# On Mismatched Oblivious Relaying

Michael Dikshtein, Nir Weinberger, and Shlomo Shamai (Shitz)
Department of Electrical and Computer Engineering, Technion, Haifa 3200003, Israel
Email: {michaeldic@, nirwein@, sshlomo@ee.}technion.ac.il

2023 IEEE International Symposium on Information Theory (ISIT) | 978-1-6654-7554-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISIT54713.2023.10206842

***Abstract*—We consider the problem of reliable communication over a *discrete memoryless channel* (DMC) with the help of a relay, termed the *information bottleneck* (IB) channel. There is no direct link between the source and the destination, and the information flows in two hops. The first hop is a noisy channel from the source to the relay. The second hop is a noiseless but limited-capacity backhaul link from the relay to the decoder. We further assume that the relay is oblivious to the transmission codebook. We examine two mismatch scenarios. In the first setting, we assume the decoder is restricted to use some fixed decoding rule, which is mismatched to the actual channel. In the second setting, we assume that the relay is restricted to use some fixed compression metric, which is again mismatched to the statistics of the relay input. We establish bounds on the *random-coding* capacity of both settings, some of which are shown to be ensemble tight.**

## I. Introduction

In this paper, we consider the point-to-point oblivious relay channel [1], in which a relay observes the transmitted signal over a noisy channel and transmits digital information to the decoder via a limited-capacity link. Our primary focus in this paper is achievable rate results under various *mismatch* conditions, specifically using random codes. We consider both mismatched decoding at the receiver as well as mismatched compression of the relay. The motivation for this analysis is that network architectures with oblivious processing at the relays serve as the fundamental building blocks of modern communication systems. A recent and comprehensive summary on oblivious communication networks can be found in [2]. The motivation for analyzing the mismatched case is that in many practical settings, .e.g., in the up-link of cellular communication with oblivious relays, the relay or the decoder only possess partial information regarding the statistical model of the channel or is restricted to operating with some specific decoding metric (and typically use it in a more elaborate channel decoding algorithm, such as belief-propagation [3]).

The capacity of the oblivious relay channel is tightly connected to the *information bottleneck* (IB) problem [4], which has been the subject of a recent extensive study, mainly due to its relation to current advances in machine learning, see e.g., [5], [6], and statistical learning [7]. In the classical rate-distortion theory of lossy source coding, a fidelity measure must be chosen that quantifies the quality of compression [8]. The IB method determines this distortion measure via an additional dependent random variable that captures the meaningful information in the data to be compressed, which can be thought of as contextual labeling of the data. Specifically, the compression quality under the IB approach is assessed via the mutual information between the compressed representation and the additional variable. As it turns out, the resulting IB matches precisely the capacity of the oblivious relay channel under consideration. In effect, it is also the single-letter rate-distortion formula for remote-source coding setting [9], [10] when the distortion measure is the log-loss [11]. It was shown in [12] that minimizing the log-loss minimizes an upper bound to any choice of loss functions for binary classification problems. It can also be shown that log-loss actually bounds general distortion measures (as stated by Linder, but to the best of our knowledge, it has not been published).

As is well known, the problem of mismatched decoding is notoriously challenging and is not fully resolved, even for standard point-to-point channels [13]. Nonetheless, analysis of random codes under mismatched encoding or decoding leads to tractable achievable bounds, and so we adopt this analysis for the oblivious relay channel (or the IB problem) studied in this paper. Coding over a DMC with mismatched decoder under the random coding regime was introduced independently in [14] and [15], where a lower bound (termed the *LM rate*) on the capacity was derived. In [16] it was shown that the LM bound is not tight. A more analytically tractable lower bound, termed *generalized mutual information* (GMI), was proposed in [17], and the random coding ensemble tightness for the LM scheme was established in [18]. In our setting, the analysis of random codes is further motivated by the desire to model codes that are not adapted to a specific communication setting, by the obliviousness nature of the relay, and by security aspects typically involved in relay communication systems.

The outline of the rest of the paper and our contributions are as follows. In Sec. II, we consider the IB problem with a mismatched decoder at the receiver and derive random coding (achievable) rates, both in the form of an LM bound, as well as a GMI bound. We then propose an algorithm for the computation of the achievable rate and exemplify its operation on a quaternary channel. We then extend the GMI rate to continuous alphabet channels and demonstrate this result for a Gaussian fading channel. Afterward, in Sec. III, we consider the setting of a relay with a mismatched compression rule. In Sec. IV, we conclude the paper.

*Related Work:* A comprehensive summary on information-theoretic foundations of mismatched decoding and encoding is provided in [13]. Beyond channel coding, mismatch has also been studied in the context of source coding [19], [20]. A successive refinement setting constrained to Gaussian codebooks with minimal Euclidean distance encoding has been proposed in [21]. An extension to general alphabets has been recently presented in [22]. The global channel knowledge at the destination setting has been studied in the context of quantized distributed reception in [23], where a simple, complex binary sign quantization has been assumed. A distributive decoding communication network

Authorized licensed use limited to: Technion Israel Institute of Technology. Downloaded on August 23,2023 at 09:04:39 UTC from IEEE Xplore. Restrictions apply.

with BPSK transmission over the AWGN channel has been considered in [24]. Uniform quantization for OFDMA-based CRAN has been considered in [25]. Outage probability in the problem of distributed reception with hard decision exchanges has been considered in [26].

