Information Bottleneck for an Oblivious Relay with Channel State Information: the Scalar Case

Giuseppe Caire[⋆], Shlomo Shamai[†], Antonia Tulino[∗], Sergio Verdu[‡], and Cagkan Yapar[⋆⋆]
[⋆] USC, Los Angeles CA, [†] Technion Haifa, [∗] Universitá Federico II Napoli, [‡] Princeton, NJ, [⋆⋆] TU Berlin

***Abstract*—We consider an extension of the information bottleneck problem where underlying Markov Chain is X–o–(Y,S)–o–Z, and where $P_{X,S,Y} = P_X P_S P_{Y|X,S}$ is the joint distribution of a source X, a channel state S independent of the source, and the channel output Y of a state-dependent channel. For the case $Y = SX + N$ with X, S and N Gaussian circularly symmetric, we provide an upper bound and two achievable lower bounds on the information bottleneck rate. We relate this problem to the case of an oblivious relay with channel state information. Our results show that simple symbol-by-symbol relay processing, possibly followed by "entropy coding" (data compression) yields a very effective method, virtually achieving the upper bound on a wide range of relevant system parameters.**

***Index Terms*—Gaussian Information Bottleneck, Channel State Information.**

I. Introduction and problem definition

The Information Bottleneck (IB) problem considers the Markov Chain X–o–Y–o–Z, where $P_{X,Y}$ is an assigned joint probability distribution on $\mathcal{X} \times \mathcal{Y}$ (the alphabet of (X,Y)) and $P_{Z|Y}$ is to be found as the solution of the constrained maximization problem

$$\text{maximize} \quad I(X;Z) \tag{1a}$$
$$\text{subject to} \quad I(Y;Z) \leq \mathsf{C}, \tag{1b}$$

where C is the bottleneck constraint parameter. The alphabet of Z may or may not be specified a priori, depending on the problem at hand [1]. This formulation was introduced by Tishby in [2], and it has been used to interpret the behavior of deep learning neural networks, where the evolution of the learning process via some training scheme (e.g., stochastic gradient back propagation) can be visualized on the IB plane of $I(X;Z)$ (relevant information on X) vs. C (representation rate of the observation Y, referred to as complexity, [3]).

From a more fundamental information theoretic viewpoint, the IB consists of a classical remote source coding problem [4], [5] under logarithmic distortion [6].

Another interesting scenario where the IB problem is relevant consists of the so-called *oblivious relay* (see [7]). Let X and Y be the input and the output of a communication channel, where a source node sends codewords from a given codebook, and the receiver is formed by a remote relay communicating to the actual decoder via an error-free link (digital pipe) of capacity C. The relay is *oblivious* in the sense that it cannot decode the information message itself. For example, the relay may represent the front-end of a digital receiver, that must produce some observation by simple signal processing, such as scalar quantization, and transmits such observation to the actual baseband processor via digital interface with constrained throughput C. A very relevant application of this scenario is when the front-end is placed in some "dumb" remote antenna head, and the processing (decoding) is performed in a centralized dedicated computer, located somewhere in the infrastructure "cloud" (see [7] and references therein). The oblivious feature of the relay can be modeled rigorously by assuming that the source and the destination make use of a codebook selected at random over a library, and that the relay is unaware of such random selection. Therefore, the relay must treat X as a random process with a distribution induced by the random selection over the codebook library (see [7] and references therein). In this case, the relay must produce some useful observation Z to the destination node, subject to the link constraint C. Then, it makes sense to find Z such that $I(X;Z)$ is maximized.

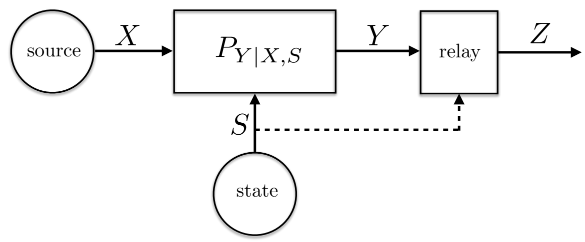

Fig. 1. Block diagram of the IB problem at hand.

