

Oblivious Processing in a Fronthaul Constrained Gaussian Channel

Adi Homri

Department of EE, Technion
Israel Institute of Technology,
Haifa, Israel
and Rafael Ltd

Michael Peleg

Department of EE, Technion
Israel Institute of Technology,
Haifa, Israel
and Rafael Ltd

Shlomo Shamai

Department of EE, Technion
Israel Institute of Technology,
Haifa, Israel

Abstract—

We consider systems for 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 frequency selective additive Gaussian noise channel. It is assumed that the transmitter can shape the spectrum and adapt the coding technique as to optimize performance. The relay operation is oblivious (nomadic transmitters [1]), that is, the specific codebooks used are unknown, while the spectral shape of the transmitted signal is available. We find the reliable information rate that can be achieved in this setting with Gaussian signaling, and to that end employ Gaussian bottleneck results [2] combined with Shannon's incremental frequency approach [3]. We also prove that unlike classical water-pouring, the allocated spectrum (power and bit-rate) of the optimal solution could be frequently discontinuous.

I. INTRODUCTION

The relaying technique makes use of intermediate nodes to help the communication between two distant nodes. Elementary relaying can be coarsely divided into compress-and-forward (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 is intended to construct universal relaying components serving many diverse users and operators and *not dependent on knowing the modulation method and coding*. Such an approach might benefit systems used in 'cloud' communication and was investigated for example in [4].

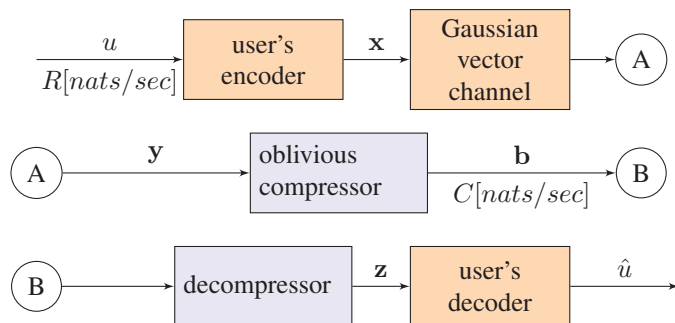


Fig. 1: The oblivious relay serving a user - Gaussian vector channel

Consider the system in Figure 1. The relay which compresses the received signal y and forwards it to the final user's destination by a finite rate C provides the user with a memoryless communication channel forwarding symbols from the

transmitter to the receiver. The user transmits the symbols x and receives symbols z while the effective channel is governed by the transition probability $P(z|x)$. We choose x to be Gaussian because of its optimality at large C and because of its ubiquitous applications. In such a setting the user faces the familiar memoryless communication channel and can choose freely how to utilize it, e.g. selecting a good error correcting code and changing the codes after the oblivious system was already implemented. The serving system is oblivious of the channel code used. See [1] for a more rigorous presentation of obliviousness. The relay performs lossy compression of the output of the Gaussian channel. 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 [2], [5] and [6]. This deviates from the classical remote rate distortion approach [7], [8] and [9] (rate distortion for sub-Nyquist sampling scheme) and [10] (Sampling stationary signals subject to bit-rate constraints), since the distortion is measured by the equivocation $H(X|Z)$ and not $MMSE = E(X - Z)^2$. Say the distribution of X is fixed, then minimizing $H(X|Z)$ means maximizing $I(X; Z) = H(X) - H(X|Z)$. In this paper, we provide a further generalization of the GIB for the case of frequency selective additive Gaussian. We find the reliable information rate that can be achieved in this setting, and to that end employ Gaussian bottleneck results [2] combined with Shannon's incremental frequency approach [3]. The reminder of this paper is organized as follows. Section II provides the required background and definitions for the GIB. In Section III, we review the main results relevant to frequency-flat channel from [5], [3] and present the new derivation for frequency selective channel and infinite-processing-time. Numerical results are provided in Section IV, conclusions and future discussions on V. *Notation:* We use boldface letters for column vectors and sequences. The expectation operator is denoted by E and we follow the notation of [11] for entropy $H(\cdot)$, differential entropy $h(\cdot)$, and mutual information $I(\cdot; \cdot)$. Furthermore, $[x]^+ \triangleq \max\{x, 0\}$. All logarithms are natural.