## II. Information Bottleneck Channel with Mismatched Decoder

### A. Discrete Memoryless Channels

In this section, we consider the 3-node point-to-point communication system with a relay depicted in Figure 1, in which the sender wishes to communicate a message $\mathsf{M}$ to the receiver with the help of the relay. We term this setting the *discrete memoryless information bottleneck channel* (DM-IBC) with mismatched decoder $\langle \mathcal{X}, \mathsf{P}_{\mathsf{Y}|\mathsf{X}}, \mathcal{Y}, \mathcal{Z}, V(z|x) \rangle$. It consists of three finite sets $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, a collection of conditional **pmf**s $\mathsf{P}_{\mathsf{Y}|\mathsf{X}}$ on $\mathcal{Y}$ (one for each input symbol $x$), and a decoding metric $V(z|x)$ on $\mathcal{Z}$.

A $(2^{nR}, 2^{nB}, n)$ code for the DM-IBC with mismatched decoder consists of:

- a message set $\mathcal{M} = [1 : 2^{nR}]$,
- a representation set $\mathcal{W} = [1 : 2^{nB}]$,
- an encoder that assigns a codeword $x^n(m)$ to each message $m \in \mathcal{M}$,
- a relay source encoder that assigns an index $w \in \mathcal{W}$ to each received sequence $y^n \in \mathcal{Y}^n$, with the respective reconstructed sequence $z^n(w)$, and
- a decoder that assigns an estimate $\hat{m}$ or an error message e to each received representation index $w \in \mathcal{W}$ according to some fixed mismatched metric

$$\hat{m}(z^n) = \arg\max_{m \in \mathcal{M}} V(z^n | x^n(m)) = \arg\max_{m \in \mathcal{M}} \prod_{i=1}^{n} V(z_i | x_i(m)). \quad (1)$$

It is assumed that the message $\mathsf{M}$ is uniformly distributed over the message set $\mathcal{M}$.

*Definition 2.1:* Consider a codebook of $2^{nR}$ $n$-dimensional sequences, $\{x^n(m)\}_{m=1}^{2^{nR}}$, where each sequence is generated at random with a memoryless **pmf** $\mathsf{P}_\mathsf{X}$ and independently of all other vectors. Also consider a compression codebook of $2^{nB}$ $n$-dimensional sequences, $\{z^n(w)\}_{w=1}^{2^{nB}}$, where each sequence is generated at random with a memoryless **pmf** $\mathsf{P}_\mathsf{Z}$ and independently of all other vectors. A pair of such channel-compression codebooks is termed a *random codebook*. Let $P_e^{(n)} = \mathsf{P}\{\hat{\mathsf{M}} \neq \mathsf{M}\}$ denote the error probability averaged over the random codebooks. A rate $R$ is said to be *achievable* for the DM-IBC at compression rate $B$, with mismatched decoding metric $V$, if $\lim_{n\to\infty} P_e^{(n)} = 0$ under the mismatched decoding rule. The *random coding capacity* $C_V(B)$ of the DM-IBC with mismatched decoder is the supremum of all achievable rates of random codebooks.

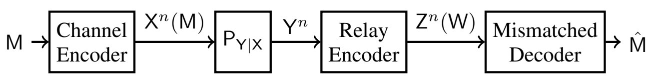


Fig. 1: Oblivious Communication System with Mismatched Decoder

Our main result of this section is stated in the following theorem, and it describes the respective LM rate [14], [15] for the oblivious relay setting.

*Theorem 1:* The random coding capacity of the DM-IBC with mismatched decoder is

$$\begin{aligned} C_V(B) = \quad & \max_{\mathsf{P}_\mathsf{X}} \max_{\mathsf{P}_{\mathsf{Z}|\mathsf{Y}}} \quad J_V(\mathsf{X};\mathsf{Z}) \\ & \text{subject to} \quad I(\mathsf{Y};\mathsf{Z}) \leq B, \end{aligned} \quad (2)$$

where

$$\begin{aligned} J_V(\mathsf{X};\mathsf{Z}) = \quad & \min_{\mathsf{Q}_{\mathsf{Z}|\mathsf{X}} \in \mathcal{P}_{\mathcal{X}\times\mathcal{Z}}} \quad I(\mathsf{P}_\mathsf{X}, \mathsf{Q}_{\mathsf{Z}|\mathsf{X}}) \\ & \text{subject to} \quad \sum_{x\in\mathcal{X}} \mathsf{Q}(x,z) = \mathsf{P}_\mathsf{Z}(z), \\ & \qquad \sum_{(x,z)\in\mathcal{X}\times\mathcal{Z}} \mathsf{Q}(x,z)\log V(z|x) \geq -D, \end{aligned} \quad (3)$$

with $\mathsf{Q}(x,z) = \mathsf{Q}_{\mathsf{Z}|\mathsf{X}}(z|x) \cdot \mathsf{P}_\mathsf{X}(x)$ and

$$D \triangleq -\sum_{x\in\mathcal{X}} \sum_{z\in\mathcal{Z}} \mathsf{P}_{\mathsf{XZ}}(x,z) \log V(z|x). \quad (4)$$

Theorem 1 provides the *exact* random coding capacity under mismatched decoding. In specific, both an upper bound and ensemble tightness are proved.

*Proof:* The proof of the direct part appears in [27, App. B] and the proof of ensemble tightness appears in [27, App. C]. ■

The resulting capacity expression of Theorem 1 is proved for the discrete case and is not easily extended to continuous channels. Therefore, in what follows, we provide an achievable random-coding rate based on the *generalized mutual information* (GMI) coding scheme [17]. As we shall see, this rate can be extended to continuous channels. Note that the main difference between the LM rate from Theorem 1 and the GMI rate provided in the following theorem is the relaxation of the marginals equality constraint, i.e., $\mathsf{Q}_\mathsf{X} = \mathsf{P}_\mathsf{X}$.