In this paper we consider a slightly augmented version of the IB problem as shown in Fig. 1. Here, the joint distribution of X and Y depends on a state variable S, independent of X, such that $P_{X,Y,S} = P_X P_S P_{Y|X,S}$, where P_X, P_S, $P_{Y|X,S}$ are all assigned by the problem. The relay node has direct observation of the state variable S, and yet must produce Z to be sent to the destination node through the digital pipe of capacity C. This modified IB problem is given by

$$\text{maximize} \quad I(X;Z) \tag{2a}$$
$$\text{subject to} \quad I(Y,S;Z) \leq \mathsf{C} \tag{2b}$$

where the optimization is with respect to $P_{Z|Y,S}$. Intuitively, the relay can simply split its link capacity into two subchannels, and convey to the destination both (a quantized version

This work has been supported by the US-Israel Binational Science Foundation (BSF).

of) the state variable S and (a quantized version of) the channel output Y. Roughly speaking, this is what happens today in "naive" implementations of remote radio head systems. Nevertheless, it is important to notice that S is not subject to any distortion constraint, i.e., it is not needed at all (explicitly) at the destination. Hence, the relay can use its knowledge of S in order to operate a convenient transformation of Y into a new channel observation, which can be then conveyed to the destination.

Motivated by the operational significance provided by the oblivious relay with channel state information discussed above, in this paper we focus on the simple Gaussian iid scalar case where X is an iid Gaussian source with components $\sim \mathcal{CN}(0,1)$, the observation at the relay node is given by $V = (Y, S)$ with

$$Y = SX + N, \tag{3}$$

where the *additive noise* N is iid $\sim \mathcal{CN}(0, \sigma^2)$ and the *channel state* S is also iid $\sim \mathcal{CN}(0,1)$. That is a standard fading channel with fading known at the relay (where the signal is received), and where decoding is done at a remote (cloud) location. Of course, for $\mathsf{C} \to \infty$ (2) yields the well-known capacity of the iid fading channel with state known to the receiver, given by $I(X; Y|S) = \mathbb{E}[\log(1 + |S|^2/\sigma^2)]$ [8].

II. Informed receiver upper bound

Notice that for given $S = s$, fixed and known to all, the problem reduces to the classical scalar Gaussian bottleneck problem:

$$\begin{aligned} &\text{maximize} && I_s(X; Z) && \text{(4a)} \\ &\text{subject to} && I_s(Y; Z) \le \mathsf{C} && \text{(4b)} \end{aligned}$$

where we use the mutual information with subscript s to indicate that the reference measure is $P_X P_{Y|X,S=s} P_{Z|Y,S=s}$. It is well-known (see [9]) that the solution of (4) is obtained by letting Z jointly Gaussian with Y (for given $S = s$). This yields the bottleneck rate [10]

$$R_{\rm gb}(\gamma, \mathsf{C}) = \log\left(1 + \gamma\right) - \log\left(1 + 2^{-\mathsf{C}}\gamma\right), \tag{5}$$

where $\gamma = |s|^2/\sigma^2$ is the channel SNR for given state $S = s$ and the subscript "gb" stands for *Gaussian bottleneck*.

An obvious upper bound to our problem (2) (where S is known only to the relay), is obtained by letting S be known to both the relay and the destination. We refer to the case where S is known also to the destination as the *informed receiver upper bound*. The problem in this case takes on the form

$$\begin{aligned} &\text{maximize} && I(X; Z|S) && \text{(6a)} \\ &\text{subject to} && I(Y; Z|S) \le \mathsf{C} && \text{(6b)} \end{aligned}$$

This decomposes a set of "parallel" Gaussian bottleneck problems, for each given $S = s$, such that the solution of (6) is obtained in a "waterfilling-like" form as

$$\begin{aligned} &\text{maximize} && \mathbb{E}\left[R_{\rm gb}(\Gamma, c(\Gamma))\right] && \text{(7a)} \\ &\text{subject to} && \mathbb{E}[c(\Gamma)] \le \mathsf{C} && \text{(7b)} \\ & && c(\gamma) \ge 0, \ \ \forall\, \gamma \in \mathbb{R}_+ && \text{(7c)} \end{aligned}$$

where $\Gamma = |S|^2/\sigma^2$ is the (random) SNR and the function $c(\gamma)$ represents the allocation of the bottleneck rate C for each channel SNR $\Gamma = \gamma$.