II. GAUSSIAN INFORMATION BOTTLENECK

A. Information rate - vector channel

The Gaussian Information Bottleneck and its derivation for the discrete-time signaling case, was thoroughly studied at [5], [12], [2], [6] and [13]. We will now give a brief overview of the GIB (the interested reader is referred to [5], [6] 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 [6].

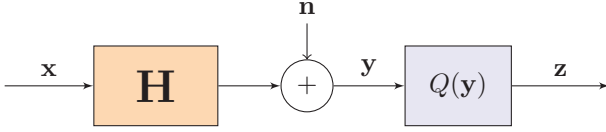


Fig. 2: Gaussian Information Bottleneck - Vector Case

Consider the system in Figure 2. Let \mathbf{x} and \mathbf{y} be jointly Gaussian zero-mean random vectors (length n) with full rank covariance matrices. Assuming that $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{\mathbf{x}\mathbf{x}})$. The channel output equals

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \quad (1)$$

where $\mathbf{H} \in \mathcal{R}^{n \times m}$ and the additive noise $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ is independent of \mathbf{x} which results

$$\mathbf{R}_{\mathbf{y}\mathbf{y}} = \mathbf{H}\mathbf{R}_{\mathbf{x}\mathbf{x}}\mathbf{H}^T + \sigma^2 \mathbf{I} \quad (2)$$

Let \mathbf{z} be a compressed representation of \mathbf{y} denoted by the conditional distribution $P(\mathbf{z}|\mathbf{y})$. It follows that $\mathbf{x} - \mathbf{y} - \mathbf{z}$ forms a Markov chain and hence by Markovity

$$P(\mathbf{z}|\mathbf{x}) = \int_{\mathcal{R}^n} P(\mathbf{z}|\mathbf{y})P(\mathbf{y}|\mathbf{x})d\mathbf{y} \quad (3)$$

The compression rate equals $I(\mathbf{z}; \mathbf{y})$. The GIB addresses the following variational problem [14]:

$$\min_{P(\mathbf{z}|\mathbf{y})} I(\mathbf{z}; \mathbf{y}) - \beta I(\mathbf{x}; \mathbf{z}) \quad (4)$$

In the context of the information bottleneck method, \mathbf{x} is called the *relevance variable* and $I(\mathbf{x}; \mathbf{z})$ 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 \mathbf{z} is jointly Gaussian with \mathbf{y} and can be written as

$$\mathbf{z} = \mathbf{A}\mathbf{y} + \xi \quad (5)$$

where \mathbf{A} is an $n \times n$ matrix and $\xi \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_\xi)$ is independent of \mathbf{y} .

Definition 1. Let $\mathbf{x} - \mathbf{y} - \mathbf{z}$ be a Markov chain. The information-rate function $I : \mathcal{R}_+ \rightarrow [0, I(\mathbf{x}; \mathbf{y})]$ is defined by [5]

$$I(C) \triangleq \max_{P(\mathbf{z}|\mathbf{y})} I(\mathbf{x}; \mathbf{z}) \text{ subject to } I(\mathbf{y}; \mathbf{z}) \leq C \quad (6)$$

$I(C)$ quantifies the maximum of the relevant information that can be preserved when the compression rate is at most C .

B. Information rate - scalar channel

We now present $I(C)$ for the channel depicted at Figure 2 for the scalar case. Since x and y are real zero-mean jointly Gaussian random variables, they obey:

$$y = \sqrt{h}x + n \quad (7)$$

where $h \in \mathcal{R}_+$ and $n \sim \mathcal{N}(0, \sigma^2)$ is independent of x . Setting $x \sim \mathcal{N}(0, P)$ yields $y \sim \mathcal{N}(0, hP + \sigma^2)$. The compressed representation of y is denoted $z = Q(y)$. By Markovity of $x - y - z$ we have

$$P(z|x) = \int_{\mathcal{R}} P(z|y)P(y|x)dy \quad (8)$$

where $P(y|x)$ is the transition pdf of the Gaussian channel and $P(z|y)$ describes the compression mapping Q . The capacity of the Gaussian channel $P(y|x)$ with average power constraint P and no channel compression equals [11] (units are [nats/channel use])

$$\mathcal{C}(\rho) \triangleq \frac{1}{2} \log(1 + \rho) \quad (9)$$

with ρ as the signal-to-noise ratio (SNR)

$$\rho \triangleq \frac{hP}{\sigma^2} \quad (10)$$

The following corollary states a closed-form expression for the information-rate function and properties [5, Theorem 2].