*Theorem 2:* The random coding capacity of the DM-IBC with mismatched decoder is lower bounded as

$$\begin{aligned} C_V(B) \geq \quad & \max_{\mathsf{P}_\mathsf{X}, \mathsf{P}_{\mathsf{Z}|\mathsf{Y}}} \quad I_{\mathsf{GMI}}(\mathsf{P}_{\mathsf{XZ}}) \\ & \text{subject to} \quad I(\mathsf{Y};\mathsf{Z}) \leq B, \end{aligned} \quad (5)$$

where

$$\begin{aligned} I_{\mathsf{GMI}}(\mathsf{P}_{\mathsf{XZ}}) = \quad & \min_{\tilde{\mathsf{P}}_{\mathsf{XZ}}} \quad \mathsf{D}\left(\tilde{\mathsf{P}}_{\mathsf{XZ}} \| \mathsf{P}_\mathsf{X} \times \mathsf{P}_\mathsf{Z}\right) \\ & \text{s.t.} \quad \sum_{x\in\mathcal{X}} \tilde{\mathsf{P}}_{\mathsf{XZ}}(x,z) = \mathsf{P}_\mathsf{Z}(z) \\ & \qquad \mathbb{E}_{\tilde{\mathsf{P}}_{\mathsf{XZ}}}\left[V(\mathsf{Z}|\mathsf{X})\right] \geq \mathbb{E}_{\mathsf{P}_{\mathsf{XZ}}}\left[V(\mathsf{Z}|\mathsf{X})\right]. \end{aligned} \quad (6)$$

In addition, the inner minimization problem has the following dual form,

$$I_{\mathsf{GMI}}(\mathsf{P}_{\mathsf{XZ}}) = \max_{\lambda\geq 0} \sum_{(x,z)\in\mathcal{X}\times\mathcal{Z}} \mathsf{P}_{\mathsf{XZ}}(x,z) \log \frac{V(z|x)^\lambda}{\sum_{x'} \mathsf{P}_\mathsf{X}(x') V(z|x')^\lambda}. \quad (7)$$

*Proof:* The proof appears in [27, App. D]. ■

Authorized licensed use limited to: Technion Israel Institute of Technology. Downloaded on August 23,2023 at 09:04:39 UTC from IEEE Xplore. Restrictions apply.

### *B. A Computationally Efficient Algorithm*

In this section, we propose an efficient algorithm to compute $C_V(B)$ by solving the optimization problem of Theorem 1. To this end, it should be noted that the inner minimization problem of computing $J_V(\mathsf{X};\mathsf{Z})$ in (3) is a convex optimization problem in $\mathsf{Q}_{\mathsf{Z}|\mathsf{X}}$ since the mutual information is a convex function of the channel and the constraints on $\mathsf{Q}_{\mathsf{Z}|\mathsf{X}}$ are linear. Therefore, $J_V(\mathsf{X};\mathsf{Z})$ can be efficiently computed using standard convex optimization solvers.

By contrast, the outer *maximization* problem in (2) over $\mathsf{P}_{\mathsf{Z}|\mathsf{Y}}$ is not concave, and solving it requires global optimization methods, e.g., a grid search. Specifically, we propose to initially compute this maximum over a coarse grid of the probability simplex. Then, we perform a refined search of the maximum in a finer grid, only at a local neighborhood of the solution of the coarse maximization. Repeating this refinement procedure in an iterative manner then leads to an improved solution at each step, and a stopping criterion may be a negligible increase in the achievable rate in the last iteration. We nonetheless emphasize that any choice of $\mathsf{P}_{\mathsf{Z}|\mathsf{Y}}$ leads to an achievable lower bound on the rate. Thus, the crucial optimization step for the validity of the solution is the convex minimization step in (3). In addition, one may also optimize the input distribution $\mathsf{P}_\mathsf{X}$, again, using search methods. In many practical applications, however, the input distribution is arbitrarily chosen, e.g., as a uniform distribution, which is typically justified by the symmetry of the problem.

The proposed algorithm involves an alternating maximization step, proposed in [4] for the original IB problem, in order to find the optimal test-channel in the case that the decoder is matched. As well known, this algorithm is based on the Blahut-Arimoto [28], [29] algorithm and is termed here *information bottleneck alternating minimization* (IBAM). A formal description can be found, e.g., in [4, Thm. 5]. Our main algorithm is detailed in Algorithm 1.

### *C. Example: A Quaternary Channel*

In this section, we demonstrate the result of Theorem 1 and the operation of Algorithm 1 in a simple setting. Concretely, suppose that the channel from $\mathsf{X}$ to $\mathsf{Y}$ is defined by the following conditional **pmf**,

$$\mathsf{P}_{\mathsf{Y}|\mathsf{X}}(y|x) = \mathsf{P}\left\{\mathsf{Y}=y|\mathsf{X}=x\right\} = \begin{cases} 1-\epsilon, & y=x \\ \frac{\epsilon}{2}, & y = x\pm 1 \\ 0, & \text{otherwise,} \end{cases} \quad (8)$$

where all addition and subtraction operations are computed modulo 4. The metric $V$ is mismatched, and specifically, is matched to a different channel. This different channel has the same correct symbol transition probability of $1-\epsilon$, yet the error can go to the other three alternative symbols with probability $\frac{\epsilon}{3}$, i.e.,

$$\mathsf{Q}_{\mathsf{Y}|\mathsf{X}}(y|x) = \mathsf{P}\left\{\mathsf{Y}=y|\mathsf{X}=x\right\} = \begin{cases} 1-\epsilon, & y=x \\ \frac{\epsilon}{3}, & y\neq x. \end{cases} \quad (9)$$

Due to symmetry, we assume that the optimal $\mathsf{P}_\mathsf{X}$ is uniform and $\mathsf{P}_{\mathsf{Z}|\mathsf{Y}}$ is symmetric. In such case, $\mathsf{P}_{\mathsf{Z}|\mathsf{Y}}(z|y)$ is a modulo-additive channel. We compare the rates obtained by a matched vs. a mismatched decoder in Figure 2, as a function of $\epsilon$. As might be expected, the difference between the rates vanishes as $\epsilon \to 0$ since the mismatch between the channels becomes milder as $\epsilon \to 0$.