The solution of (7a) is readily obtained using standard Lagrange multipliers and KKT conditions. This yields the "informed" upper bound which shall be denoted as UB0 in the following:

$$R^{\rm ub0}(\mathsf{C}) \triangleq \mathbb{E}\left[\log(1+\Gamma) - \log(1+\nu) \middle| \, \Gamma \ge \nu\right] \mathbb{P}(\Gamma \ge \nu), \tag{8}$$

where the optimized bottleneck capacity allocation is $c^\star(\gamma) = [\log(\gamma/\nu)]_+$, and the Lagrange multiplier value ν is the solution of

$$\mathbb{E}\left[\log(\Gamma/\nu) \middle| \, \Gamma \ge \nu\right] \mathbb{P}(\Gamma \ge \nu) = \mathsf{C}. \tag{9}$$

Noticing that $|S|^2$ is an exponentially distributed random variable with mean 1 and using the results [11]

$$\begin{aligned} \int_1^\infty \ln(t) e^{-at}\, dt &= \frac{1}{a} \mathrm{Ei}(1, a) && \text{(10a)} \\ \int_0^\infty \ln(1+t) e^{-at}\, dt &= \frac{e^a}{a} \mathrm{Ei}(1, a) && \text{(10b)} \end{aligned}$$

with $\mathrm{Ei}(n, a) = \int_1^{+\infty} \frac{e^{-at}}{t^n} dt$, $a > 0$, $n \ge 1$, and some simple change of variables, we arrive at [1]

$$R^{\rm ub0}(C) = \frac{e^{\sigma^2}}{\ln(2)} \mathrm{Ei}(1, \sigma^2) - \sigma^2 \int_0^\nu \log(1+\gamma) e^{-\sigma^2\gamma} d\gamma - \log(1+\nu) e^{-\sigma^2\nu} \tag{11}$$

with ν solution of

$$\mathrm{Ei}(1, \sigma^2\nu) = \ln(2)\mathsf{C}. \tag{12}$$

III. Achievable schemes

A key observation in oblivious relay processing is that the relay does not need to convey to the destination node information on both the channel output Y and the channel state S. In fact, it can pre-process (Y, S) and convey some function Z that maximizes the mutual information $I(X; Z)$. As a simple extreme example, suppose that S reduces to phase-noise only, i.e., $S = e^{j\Phi}$ for some phase random variable Φ. Thanks to the fact that the noise N is rotationally invariant, in this case the relay can perfectly "undo" the effect of the channel by computing $Y' = e^{-j\Phi} Y = X + N'$, where N' has the same statistics of N. At this point, the problem reduces to a standard Gaussian bottleneck, for which the optimal solution is known. More in general, "undoing" the channel state incurs some costs. In the following we present two simple achievable schemes based on the idea of simple scalar relay processing.

In passing, it should also be noticed here that in our setting, when $S \sim \mathcal{CN}(0,1)$, the strategy of canceling the phase of S without changing the noise statistics yields the channel $Y' = |S|X + N'$, where $|S|$ is a Rayleigh distributed random variable. At this point, even if the relay conveys Y' directly,

[1]We express all rates in bits, log and ln denote base-2 and natural logarithms.

i.e., lets $Z = Y'$ (obviously violating the IB constraint), the resulting channel seen at the end receiver is a channel with iid Rayleigh fading unknown to the receiver, whose high-SNR capacity behavior is know to be doubly logarithmic in the channel SNR [12], i.e., we have $I(X;Y') \leq \log\log(1/\sigma^2) + O(1)$. This shows that conveying implicitly or explicitly some information on the amplitude fading is extremely important and in fact, as simple as it may appear, a non-trivial problem in this context.

A. Quantized channel inversion at the relay

Our first proposed approach consists of using the relay to invert the channel on a symbol by symbol basis, i.e., multiplying by $S^*/|S|^2$. The resulting channel becomes

$$Y' = X + \sqrt{\xi}N', \tag{13}$$

where we define the random variable $\xi = |S|^{-2}$. At this point, the relay forces the channel to belong to a finite set of Gaussian channels by adding artificial noise (i.e., by introducing physical degradation). We fix a finite grid of K positive quantization points $\mathcal{B} = \{b_1 \leq b_2 \cdots \leq b_{K-1} < +\infty\}$ where $b_K = +\infty$, and define the ceiling operation

$$\lceil \xi \rceil_{\mathcal{B}} = \arg\min_{b \in \mathcal{G}} \{b \geq \xi\}. \tag{14}$$

Then, the degraded channel is given by

$$Y'' = X + \underbrace{\sqrt{\xi}N' + \sqrt{\lceil \xi \rceil_{\mathcal{B}} - \xi}W}_{N''}, \tag{15}$$

where $W \sim \mathcal{CN}(0,\sigma^2)$, independent of everything else. It follows that the variance of the equivalent noise is given by

$$\mathrm{Var}(N'') = \sigma^2(\xi + \lceil \xi \rceil_{\mathcal{B}} - \xi) = \sigma^2 \lceil \xi \rceil_{\mathcal{B}}.$$

Notice that this can be $+\infty$, such that the corresponding Gaussian channel has zero capacity.