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

$$I(C) = \frac{1}{2} \log \left(\frac{1 + \rho}{1 + \rho e^{-2C}} \right) \quad (11)$$

$I(C)$ has the following properties:

- (a) $I(C)$ is strictly concave in \mathcal{R}_+
- (b) $I(C)$ is strictly increasing in C .
- (c) $I(C) \leq \min\{C, \mathcal{C}(\rho)\}$
- (d) $I(0) = 0$ and $\lim_{C \rightarrow \infty} I(C) = \mathcal{C}(\rho)$
- (e) $\frac{\partial I(C)}{\partial C} = (1 + e^{2C} \rho^{-1})^{-1} \leq \frac{\partial I(C)}{\partial C} |_{C=0} = (1 + \rho^{-1})^{-1}$

The proof is within [5]. It should be noted that it can also be proved using the I-MMSE relation [15, Chapter 5, Section 7.1.3].

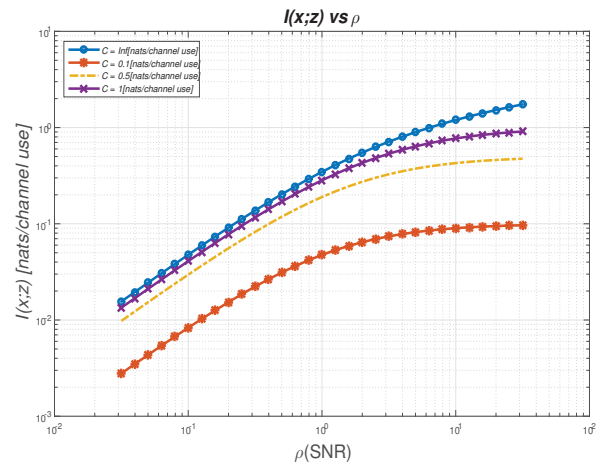


Fig. 3: $I(x; z)$ vs. Rate and $\rho(\text{SNR})$

Figure 3 illustrates the effect of limited-rate processing. It is clear that the total mutual information is upper bounded by the capacity for AWGN channels (which can be achieved by the water-pouring approach) presented by Shannon [3].

III. GENERALIZING THE GIB - FREQUENCY SELECTIVE CHANNEL INFINITE PROCESSING TIME

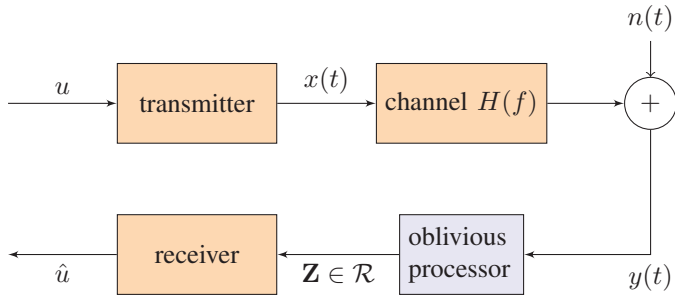


Fig. 4: Finite rate fronthaul Gaussian channel

Consider the system depicted in figure 4. $x(t)$ - the input signal assumed to be Gaussian, $H(f)$ - the frequency response of the channel linear filter, and impulse response $\mathcal{F}^{-1}(H(f)) = h(t)$ (here, $\mathcal{F}, \mathcal{F}^{-1}$ designate the Fourier response and its inverse)

$$y(t) = x(t) * h(t) + n(t) \quad (12)$$

where $n(t)$ is a normalized additive white Gaussian noise with one-sided power spectral density $N_0 = 1[\text{Watt}/\text{Hz}]$, and $*$ designates convolution.

We are interested in the normalized mutual-information when standard coding theorems [16] guarantee that the associated rate can be reliably transmitted through the system.

$$\lim_{T \rightarrow \infty} \frac{1}{2T} I(X_{-T}^T; \mathbf{Z}) \triangleq I_n^C(x; z) \quad (13)$$

where $x_a^b : (x(t), a \leq t \leq b)$ and where \mathbf{Z} is the output binary vector that is constrained to be $\mathcal{C}[\text{nats}/\text{sec}]$. The information in (13) is also measured in terms of $[\text{nats}/\text{sec}]$. n denotes the n -symbol block transmitted and is the dimension of \mathbf{Z} .