**Algorithm 1:** MMIB algorithm

**Input:** $\mathsf{P}_{\mathsf{XY}}, V(z|x), B, \mathsf{RES}$
$\mathsf{P}^{opt}_{\mathsf{Z}|\mathsf{Y}} = \text{IBAM}(\mathsf{P}_{\mathsf{XY}}, C, args)$
$\mathcal{P}_{\mathcal{Z}|\mathcal{Y}} = \{\mathsf{P}_{\mathsf{Z}|\mathsf{Y}} \colon I(\mathsf{Y};\mathsf{Z}) = B\}$
$\mathcal{P}^{\text{coarse}}_{\mathcal{Z}|\mathcal{Y}} = \mathcal{P}_{\mathcal{Z}|\mathcal{Y}} \cap \mathsf{RES}\cdot\mathbb{Z}^{|\mathcal{Z}|}$
**for** $\mathsf{P}_{\mathsf{Z}|\mathsf{Y}} \in \mathcal{P}^{coarse}_{\mathcal{Z}|\mathcal{Y}}$ **do**
  Compute $D$ according to (4) with $\mathsf{P} = \mathsf{P}_{\mathsf{Z}|\mathsf{Y}}\cdot\mathsf{P}_{\mathsf{XY}}$
  $J_V(\mathsf{X};\mathsf{Z}) = \min_Q I(\mathsf{P}_\mathsf{X}, \mathsf{Q}_{\mathsf{Z}|\mathsf{X}})$
  subject to:
  $\mathbb{E}_{\mathsf{P}_Q}\left[\log V(\mathsf{Z}|\mathsf{X})\right] \geq -D$
Find $\mathsf{P}^*_{\mathsf{Z}|\mathsf{Y}} \in \mathcal{P}^{\text{coarse}}_{\mathcal{Z}|\mathcal{Y}}$ that maximizes $J_V(\mathsf{X};\mathsf{Z})$.
**while** $\Delta R \geq \epsilon$ **do**
  RES = RES $*$ FINE
  $\mathcal{P}^{\text{fine}}_{\mathcal{Z}|\mathcal{Y}} = \mathcal{P}_{\mathcal{Z}|\mathcal{Y}} \cap \mathsf{FINE}\cdot\mathbb{Z}^{|\mathcal{Z}|} \cap \mathcal{B}(\mathsf{P}^*_{\mathsf{Z}|\mathsf{Y}}, 1/\text{RES})$
  **for** $\mathsf{P}_{\mathsf{Z}|\mathsf{Y}} \in \mathcal{P}^{fine}_{\mathcal{Z}|\mathcal{Y}}$ **do**
    Compute $D$ according to (4) with $\mathsf{P} = \mathsf{P}_{\mathsf{Z}|\mathsf{Y}}\cdot\mathsf{P}_{\mathsf{XY}}$
    $J_V(\mathsf{X};\mathsf{Z}) = \min_Q I(\mathsf{P}_\mathsf{X}, \mathsf{Q}_{\mathsf{Z}|\mathsf{X}})$
    subject to:
    $\mathbb{E}_{\mathsf{P}_Q}\left[\log V(\mathsf{Z}|\mathsf{X})\right] \geq -D$
  Find $\mathsf{P}^*_{\mathsf{Z}|\mathsf{Y}} \in \mathcal{P}^{\text{fine}}_{\mathcal{Z}|\mathcal{Y}}$ that maximizes $J_V(\mathsf{X};\mathsf{Z})$.
**Output:** $\mathsf{P}^*_{\mathsf{Z}|\mathsf{Y}}$, $R$

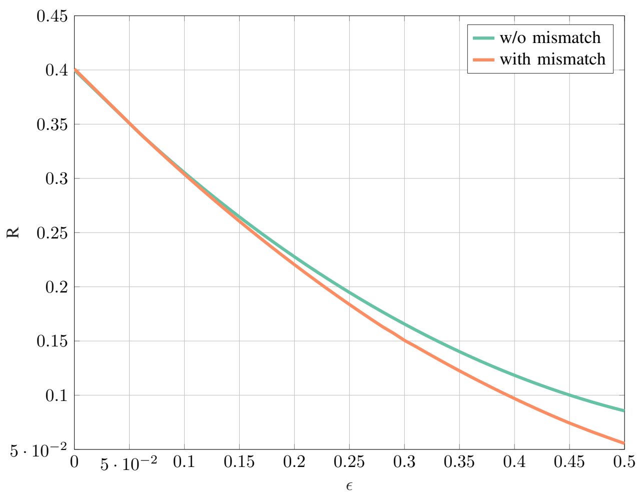


Fig. 2: Mismatched performance of the Quaternary example.

### *D. Continuous-Alphabet Memoryless Channels*

In this section, we modify the bound of the previous section to continuous alphabet Gaussian channels. As mentioned, the discrete alphabet assumption is crucial to the derivation of the LM rate in Theorem 1. Indeed, in the standard point-to-point communication setting, without mismatch, coding theorems for continuous alphabets are obtained by taking the limit of

Authorized licensed use limited to: Technion Israel Institute of Technology. Downloaded on August 23,2023 at 09:04:39 UTC from IEEE Xplore. Restrictions apply.

fine quantization of the continuous inputs and outputs of the channel. Unfortunately, this technique is not applicable in the mismatched case, since quantization of the output changes the decoder which should be fixed by assumption, and it becomes very challenging to track and evaluate the impact on the final result. For this purpose, we have also derived the corresponding GMI rate in Theorem 2, as a lower bound, which, as we shall see, is amenable to modification from discrete alphabet channels to continuous alphabet channels.