Let L denote the random state quantization index, with pmf

$$P_L(\ell) = \mathbb{P}(\lceil \xi \rceil_{\mathcal{B}} = b_\ell), \tag{16}$$

such that $\sum_{\ell=1}^{K} P_L(\ell) = 1$. The number of quantization bits per channel use necessary to compress the channel quantization index L (treated as a discrete memoryless source) is given by $H(L)$. Thus, the number of bits per channel use available to represent the (quantized) channel observations is given by $\mathsf{C} - H(L)$. We treat the K channels in the form (15), as a parallel Gaussian channel model, for which we apply the standard parallel Gaussian bottleneck result. Namely, for each given channel state quantization index $L = \ell$, we have the achievable rate

$$R_\ell = \log\left(1 + \frac{1}{b_\ell \sigma^2}\right) - \log\left(1 + \frac{2^{-r_\ell}}{b_\ell \sigma^2}\right), \tag{17}$$

where r_ℓ denotes the partial bottleneck rate, i.e., the rate allocated to encode the output of the ℓ-th parallel channel. By construction, $R_K = 0$ and therefore we let $r_K = 0$. Then, the ergodic achievable bottleneck rate is $\mathbb{E}[R_L] = \sum_{\ell=1}^{K-1} P_L(\ell) R_\ell$. Define $\rho_\ell = 1/(b_\ell \sigma^2)$, for $\ell = 1, \ldots, K$, with $\rho_K = 0$. In order to optimize the bottleneck ergodic rate subject to the bottleneck (average rate) constraint, we need to solve the modified Waterfilling problem

$$\begin{aligned} \text{maximize} \quad & \sum_{\ell=1}^{K-1} P_L(\ell)\left(\log(1+\rho_\ell) - \log(1 + 2^{-r_\ell}\rho_\ell)\right) \\ \text{subject to} \quad & \sum_{\ell=1}^{K-1} P_L(\ell) r_\ell \leq \mathsf{C} - H(L) \\ & r_\ell \geq 0, \quad \text{for } \ell = 1, 2, \ldots, K-1. \end{aligned} \tag{18}$$

Notice that choosing the quantization levels as quantiles, we obtain the uniform pmf $P_L(\ell) = 1/K$ for all $\ell = 1, \ldots, K$, such that $H(L) = \log(K)$ and the solution of the waterfilling problem can be slightly simplified (see later).

In order to solve (18), we form the Lagrangian function

$$\begin{aligned} \mathcal{L}(\mathbf{r}, \lambda) = & \sum_{\ell=1}^{K-1} P_L(\ell)\left(\log(1+\rho_\ell) - \log(1+2^{-r_\ell}\rho_\ell)\right) \\ & - \lambda\left(\sum_{\ell=1}^{K-1} P_L(\ell) r_\ell - \mathsf{C} + H(L)\right) \end{aligned} \tag{19}$$

and set its partial derivatives with respect to the primal variables $\mathbf{r}$ to zero. The resulting equations must be discussed with respect to the KKT conditions for the non-negativity constraints $r_\ell \geq 0$. We have

$$\frac{\partial \mathcal{L}}{\partial r_k} = P_L(k)\left(\frac{2^{-r_k}\rho_k}{1 + 2^{-r_k}\rho_k} - \lambda\right) \tag{20}$$