Again, we'd like to find the (one-sided) power spectral density of the input Gaussian process $S_x(f)$ which maximizes $I_n^C(x; z)$ under an average power constraint under some bandwidth W :

$$\int_0^W S_x(f) df \leq P \quad (14)$$

A. Water pouring

First we shall recall the classical water pouring approach which yields the maximum $I_n^\infty(x; z)$ for $C \rightarrow \infty$.

The idea of splitting the channel into incremental bands appears in [3] and [11] where each incremental band of bandwidth df is treated as ideal (independent due to Gaussianity) band-limited channel with response $H(f)df$, and the result yields:

$$\lim_{C \rightarrow \infty} I_n^C(x; z) \triangleq I_n^\infty(x; z) = \int_0^W \log [1 + S_x(f)|H(f)|^2] df \quad (15)$$

Optimizing this over $S_x(f)$ under the power constraint yields (using the standard Euler-Lagrange [17])

$$\frac{|H(f)|^2}{1 + S_x(f)|H(f)|^2} = \frac{1}{b} \quad (16)$$

Thus, the result is (see [3, chapter 8]).

$$I_{\text{water-pouring}} \triangleq I_n^\infty(x; z) = \int_{\mathcal{B}} \log [b|H(f)|^2] df \quad (17)$$

and where the frequency region \mathcal{B} is given by

$$\mathcal{B} = \{f : b - \frac{1}{|H(f)|^2} \geq 0\} \quad (18)$$

B. Processing under limited bit-rate C

As before, we adopt the 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 $[\text{nats}/\text{channel use}]$ assigned for delivering (processing) the band $(f, f + df)$. Since we do have $2df$ 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,

$$\int_0^W C(f)df = C \quad (19)$$

culminating this view and incorporating (11), we reach the equation:

$$I[f, S_x(f), C(f)] = \int_0^W \log [1 + S_x(f)|H(f)|^2] df - \int_0^W \log [1 + S_x(f)|H(f)|^2 e^{-C(f)}] df \quad (20)$$

Leading to the following optimization problem:

$$\begin{aligned} \max_{S_x(f), C(f)} & \int_0^W \log \left(\frac{1 + S_x(f)|H(f)|^2}{1 + S_x(f)|H(f)|^2 e^{-C(f)}} \right) df \\ \text{s.t.} & \int_0^W S_x(f)df = P, \int_0^W C(f)df = C \end{aligned} \quad (21)$$

The result follows the standard Euler-Lagrange [17] reasoning. To that end, we follow the notation presented in [17]. Assuming $I[f, \hat{S}_x, \hat{C}] \triangleq \log \left(\frac{1 + S_x(f)|H(f)|^2}{1 + S_x(f)|H(f)|^2 e^{-C(f)}} \right)$, is the mutual information spectral density $[\text{nats}/\text{sec}/\text{Hz}]$. Also, $\hat{S}_x \triangleq S_x(f), \hat{C} \triangleq C(f)$. The Lagrangian is:

$$L[f, \hat{S}_x, \hat{C}] = I[f, \hat{S}_x, \hat{C}] - \lambda_c \cdot \hat{C} - \lambda_s \cdot \hat{S}_x, \{\lambda_c, \lambda_s\} \in \mathfrak{R} \quad (22)$$

Differentiating $L[f, \hat{S}_x, \hat{C}]$ with respect to \hat{C}, \hat{S}_x leads:

$$\frac{\partial I[f, \hat{S}_x, \hat{C}]}{\partial \hat{S}_x} - \lambda_s = 0 \quad (23a)$$

$$\frac{\partial I[f, \hat{S}_x, \hat{C}]}{\partial \hat{C}} - \lambda_c = 0 \quad (23b)$$

