|$, which is dependent upon C. If e^C is a natural number, the alphabet size would be $|Z| = e^C$ and the probability for each output would be e^{-C}. If not, we define the alphabet size $|Z| = \lceil e^C \rceil$. Setting equal probability to this alphabet would yield output entropy $H(Z) > C$. At this point, we can reach the desired entropy by gradually decreasing the probability of one of the outcomes, say $Z = 1$, and increasing (uniformly) the others, thus reaching the desired entropy $(H(Z) = C)$.

References

[1] A. de la Oliva, J. A. Hernandez, D. Larrabeiti, and A. Azcorra, "An overview of the CPRI specification and its application to C-RAN-based LTE scenarios," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 152–159, Feb. 2016.

[2] A. Sanderovich, S. Shamai (Shitz), and Y. Steinberg, "Distributed MIMO receiver—Achievable rates and upper bounds," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4419–4438, Oct. 2009.

[3] S.-H. Park, O. Simeone, O. Sahin, and S. Shamai (Shitz), "Fronthaul compression for cloud radio access networks: Signal processing advances inspired by network information theory," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 69–79, Nov. 2014.

[4] A. Sanderovich, S. Shamai (Shitz), Y. Steinberg, and G. Kramer, "Communication via decentralized processing," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3008–3023, Jul. 2008.

[5] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2148–2177, Oct. 1998.

[6] G. Chechik, A. Globerson, N. Tishby, and Y. Weiss, "Information bottleneck for Gaussian variables," *J. Mach. Learn. Res.*, vol. 6, pp. 165–188, Jan. 2004.

[7] A. Winkelbauer and G. Matz, "Rate-information-optimal Gaussian channel output compression," in *Proc. 48th Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2014, pp. 1–5, doi: 10.1109/CISS.2014.6814120.

[8] A. Winkelbauer, S. Farthofer, and G. Matz, "The rate-information trade-off for Gaussian vector channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun./Jul. 2014, pp. 2849–2853.

[9] T. Berger, *Rate-Distortion Theory*. Hoboken, NJ, USA: Wiley, 2003.

[10] R. Dobrushin and B. Tsybakov, "Information transmission with additional noise," *IRE Trans. Inf. Theory*, vol. 8, no. 5, pp. 293–304, Sep. 1962.

[11] A. Kipnis, A. J. Goldsmith, Y. C. Eldar, and T. Weissman, "Distortion rate function of sub-Nyquist sampled Gaussian sources," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 401–429, Jan. 2016.

[12] C. Tian and J. Chen, "Remote vector Gaussian source coding with decoder side information under mutual information and distortion constraints," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4676–4680, Oct. 2009.

[13] A. Kipnis, Y. C. Eldar, and A. J. Goldsmith. (2016). "Fundamental distortion limits of analog-to-digital compression." [Online]. Available: https://arxiv.org/abs/1601.06421

[14] H. Gish and J. Pierce, "Asymptotically efficient quantizing," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 5, pp. 676–683, Sep. 1968.

[15] P. Noll and R. Zelinski, "Bounds on quantizer performance in the low bit-rate region," *IEEE Trans. Commun.*, vol. COMM-26, no. 2, pp. 300–304, Feb. 1978.

[16] N. Farvardin and J. W. Modestino, "Optimum quantizer performance for a class of non-Gaussian memoryless sources," *IEEE Trans. Inf. Theory*, vol. 30, no. 3, pp. 485–497, May 1984.

[17] A. Homri, M. Peleg, and S. Shamai (Shitz), "Oblivious processing in a fronthaul constrained Gaussian channel," in *Proc. IEEE Int. Conf. Sci. Elect. Eng. (ICSEE)*, Nov. 2016, pp. 1–5.

[18] C. E. Shannon, "Communication in the presence of noise," *Proc. Inst. Radio Eng.*, vol. 37, no. 1, pp. 10–21, Jan. 1949.

[19] A. Kipnis, Y. C. Eldar, and A. J. Goldsmith, "Analog-to-digital compression: A new paradigm for converting signals to bits," *IEEE Signal Process. Mag.*, vol. 35, no. 3, pp. 16–39, May 2018.

[20] K. Eswaran and M. Gastpar. (2018). "Remote source coding under Gaussian noise: Dueling roles of power and entropy power." [Online]. Available: https://arxiv.org/abs/1805.06515

[21] S. Ray, M. Medard, and L. Zheng, "Fiber aided wireless network architecture," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 6, pp. 1284–1294, Jun. 2011.

[22] S. Ray, M. Medard, and L. Zheng, "A SIMO fiber aided wireless network architecture," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2006, pp. 2904–2908.

[23] D. J. Strouse and D. J. Schwab, "The deterministic information bottleneck," *Neural Comput.*, vol. 29, no. 6, pp. 1611–1630, 2017.

[24] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley Series in Telecommunications and Signal Processing), 10th ed. Hoboken, NJ, USA: Wiley, Aug. 1991.

[25] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.

[26] T. Koch, "On the dither-quantized Gaussian channel at low SNR," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun./Jul. 2014, pp. 186–190. [Online]. Available: arxiv:1401.6787

[27] J. Singh, O. Dabeer, and U. Madhow, "Capacity of the discrete-time AWGN channel under output quantization," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2008, pp. 1218–1222.