Letting $a_k = 2^{-r_k}\rho_k$ and $\nu = \frac{\lambda}{1-\lambda}$, we find the solution $a_k = \nu$ for all k, yielding

$$r_k = \log\frac{\rho_k}{\nu}. \tag{21}$$

If $\rho_k \geq \nu$, then this solution is positive and therefore consistent with the KKT conditions. If instead $\rho_k < \nu$, this solution is not consistent. However, we notice that in this case letting $r_k = 0$ yields the derivative expression equal to

$$\frac{\rho_k}{1+\rho_k} - \frac{\nu}{1+\nu}. \tag{22}$$

Since the function $x/(1+x)$ is monotonically increasing for $x \in \mathbb{R}_+$, we have that if $\rho_k < \nu$ then the difference in (22) is negative. Hence, a negative derivative at the boundary implies that the objective function is maximized w.r.t. r_k by letting r_k equal to the boundary, i.e., $r_k = 0$. Summarizing, we have obtained the optimal value of $\mathbf{r}$ as

$$r_k^\star = \left[\log\frac{\rho_k}{\nu}\right]_+ \tag{23}$$

where $[x]_+$ denotes the positive part.

The value of the threshold ν can be found by imposing the average bottleneck constraint

$$\sum_{\ell=1}^{K-1} P_L(\ell)\left[\log\frac{\rho_\ell}{\nu}\right]_+ = \mathsf{C} - H(L). \tag{24}$$

This can be solved using a bisection line search method with respect to the one-dimensional parameter $\nu > 0$. In the special case of uniform probabilities $P_L(\ell) = 1/K$, the line search can be avoided and the solution can be found in a finite number of steps as follows. Without loss of generality, sort the values $\{\rho_\ell\}$ in non-increasing order. Then, there is some integer $1 \leq k \leq K-1$ for which

$$\sum_{\ell=1}^{k} \log \frac{\rho_\ell}{\nu} = K(\mathsf{C} - \log K), \tag{25}$$

and

$$\rho_{k+1} \leq \nu, \tag{26}$$

(counting $r_K = 0$), such that

$$\log \nu = \frac{1}{k} \sum_{\ell=1}^{k} \log \rho_\ell - \frac{K(\mathsf{C} - \log K)}{k}. \tag{27}$$

From the strict convexity of the problem and the uniqueness/existence of the solution, we have that such index k must exist and it is unique. Hence, we have just to test the above condition (for sorted values $\{\rho_\ell\}$), for $k = 1, 2, 3, \ldots$ till the conditions (25) and (26), with (27), are satisfied.

Replacing back the solution into the objective function in (18), we find the achievable bottleneck rate as

$$R^{\text{q-ch-inv}}(\mathsf{C}, K) = \sum_{\ell=1}^{K-1} P_L(\ell) \big[\log(1+\rho_\ell) - \log(1+\nu) \big]_+ . \tag{28}$$

B. Quantization of the MMSE estimate at the relay

Our second achievable scheme works as follows: the relay produces the MMSE estimate of X given (Y, S) and simply source-encode such estimate $\widehat{X}$. We let

$$\widehat{X} = \mathbb{E}[X|Y,S] \tag{29}$$

and treat $\widehat{X}$ as the new relay observation. At this point, we consider the modified problem

$$\begin{aligned} &\text{maximize} && I(X;Z) && (30a)\\ &\text{subject to} && I(\widehat{X};Z) \leq \mathsf{C}. && (30b) \end{aligned}$$

which falls into the "classical" bottleneck class of problems since $X \to \widehat{X} \to Z$.

In the case at hand, using $X \sim \mathcal{CN}(0,1)$ and $N \sim \mathcal{CN}(0,\sigma^2)$ yields explicitly

$$\widehat{X} = \frac{S^*}{\sigma^2 + |S|^2} Y = \frac{|S|^2}{\sigma^2 + |S|^2} X + \frac{S^*}{\sigma^2 + |S|^2} N. \tag{31}$$

Notice that $\widehat{X}$ is conditionally Gaussian given S, but it is non-Gaussian when removing such conditioning.

Then, we consider the source coding problem at the relay, where the relay encodes blocks of the estimated signal $\widehat{X}$ given in (31). Letting Z denote the representation variable for this quantization problem, standard rate-distortion theory yields that for any conditional distribution $P_{Z|\widehat{X}}$ the following rate-distortion pair is achievable:

$$R = I(\widehat{X}; Z), \quad D = \mathbb{E}[|\widehat{X} - Z|^2]. \tag{32}$$

We choose $P_{Z|\widehat{X}}$ to be a conditional Gaussian distribution, i.e., we let

$$Z = \widehat{X} + Q, \tag{33}$$

where $Q \sim \mathcal{CN}(0, D)$ is independent of anything else (following the random coding argument, this means that we use P_Z, i.e., the marginal of $P_{Z|\widehat{X}} P_{\widehat{X}}$ to generate a codebook ensemble of codes of length n and 2^{nR} points).