Let $\hat{Q} \triangleq e^{-\hat{C}}, Q(f) \triangleq e^{-C(f)}$ and solving the quadratic equation that follows from (23), we have two sets of solutions for $\{\hat{S}_x, \hat{Q}\}$ (see [18] for full derivation) Define:

$$\psi_i = \begin{cases} 1 & i = 1 \\ -1 & i = 2 \end{cases} \quad (24)$$

$$X(f; \lambda_c, \lambda_s) \equiv H(f)^2 - \lambda_s - \lambda_c H(f)^2 \quad (25)$$

$$S_{x,i}(f) = \frac{X(f; \lambda_c, \lambda_s) + \psi_i \sqrt{X(f; \lambda_c, \lambda_s)^2 - 4H(f)^2 \lambda_c \lambda_s}}{2H(f)^2 \lambda_s} \quad (26a)$$

$$Q_i(f) = \frac{X(f; \lambda_c, \lambda_s) - \psi_i \sqrt{X(f; \lambda_c, \lambda_s)^2 - 4H(f)^2 \lambda_c \lambda_s}}{2H(f)^2 (1 - \lambda_c)} \quad (26b)$$

Although equation (26) produce two curve sets, we discard the $\{S_{x,2}(f), Q_2(f)\}$ curve, since for each frequency (regardless of $H(f)$), $I[f, \hat{S}_x, \hat{Q}]$ is not concave in the pair $\{S_{x,2}(f), Q_2(f)\}$, meaning that one could improve the "optimal" $I[f, \hat{S}_x, \hat{Q}]$ simply by splitting the incremental frequency band, a rigorous proof can be found within [18].

In stark contrast to classical water-pouring [11] and [10], the optimal solution will be discontinuous frequently. An outstanding example is $H(f)$ constant over f , sufficient SNR and rather low C . In this case an attempt to use frequency-constant $S_x(f)$, $C(f)$ will place us in the non-concave region and a better performing solution will use only part of the available spectra to utilize the available bits better by transmitting less information about the channel noise, and since $C(f)$ and $S(f)$ will never fall gradually down to zero, the transition will have always an abrupt part.

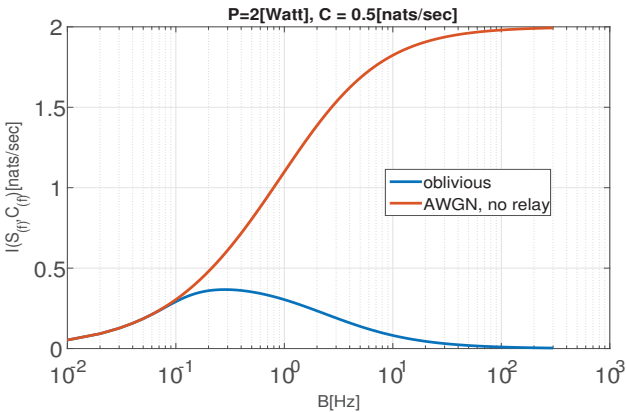


Fig. 5: Information rate as a function of allocated bandwidth

Figure 5 demonstrates this idea. Assuming a flat channel (i.e. $H(f) = 1$). For a given total power $P = 2$ [Watt] and capacity $C = 0.5$ [nats/sec] we calculated the Mutual Information Rate when distributing the power and bit-rate uniformly over the bandwidth B .

$$I[S_x(f), C(f)] = B \log \left[\frac{1 + P/B}{1 + P/B e^{-C/B}} \right] \quad (27)$$

It is clear that the best course would not be to spread the power and bit-rate all over the spectrum.

IV. NUMERICAL ANALYSIS

The algorithm has been tested for different types of channels (assigned as "Channel A") of the form $H_A(f) \equiv \alpha_1 N(f_1, 1) + \alpha_2 N(f_2, 1)$ (when $N(\mu, \sigma^2)$ is the Gaussian curve) with $P = 100$, $C = 9$ [nats/sec]. We also tested the "reciprocal" channel - assigned as "channel B" (i.e., $H_B(f) = \max[H_A(f)] -$

$H_A(f)$) In each scenario, we've compared the overall information rate using the following algorithms:

- The proposed algorithm
- Uniform allocation of rate and power
- Classical water pouring as presented in [3], for the case of $C \rightarrow \infty$
- "Limited Rate Water Pouring", which is:
 - A) Calculate $S_x(f)$ using the classical water pouring approach.
 - B) The allocated rate is: $C(f) = \frac{C}{P} S_x(f)$

The results are summarized at the next figures and table. Each figure contains

- a Normalized curves of $S_x(f)$, $C(f)$ of the proposed algorithm with respect to $H(f)$
- b Comparison between the allocated power using the proposed approach and the classical water pouring