From a technical notation perspective, we first replace all **pmf**s to the form of probability densities. Furthermore, the decoding metric $V(z|x)$ will also be defined on $\mathbb{R} \times \mathbb{R}$. In addition, we add a constraint on the input of the channel, otherwise, as happens in most continuous channels, the capacity may be unbounded. Thus, every transmitted sequence $x^n$ must satisfy $\frac{1}{n}\sum_{i=1}^{n} c(x_i) \leq \Gamma$, for some cost function $c(x)$ and threshold $\Gamma$. Usually, we take $c(x) = x^2$ and $\Gamma$, both represent a power constraint and the maximum permitted power per-symbol.

*Theorem 3:* For any continuous oblivious relay channel with mismatched decoding metric $V$, input cost function $c(\cdot)$ and input cost threshold $\Gamma$, the random coding error probability vanishes for rate $R$ that satisfies:

$$\begin{aligned} R \leq \quad & \max_{f_{\mathsf{XZ}}, f_{\mathsf{Z|Y}}} \quad I_{\mathsf{GMI}}(f_{\mathsf{XZ}}) \\ & \text{s.t.} \qquad I(\mathsf{Y};\mathsf{Z}) \leq B, \quad \mathbb{E}\left[c(\mathsf{X})\right] \leq \Gamma. \end{aligned} \tag{10}$$

where

$$I_{\mathsf{GMI}}(\mathsf{P}_{\mathsf{XZ}}) = \max_{\lambda \geq 0} \int \mathrm{d}x \mathrm{d}z f_{\mathsf{XZ}}(x,z) \log \frac{V(z|x)^{\lambda}}{\int \mathsf{P}_{\mathsf{X}}(x') V(z|x')^{\lambda} \mathrm{d}x'}. \tag{11}$$

*Proof:* The dual expression from (7) can also be derived directly (rather than deriving the dual optimization problem as shown in [27, App. D]) using a similar analysis to that of Gallager [30] for maximum-likelihood decoding. The former involves replacing summations with integrals and using a standard expurgation argument to construct a sub-codebook with feasible codewords from the randomly generated codebook. ■

### E. Example: Fading Channel

We next exemplify our result on a fading channel, a fundamental wireless communication channel model. As common, we assume that the channel is complex-valued, and the additive noise is *circularly symmetric complex Gaussian* (CSCG). Specifically, we consider a memoryless time-varying fast-fading model of the form $\mathsf{Y}_i = \mathsf{H}_i\mathsf{X}_i + \mathsf{N}_i$, where $\mathsf{X}_i \in \mathbb{C}$ is the input, $\mathsf{N}_i \in \mathbb{C}$ is additive noise, and $\mathsf{H}_i \in \mathbb{C}$ is a fading coefficient. We assume that $\{\mathsf{N}_i\}_{i=1}^n$ are i.i.d. distributed according to $\mathcal{CN}(0,\sigma^2)$, and that $\mathsf{H}_i$ is an i.i.d. sequence with density function $S_{\mathsf{H}}$.

*1) Perfect Channel Knowledge:* If each random realization $\mathsf{H}_i = h_i$ is perfectly known at the decoder, then due to Gaussianity of the noise and the optimality of Gaussian compression assuming Gaussian input distribution, the optimal decoding rule would be the following weighted version of the nearest-neighbor rule:

$$\hat{m} = \arg\min_{j=1,\dots,M} \sum_{i=1}^{n} |z_i - h_i x_i^{(j)}|^2. \tag{12}$$

Similarly, under a power constraint $\mathbb{E}\left[|\mathsf{X}|^2\right] \leq \Gamma$ (i.e. $c(x) = |x|^2$) and a Gaussian input distribution assumption, the optimal rate is achieved using a Gaussian test channel from $\mathsf{Y}$ to $\mathsf{Z}$, and is given by

$$R_{\text{IB}}^{CG}(\Gamma,\sigma^2,S_{\mathsf{H}}) = \mathbb{E}\left[\log \frac{|\mathsf{H}|^2\Gamma + \sigma^2 + q}{\sigma^2 + q}\right]. \tag{13}$$

Evidently, in a fast-fading channel, it is unrealistic to assume that the decoding is matched, since this requires perfect knowledge of $\mathsf{H}_i$ at any time point.

*2) Imperfect Channel Knowledge:* As said, assuming Gaussian signaling, the pair $(\mathsf{X},\mathsf{Y})$ is jointly Gaussian. In the standard IB setting without mismatch, the optimal test-channel from $\mathsf{Y}$ to $\mathsf{Z}$ is also Gaussian in such case. Therefore, we also adopt this test channel for the mismatched setting, and assume that the channel from $\mathsf{Y}$ to $\mathsf{Z}$ is Gaussian, i.e., there exists a $\mathsf{W} \sim \mathcal{CN}(0,q)$ such that $\mathsf{Z} = \mathsf{Y}+\mathsf{W} = \mathsf{H}\mathsf{X}+\mathsf{N}+\mathsf{W}$. The value of $q$ is determined as the solution of the mutual information constraint equation, i.e.,

$$B \geq I(\mathsf{Y};\mathsf{Z}) = \mathbb{E}_{\mathsf{H}}\left[\log \frac{|\mathsf{H}|^2\Gamma + \sigma^2 + q}{q}\right]. \tag{14}$$

Let $q^*$ be the solution to the above equation. We use it to find the GMI rate of Theorem 2 using the dual form.