Now, using the fact that Gaussian input maximizes the mutual information for a Gaussian additive noise channel, we have the upper bound

$$I(\widehat{X}; Z) \leq I(\widehat{X}_g; Z) = \log\left(1 + \frac{\mathbb{E}[|\widehat{X}|^2]}{D}\right), \tag{34}$$

where $\widehat{X}_g$ denotes a Gaussian zero-mean circularly symmetric random variable with the same second moment of $\widehat{X}$. It follows that imposing $\log\left(1 + \frac{\mathbb{E}[|\widehat{X}|^2]}{D}\right) = \mathsf{C}$ automatically satisfies the bottleneck constraint in (30b). This yields the quantization noise variance

$$D = \frac{\mathbb{E}[|\widehat{X}|^2]}{2^{\mathsf{C}} - 1}. \tag{35}$$

The last step consists of evaluating the resulting achievable bottleneck rate, i.e., the mutual information $I(X;Z)$. We write

$$I(X;Z) = I(X,S;Z) - I(S;Z|X). \tag{36}$$

Noticing that

$$I(X,S;Z) \geq I(X;Z|S) = h(Z|S) - h(Z|X,S),$$

we obtain the lower bound

$$I(X;Z) \geq I(X;Z|S) - I(S;Z|X) = h(Z|S) - h(Z|X). \tag{37}$$

The conditional differential entropy $h(Z|S)$ is obtained exactly from (33) since by conditioning on S this is just a simple Gaussian additive noise channel. Therefore,

$$h(Z|S) = \mathbb{E}\left[\log\left(\pi e \left(\frac{|S|^2}{\sigma^2 + |S|^2} + D\right)\right)\right]. \tag{38}$$

In order to evaluate the term $h(Z|X)$, we can write an upper bound by replacing Z with a conditionally Gaussian Z_g given X with the same conditional variance of Z. Namely, we have

$$h(Z|X) \leq h(Z_g|X) = \mathbb{E}\left[\log\left(\pi e \mathrm{Var}(Z|X)\right)\right]. \tag{39}$$

From (33), defining for simplicity of notation the random variable $U = \frac{|S|^2}{\sigma^2+|S|^2}$, we have

$$\mathrm{Var}(Z|X) = \mathrm{Var}(U)|X|^2 + \mathbb{E}\left[\frac{U\sigma^2}{\sigma^2 + |S|^2}\right] + D. \tag{40}$$

Putting all together, we find the achievable bottleneck rate as

$$R^{\text{q-mmse}}(\mathsf{C}) = \mathbb{E}\left[\log\left(U + D\right)\right] - \mathbb{E}\left[\log\left(\mathrm{Var}(U)|X|^2 + \mathbb{E}\left[\frac{U\sigma^2}{\sigma^2 + |S|^2}\right] + D\right)\right] \tag{41}$$

IV. Numerical results

We evaluate the performance of the proposed achievable schemes in III and compare them with the upper bound derived in section II. For the quantized channel inversion scheme (Section III-A), we choose the quantization levels as quantiles. The channel SNR is defined as $\mathsf{SNR} = 1/\sigma^2$.

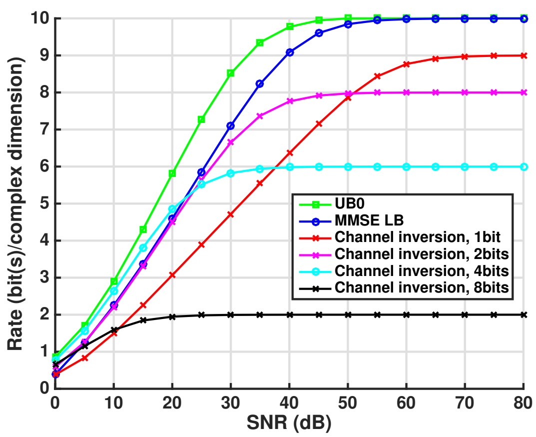

Fig. 2. Bottleneck rate $\mathsf{C} = 10$ bit/complex dimension.