It is clear from the results that:

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

$$I_n^\infty(x, z) \geq I_n^C(x, z) \geq I_n^C(x, z)|_{\text{Limited Rate Water Pouring}} \geq I_n^C(x, z)|_{\text{Uniform Allocation}} \quad (28)$$

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

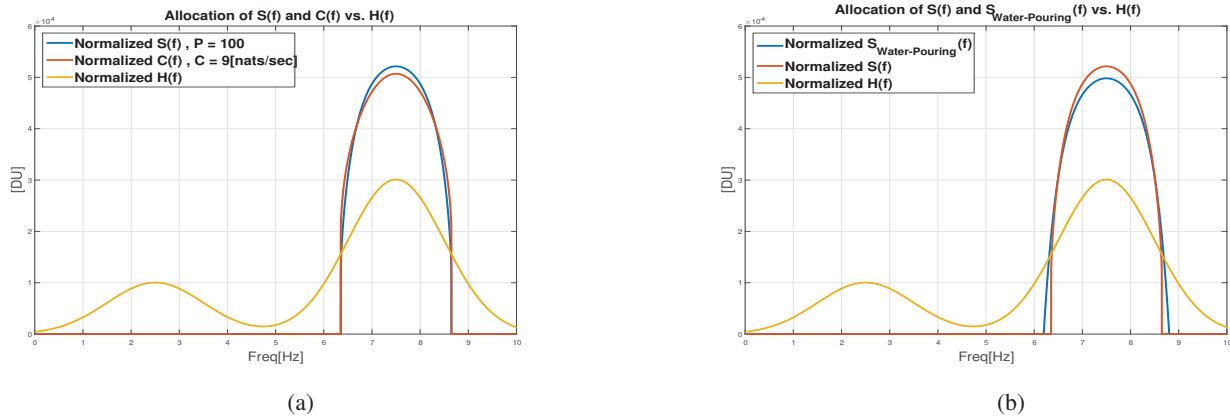
V. CONCLUSIONS AND FUTURE DIRECTIONS

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 as well as to the Gaussian bottleneck frameworks. The advantage of this approach over other methods was demonstrated. Our results apply directly also to the frequency dependent vector (MIMO) channels presented in figure 1 and in section II above. Such channels can be transformed to a set of parallel independent channels [19]. Thus equation (26) and the optimization algorithms can be applied on those with no need of 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 [20]).

An area under current investigation is the Constrained $H(z)$ for general memoryless channels. Let us replace the constraint $I(y; z) \leq C$ by

$$H(z) \leq C \quad (29)$$

It should be emphasized that the entropy constraint is per *symbol*. In this case we can apply insights from the results of [21]. This will result in a universal compression scheme, where standard algorithms, as the Lempel Ziv, Arithmetic coding and others can be applied [11]. This will also enforce z to a discrete alphabet since otherwise $H(z)$ is infinite. Open questions:

Fig. 6: $\alpha_1 = 0.25, f_1 = 0.25W, \alpha_2 = 0.25, f_2 = 0.75W$

Case	α_1	f_1	α_2	f_2	Remark	$I_n^C(x, z)$	$I_n^\infty(x, z)$	$I_n^C(x, z)$ - Uni- form Allocation	$I_n^C(x, z)$ - Limited Rate Water Pouring
1	0.25	0.25W	0.75	0.75W	Channel A , Fig. 6	2.83	2.94	0.77	2.80
2	0.25	0.25W	0.75	0.75W	Channel B	3.40	4.53	1.98	3.03
3	0	0	1	0.5W	Channel A	3.73	3.98	0.98	3.68
4	0	0	1	0.5W	Channel B	4.98	7.85	3.28	4.62
5	-	-	-	-	Allpass channel	7.92	23.98	7.75	7.75

TABLE I: Comparison Between Channels (all units are [nats/sec])

- A) Is the curve of the maximal $I(x; z)$ as a function of the maximal permitted $H(z)$ concave? It is of cause concave if time sharing between two different quantizers is allowed.
- B) In which cases does the optimal mapping $y \rightarrow z$ become deterministic? In some cases it is stochastic, for example if y is binary there are only two possible deterministic quantizers, one of which is $z = y$ and the other $z = const.$ So $H(z)$ has only two possible values. In such cases the quantizer will be stochastic or time sharing between different deterministic quantizers will be used. With deterministic quantizer the scheme becomes similar to Lloyd quantizer concatenated with entropy compression but may be superior to it because the quantizer is optimized jointly for $I(x; z)$ and $H(z)$. On the other hand there is a trivial way to transfer any stochastic quantizer to a deterministic one if the $x \rightarrow y$ channel is an additive noise channel.