We adopt a simple uncertainty model in which

$$\mathsf{H}_i = \hat{\mathsf{H}}_i + \Delta_i, \quad \mathbb{E}\left[\Delta|\hat{\mathsf{H}}_i\right] = 0, \tag{15}$$

where $\hat{\mathsf{H}}_i$ is a possibly-random estimate of $\mathsf{H}$ known at the decoder, and $\Delta$ represents an unknown conditionally zero-mean error term. We make the simplifying assumption that the pairs $\{(\hat{\mathsf{H}}_i,\Delta_i)\}_{i=1}^n$ are i.i.d. with respect to $i = 1,\dots,n$, and independent of the channel input and noise.

In the case that the joint density function of $(\hat{\mathsf{H}}_i,\Delta_i)$ is unknown (or even when it is known but difficult to design a corresponding optimal coding scheme), it is natural to apply weighted nearest-neighbor coding

$$\hat{m} = \arg\min_{j=1,\dots,M} \sum_{i=1}^{n} |z_i - \hat{h}_i x_i^{(j)}|^2. \tag{16}$$

This is a mismatched decoding rule, in the sense that is would be optimal under a model of the form $\mathsf{Y} = \hat{\mathsf{H}}\mathsf{X} + \mathsf{N}$. The corresponding decoding metric is given by $V(x,z) = e^{-|z-\hat{h}x|^2}$.

*Theorem 4:* Consider the complex-valued channel fading setup with a known estimate $|\hat{\mathsf{H}}|$ at the output and a conditionally zero-mean error term $\Delta$. Under i.i.d. random coding with $\mathsf{X} \sim \mathcal{CN}(0,\Gamma)$, along with weighted nearest-neighbor decoding, the GMI rate is given by

$$\begin{aligned} C(B) = \quad & \max_{q} \quad \mathbb{E}\left[\log\left(1 + \frac{|\hat{\mathsf{H}}|^2\Gamma}{\mathbb{E}\left[|\Delta|^2|\hat{\mathsf{H}}\right]\Gamma + \sigma^2 + q}\right)\right] \\ & \text{s.t.} \qquad \mathbb{E}_{\mathsf{H}}\left[\log \frac{|\mathsf{H}|^2\Gamma + \sigma^2 + q}{q}\right] \leq B. \end{aligned} \tag{17}$$

*Proof:* The proof is omitted here due to lack of space. ■

We illustrate Theorem 4 using a numerical example. We choose $\Gamma = \sigma = 1$ and $\hat{\mathsf{H}}$ and $\Delta$ to be independent Rayleigh

Authorized licensed use limited to: Technion Israel Institute of Technology. Downloaded on August 23,2023 at 09:04:39 UTC from IEEE Xplore. Restrictions apply.

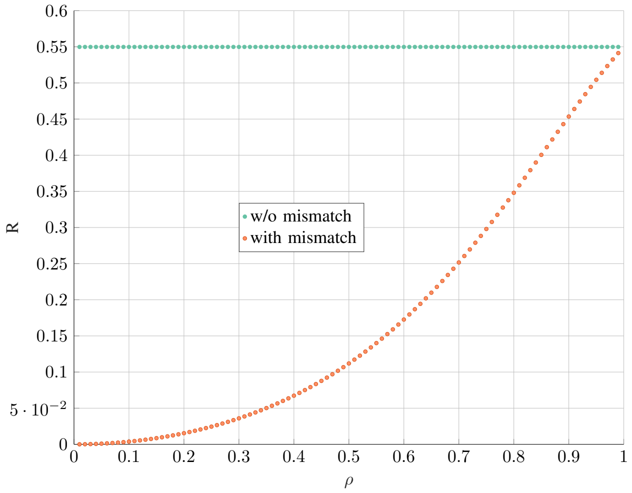


Fig. 3: Mismatched performance of the fading example.

random variables with $\sigma^2_{\hat{\mathsf{H}}} = \rho^2$ and $\sigma^2_{\Delta} = 1 - \rho^2$. The results are shown in Figure 3.

## III. Information Bottleneck Channel with Mismatched Relay

In this section, we consider a different relay model with a mismatch. Specifically, let us consider the 3-node point-to-point communication system with a mismatched relay depicted in Figure 4. In this model, the sender wishes to communicate a message $\mathsf{M}$ to the receiver with the help of the relay. We consider the *discrete memoryless information bottleneck channel* (DM-IBC) with mismatched relay $\langle \mathsf{X}, \mathsf{P}_{\mathsf{Y}|\mathsf{X}}, \mathcal{Y}, \mathcal{Z}, d_0(y,z) \rangle$ that consists of three finite sets $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, a collection of conditional **pmf**s $\mathsf{P}_{\mathsf{Y}|\mathsf{X}}$ on $\mathcal{Y}$ (one for each $x$), and an encoding metric $d_0(y,z)$ on $\mathcal{Y} \times \mathcal{Z}$.

A $(2^{nR}, 2^{nB}, n)$ code for the DM-IBC with mismatched relay defined in a similar manner to Section II with the following exceptions:

- a mismatched relay encoder that assigns an index $\hat{w} \in \mathcal{W}$ to each received sequence $y^n \in \mathcal{Y}^n$ according to $\hat{w} = \arg\min_{w \in \mathcal{W}} d_0^n(y^n, z^n(w))$ where

$$d_0^n(y^n, z^n) = \frac{1}{n} \sum_{i=1}^{n} d_0(y_i, z_i), \tag{18}$$

- a decoder that assigns an estimate $\hat{m}$ or an error message e to each received representation index $w \in \mathcal{W}$.

We assume that the decoder knows the channel and mismatched relay's codebook. Furthermore, it is assumed that the message $\mathsf{M}$ is uniformly distributed over the message set $\mathcal{M}$.