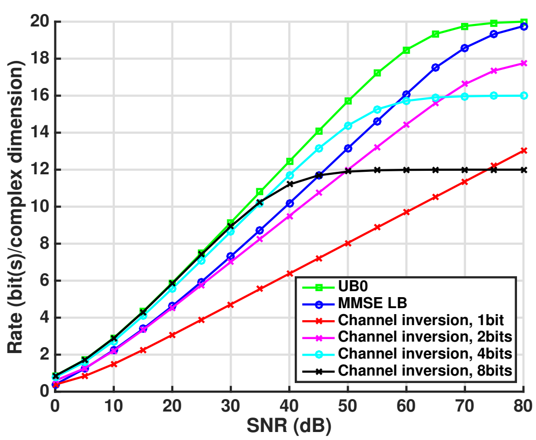

Fig. 3. Bottleneck rate $\mathsf{C} = 20$ bit/complex dimension.

We notice that for relatively large bottleneck constraint C and not extremely large SNR (see Fig. 2 vs. Fig. 3), the quantized channel inversion scheme can essentially match the informed receiver upper bound. In addition, there is a non-trivial optimal number of quantization bits (i.e., an optimal number of discrete SNR levels) which in general depends on C and on the operating SNR (see Fig. 2). This simple example suggests that some symbol-by-symbol processing at the oblivious relay can yield very effective schemes for this problem.

V. Conclusions

This work represents a preliminary study of a generally relevant problem, namely, optimal oblivious relay processing in the case where the channel state can be easily estimated at the relay node, but conveying such estimate to the destination has a rate cost, since the whole communication between the relay and the destination is constrained by an information bottleneck link. Although we have focused on a very simple iid Gaussian model with multiplicative (fading) state, we plan to extend the problem to considering the case of MIMO channels (multi-antenna relay) and, going further, to the case of multiple parallel relays [13], where the latter is particularly relevant to the centralized processing of multiple remote antennas, as in the so-called "cloud RAN" architectures (see [7]). The preliminary results exposed in this paper hint that very simple symbol-by-symbol processing at the relay nodes, followed by "entropy coding" data compression of quantization indices (in our case, achieving the rate $\log K$ bits in order to represent the K discrete quantization levels, as explained in Section III-A), can be a very effective approach.

References

[1] G. Pichler and G. Koliander, "Information bottleneck on general alphabets," *arXiv preprint arXiv:1801.01050*, 2018.

[2] N. Tishby, F. C. Pereira, and W. Bialek, "The information bottleneck method," *arXiv preprint physics/0004057*, 2000.

[3] R. Shwartz-Ziv and N. Tishby, "Opening the black box of deep neural networks via information," *arXiv preprint arXiv:1703.00810*, 2017.

[4] R. Dobrushin and B. Tsybakov, "Information transmission with additional noise," *IRE Transactions on Information Theory*, vol. 8, no. 5, pp. 293–304, 1962.

[5] H. Witsenhausen, "Indirect rate distortion problems," *IEEE Transactions on Information Theory*, vol. 26, no. 5, pp. 518–521, 1980.

[6] T. A. Courtade and T. Weissman, "Multiterminal source coding under logarithmic loss," *IEEE Transactions on Information Theory*, vol. 60, no. 1, pp. 740–761, 2014.

[7] I. E. Aguerri, A. Zaidi, G. Caire, and S. Shamai, "On the capacity of cloud radio access networks with oblivious relaying," *arXiv preprint arXiv:1710.09275*, 2017.

[8] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2619–2692, 1998.

[9] G. Chechik, A. Globerson, N. Tishby, and Y. Weiss, "Information bottleneck for gaussian variables," *Journal of machine learning research*, vol. 6, no. Jan, pp. 165–188, 2005.

[10] A. Winkelbauer and G. Matz, "Rate-information-optimal gaussian channel output compression," in *Information Sciences and Systems (CISS), 2014 48th Annual Conference on*. IEEE, 2014, pp. 1–5.

[11] M. S. Alouini and A. J. Goldsmith, "Capacity of rayleigh fading channels under different adaptive transmission and diversity-combining techniques," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 4, pp. 1165–1181, Jul 1999.

[12] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2426–2467, 2003.

[13] I. Estella Aguerri and A. Zaidi, "Distributed information bottleneck method for discrete and gaussian sources," in *The International Zurich Seminar on Information and Communication (IZS 2018) Proceedings*. ETH Zurich, 2018, pp. 35–39.