ACKNOWLEDGEMENT

This project has received funding from the European Union's Horizon 2020 Research And Innovation Programme under grant agreement no. 694630.

REFERENCES

- [1] Shamai S., Steinberg Y., Sanderovich A., and Kramer G. Communication via decentralized processing. *IEEE Trans. Information Theory*, 54(7):3008–3023, jul 2008.
- [2] A. Globerson G. Chechik, Y. Weiss and N. Tishbi. Information bottleneck for gaussian variables. *Machine Learning Research*, 6:165–188, jan 2005.
- [3] C.E. Shannon. Communication in the presence of noise. *Proc. IRE*, 37(1):10–21, jan 1949.
- [4] A. Sanderovich, S. Shamai, and Y. Steinberg. Distributed mimo receiver-achievable rates and upper bounds. *IEEE Trans. Information Theory*, 55(10):4419–4438, oct 2009.
- [5] A. Winkelbauer and G. Matz. Rate-information-optimal gaussian channel output compression. *48th Annual Conference on Information Sciences and Systems (CISS)*, (1-5), aug 2014. doi: 10.1109/CISS.2014.6814120.
- [6] S. Farthofer A. Winkelbauer and G. Matz. The rate-information trade-off for gaussian vector channels. *IEEE International Symposium on Information Theory (ISIT)*, pages 2849 – 2853, jul 2014.
- [7] T. Berger. *Rate-Distortion Theory*. Wiley Online Library, 2003.
- [8] R. Dobrushin and B. Tsybakov. information transmission with additional noise. *IRE Transactions on Information Theory*, 8(5):293–304, 1962.
- [9] A. Kipnis, A.J. Goldsmith, Y.C. Eldar, and T. Weissman. Distortion-rate function of sub-nyquist sampled gaussian sources. *IEEE Transactions on Information Theory*, 62:401–429, Jan 2016.
- [10] A. Kipnis, Y.C. Eldar, and A.J. Goldsmith. Sampling stationary signals subject to bitrate constraints. *arXiv:1601.06421*, 2016.
- [11] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Elements in Communications and Signal Processing. Wiley, tenth edition, aug 1991.
- [12] A. Globerson and N. Tishbi. On the optimality of the gaussian information bottleneck curve. *Unpublished*, 2003. HUJI TR.
- [13] N. Tishbi, F. Pereira, and W. Bialek. Information bottleneck method. *Proc. 37th Allerton Conf. on Communication, Control and Computing*, pages 368–377, sept 1999.
- [14] G. Chechik, A. Globerson, N. Tishby, and Y. Weiss. Information bottleneck for gaussian variables. *Journal of Machine Learning Research*, (6):165–188, 2005.
- [15] S. Shamai D. Guo and S. Verdú. The interplay between information and estimation measures. *Foundation and Trends in Signal Processing*, 6(4):243–429, 2012.
- [16] R. G. Gallager. *Information Theory and Reliable Communications*. John Wiley, New York, 1968.
- [17] I. M. Gelfand and S. V. Fomin. *Calculus of Variations*. Prentice-Hall, Englewood Cliffs, NJ, 1963.
- [18] A. Homri, M. Peleg, and S. Shamai. Oblivious processing in a fronthaul constrained gaussian channel. *arXiv:1510.08202*, 2015.
- [19] L. Brandenburg and A. Wyner. Capacity of the gaussian channel with memory: The multivariate case. *Bell System Technical Journal*, 53:745–779, may 1974.
- [20] J. Mietzner, R. Schober, L. Lampe, W. H. Gerstacker, and P. A. Hoeher. Multiple-antenna techniques for wireless communications - a comprehensive literature survey. *IEEE Communications Surveys & Tutorials*, 11(2), 2009.
- [21] H. Gish and J.N. Pierce. Asymptotically efficient quantization. *IEEE Transactions on Information Theory*, 14(5):676–684, sep 1968.