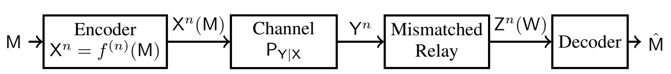


Fig. 4: Oblivious Communication System with Mismatched Relay

*Remark 1:* The mismatched compression problem is somewhat simpler in case the mismatch at the relay is the result of a wrong test-channel for compression, that is, the relay is constrained to joint typicality encoding with $\mathsf{Q}_{\mathsf{Z}|\mathsf{Y}}$. This is equivalent to a choice of nonoptimal test-channel in the standard IB problem. In such case, the resulting capacity is given by

$$C(B, \mathsf{Q}_{\mathsf{Z}|\mathsf{Y}}) = \begin{cases} \max_{\mathsf{P}_\mathsf{X}} I(\mathsf{X};\mathsf{Z}), & I(\mathsf{P}_\mathsf{Y}, \mathsf{Q}_{\mathsf{Z}|\mathsf{Y}}) \le B \\ 0, & \text{otherwise} \end{cases} \tag{19}$$

where $\mathsf{Q}_{\mathsf{Z}|\mathsf{X}}(z|x) = \sum_{y \in \mathcal{Y}} \mathsf{Q}_{\mathsf{Z}|\mathsf{Y}}(z|y) \mathsf{P}_{\mathsf{Y}|\mathsf{X}}(y|x)$. Note the difference between this setting and the mismatched relay compression setting. In the standard IB problem, a test-channel $\mathsf{P}_{\mathsf{Z}|\mathsf{Y}}$ is optimized, and a random compression codebook is generated according to $\mathsf{P}_\mathsf{Z}$. Given this codebook, the compressed index is chosen based on joint typicality encoding with the given $\mathsf{P}_{\mathsf{YZ}}$. By contrast, in the mismatched relay setting, the compressed codeword is chosen according to the given fixed (mismatched) metric, which is not necessarily matched to $\mathsf{P}_{\mathsf{YZ}}$.

Our main result for this setting is stated in the following theorem.

*Theorem 5:* The capacity of the mismatched relay channel is lower bounded as:

$$C(B) \ge \max_{\mathsf{P}_\mathsf{X}, \mathsf{P}_\mathsf{Z}} \min_{\mathsf{Q}_{\mathsf{YZ}} \in \mathcal{Q}} I(\mathsf{X};\mathsf{Z}) \tag{20}$$

where $\mathcal{Q} = \arg\min_{\mathsf{Q}_{\mathsf{YZ}} \in \mathcal{D}} \sum_{y \in \mathcal{Y}} \sum_{z \in \mathcal{Z}} \mathsf{Q}_{\mathsf{YZ}}(y,z) d_0(y,z)$ and

$$\mathcal{D} = \left\{ \substack{\mathsf{Q}_{\mathsf{YZ}}:\ \mathsf{Q}_\mathsf{Y} = \mathsf{P}_\mathsf{Y}, \mathsf{Q}_\mathsf{Z} = \mathsf{P}_\mathsf{Z}, \\ \mathsf{Q}_{\mathsf{XZ}}(x,z) = \sum_{y \in \mathcal{Y}} \mathsf{P}_{\mathsf{X}|\mathsf{Y}}(x|y) \mathsf{Q}_{\mathsf{YZ}}(y,z), I(\mathsf{Y};\mathsf{Z}) \le B} \right\}. \tag{21}$$

*Proof:* The proof appears in [27, App. E]. ■

## IV. Summary and Outlook

We considered the problem of reliable communication in a point-to-point oblivious-relay communication system with a mismatch. In particular, we considered mismatch at the relay or at the decoder. We have established ensemble tight achievable rates and their dual representations. We further specialized those results to particular instances: the quaternary channel and the fading channel. We proposed an alternating algorithm to find those rates.

For future work, it would be interesting to consider converse bounds to this problem, e.g., using the methods described in [13] and [31]. Alternatively, it would be interesting to find relations between the mismatch capacity of the channel $\mathsf{P}_{\mathsf{Y}|\mathsf{X}}$ to that of the entire channel $\mathsf{P}_{\mathsf{Z}|\mathsf{X}}$. Another possibility is to generalize the results to a state-dependent channel [32], [33], where the relay knows the state sequence (which may also be assumed to be i.i.d.), and add a description of the state as part of its message, in an effort to provide the receiver with channel state information, thus aiding its decoding performance.

## Acknowledgment

The work has been supported by the European Union's Horizon 2020 Research And Innovation Programme, grant agreement no. 694630 and by the ISF under Grant 1791/17. The research of N. Weinberger was supported by the Israel Science Foundation (ISF), grant no. 1782/22.

Authorized licensed use limited to: Technion Israel Institute of Technology. Downloaded on August 23,2023 at 09:04:39 UTC from IEEE Xplore. Restrictions apply.

## References


[1] A. Sanderovich, S. Shamai, Y. Steinberg, and G. Kramer, “Communication via decentralized processing,” *IEEE Trans. Inf. Theory*, vol. 54, pp. 3008–3023, July 2008.

[2] I. Estella Aguerri, A. Zaidi, G. Caire, and S. Shamai Shitz, “On the capacity of cloud radio access networks with oblivious relaying,” *IEEE Trans. Inf. Theory*, vol. 65, pp. 4575–4596, July 2019.

[3] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[4] N. Tishby, F. C. N. Pereira, and W. Bialek, “The information bottleneck method,” in *37th Annu. Allerton Conf. Commun. Control Comput.*, pp. 368—-377, Sept. 1999.

[5] A. Zaidi, I. Estella-Aguerri, and S. Shamai (Shitz), “On the information bottleneck problems: Models, connections, applications and information theoretic views,” *Entropy*, vol. 22, no. 2, 2020.