[28] G. Kindler, R. O'Donnell, and D. Witmer. (2016). "Remarks on the most informative function conjecture at fixed mean." [Online]. Available: https://arxiv.org/abs/1506.03167

[29] H. S. Witsenhausen and A. D. Wyner, "A conditional entropy bound for a pair of discrete random variables," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 5, pp. 493–501, Sep. 1975.

[30] A. Homri, M. Peleg, and S. Shamai (Shitz). (2015). "Oblivious fronthaul-constrained relay for a Gaussian channel." [Online]. Available: https://arxiv.org/abs/1510.08202

[31] N. Tishbi, F. Pereira, and W. Bialek. (2000). "Information bottleneck method." [Online]. Available: https://arxiv.org/abs/physics/0004057

[32] D. Guo, S. Shamai (Shitz), and S. Verdú, "The interplay between information and estimation measures," *Found. Trends Signal Process.*, vol. 6, no. 4, pp. 243–429, 2012. [Online]. Available: http://dx.doi.org/10.1561/2000000018

[33] I. M. Gelfand and S. V. Fomin, *Calculus of Variations*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1963.

[34] S. Hassanpour, D. Wuebben, and A. Dekorsy, "Overview and investigation of algorithms for the information bottleneck method," in *Proc. 11th Int. ITG Conf. Syst., Commun. Coding (SCC)*, Feb. 2017, pp. 1–6.

[35] J. L. Massey, "All signal sets centered about the origin are optimal at low energy-to-noise ratios on the AWGN channel," in *Proc. ISIT*, Jun. 1976, pp. 80–81.

[36] L. H. Brandenburg and A. D. Wyner, "Capacity of the Gaussian channel with memory: The multivariate case," *Bell Syst. Techn. J.*, vol. 53, no. 5, pp. 745–778, May/Jun. 1974.

[37] J. Mietzner, R. Schober, L. Lampe, W. H. Gerstacker, and P. A. Hoeher, "Multiple-antenna techniques for wireless communications—A comprehensive literature survey," *IEEE Commun. Surveys Tuts.*, vol. 11, no. 2, pp. 87–105, 2nd Quart., 2009.

Adi Homri received the B.Sc. degree from Tel Aviv University in 2010 and the M.Sc. degree from the Technion–Israel Institute of Technology, Haifa, Israel, in 2018. He was with Rafael Advanced Defense Systems Ltd. He is currently with Apple Ltd. His research interests include wireless communications and multi-antenna systems.

Michael Peleg (SM'98) received the B.Sc. and M.Sc. degrees from the Technion–Israel Institute of Technology, Haifa, Israel, in 1978 and 1986 respectively.

From 1980 to 2000, he was with the communication research facilities of the Israel Ministry of Defense. Since 2000, he has been with Rafael Advanced Defense Systems Ltd. He is associated with the Department of Electrical Engineering, Technion–Israel Institute of Technology, where he is involved in communications and information theory.

His research interests include wireless communications, iterative decoding, multi-antenna systems, and radiation safety.

Shlomo Shamai (Shitz) (F'94) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Technion–Israel Institute of Technology, in 1975, 1981, and 1986 respectively. From 1975 to 1985, he was with the Communications Research Laboratory as a Senior Research Engineer. Since 1986, he has been with the Department of Electrical Engineering, Technion–Israel Institute of Technology, where he is currently a Technion Distinguished Professor, and holds the William Fondiller Chair of Telecommunications. His research interests include wide spectrum of topics in information theory and statistical communications.

Dr. Shamai is a URSI Fellow, a member of the Israeli Academy of Sciences and Humanities, and a Foreign Member of the U.S. National Academy of Engineering. He was a recipient of the 2011 Claude E. Shannon Award, the 2014 Rothschild Prize in mathematics/computer sciences and engineering, and the 2017 IEEE Richard W. Hamming Medal.

He was a recipient of the 1985 Alon Grant for distinguished young scientists, the 1999 van der Pol Gold Medal of the Union Radio Scientifique Internationale, and the 2000 Technion Henry Taub Prize for Excellence in Research. He was a co-recipient of the 2000 IEEE Donald G. Fink Prize Paper Award, the 2003 and 2004 Joint IT/COM Societies Paper Award, the 2007 IEEE Information Theory Society Paper Award, the 2009 and 2015 European Commission FP7, Network of Excellence in Wireless COMmunications (NEWCOM++) Best Paper Awards, the 2010 Thomson Reuters Award for international excellence in scientific research, the 2014 EURASIP Best Paper Award (for the *EURASIP Journal on Wireless Communications and Networking*), and the 2015 IEEE Communications Society Best Tutorial Paper Award. He has served as an Associate Editor for the Shannon Theory of the IEEE TRANSACTIONS ON INFORMATION THEORY and has also served twice on the Board of the Governors of the Information Theory Society. He has also served on the Executive Editorial Board of the IEEE TRANSACTIONS ON INFORMATION THEORY and the IEEE Information Theory Society Nominations and Appointments Committee.