[6] Z. Goldfeld and Y. Polyanskiy, “The information bottleneck problem and its applications in machine learning,” *IEEE J. Sel. Areas Inf. Theory*, vol. 1, pp. 19–38, May 2020.

[7] I. E. Aguerri and A. Zaidi, “Distributed variational representation learning,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, pp. 120–138, Jan. 2021.

[8] C. E. Shannon *et al.*, “Coding theorems for a discrete source with a fidelity criterion,” *IRE Nat. Conv. Rec*, vol. 4, no. 142-163, p. 1, 1959.

[9] R. Dobrushin and B. Tsybakov, “Information transmission with additional noise,” *IRE Transactions on Information Theory*, vol. 8, pp. 293–304, Sept. 1962.

[10] J. Wolf and J. Ziv, “Transmission of noisy information to a noisy receiver with minimum distortion,” *IEEE Trans. Inf. Theory*, vol. 16, pp. 406–411, July 1970.

[11] T. A. Courtade and T. Weissman, “Multiterminal source coding under logarithmic loss,” *IEEE Trans. Inf. Theory*, vol. 60, pp. 740–761, Jan. 2014.

[12] A. Painsky and G. Wornell, “On the universality of the logistic loss function,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT 2018)*, pp. 936–940, May 2018.

[13] J. Scarlett, A. G. i Fàbregas, A. Somekh-Baruch, and A. Martinez, “Information-theoretic foundations of mismatched decoding,” *Foundations and Trends® in Communications and Information Theory*, vol. 17, no. 2–3, pp. 149–401, 2020.

[14] J. Y. N. Hui, *Fundamental issues of multiple accessing*. PhD thesis, Massachusetts Institute of Technology, 1983.

[15] I. Csiszar and J. Korner, “Graph decomposition: A new key to coding theorems,” *IEEE Trans. Inf. Theory*, vol. 27, pp. 5–12, Jan. 1981.

[16] I. Csiszar and P. Narayan, “Channel capacity for a given decoding metric,” *IEEE Trans. Inf. Theory*, vol. 41, pp. 35–43, Jan. 1995.

[17] G. Kaplan and S. Shamai, “Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment,” *AEU. Archiv für Elektronik und Übertragungstechnik*, vol. 47, no. 4, pp. 228–239, 1993.

[18] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai Shitz, “On information rates for mismatched decoders,” *IEEE Trans. Inf. Theory*, vol. 40, pp. 1953–1967, Nov. 1994.

[19] A. Lapidoth, “On the role of mismatch in rate distortion theory,” *IEEE Trans. Inf. Theory*, vol. 43, pp. 38–47, Jan. 1997.

[20] L. Zhou, V. Y. F. Tan, and M. Motani, “Refined asymptotics for rate-distortion using Gaussian codebooks for arbitrary sources,” *IEEE Trans. Inf. Theory*, vol. 65, pp. 3145–3159, May 2019.

[21] L. Bai, Z. Wu, and L. Zhou, “Achievable refined asymptotics for successive refinement using Gaussian codebooks,” *arXiv preprint arXiv:2208.03926*, 2022.

[22] M. Kanabar and J. Scarlett, “Mismatched rate-distortion theory: Ensembles, bounds, and general alphabets,” *arXiv preprint arXiv:2203.15193*, 2022.

[23] J. Choi, D. J. Love, D. R. Brown, and M. Boutin, “Quantized distributed reception for MIMO wireless systems using spatial multiplexing,” *IEEE Trans. Signal Process.*, vol. 63, pp. 3537–3548, July 2015.

[24] D. R. Brown, M. Ni, U. Madhow, and P. Bidigare, “Distributed reception with coarsely-quantized observation exchanges,” in *2013 47th Annual Conference on Information Sciences and Systems (CISS)*, pp. 1–6, 2013.

[25] L. Liang, S. Bi, and R. Zhang, “Joint power control and fronthaul rate allocation for throughput maximization in OFDMA-based cloud radio access network,” *arXiv preprint arXiv:1407.3855*, 2014.

[26] R. Wang, D. Richard Brown, M. Ni, U. Madhow, and P. Bidigare, “Outage probability analysis of distributed reception with hard decision exchanges,” in *2013 Asilomar Conference on Signals, Systems and Computers*, pp. 597–601, 2013.

[27] M. Dikshtein, N. Weinberger, and S. S. Shitz, “On mismatched oblivious relaying,” *arXiv preprint arXiv:2305.00930*, 2023.

[28] R. Blahut, “Computation of channel capacity and rate-distortion functions,” *IEEE Trans. Inf. Theory*, vol. 18, pp. 460–473, July 1972.

[29] S. Arimoto, “An algorithm for computing the capacity of arbitrary discrete memoryless channels,” *IEEE Trans. Inf. Theory*, vol. 18, pp. 14–20, Jan. 1972.

[30] R. G. Gallager, *Information theory and reliable communication / Robert G. Gallager.* New York: Wiley, 1968.

[31] A. Somekh-Baruch, “Upper bounds on the mismatched reliability function and capacity using a genie receiver,” *arXiv preprint arXiv:2203.08524*, 2022.

[32] H. Xu, T. Yang, G. Caire, and S. Shamai (Shitz), “Information bottleneck for a Rayleigh fading MIMO channel with an oblivious relay,” *Information*, vol. 12, no. 4, 2021.

[33] H. Xu, T. Yang, G. Caire, and S. Shamai Shitz, “Information bottleneck for an oblivious relay with channel state information: the vector case,” in *IEEE Int. Symp. Inf. Theory (ISIT 2021)*, pp. 2483–2488, July 2021.


Authorized licensed use limited to: Technion Israel Institute of Technology. Downloaded on August 23,2023 at 09:04:39 UTC from IEEE Xplore. Restrictions apply.