

Nonasymptotic Oblivious Relaying and Variable-Length Noisy Lossy Source Coding

Yanxiao Liu, Sepehr Heidari Advary and Cheuk Ting Li

Abstract

The information bottleneck channel (or the oblivious relay channel) concerns a channel coding setting where the decoder does not directly observe the channel output. Rather, the channel output is relayed to the decoder by an oblivious relay (which does not know the codebook) via a rate-limited link. The capacity is known to be given by the information bottleneck. We study finite-blocklength achievability results of the channel, where the relay communicates to the decoder via fixed-length or variable-length codes. These two cases give rise to two different second-order versions of the information bottleneck. Our proofs utilize the nonasymptotic noisy lossy source coding results by Kostina and Verdú, the strong functional representation lemma, and the Poisson matching lemma. Moreover, we also give a novel nonasymptotic variable-length noisy lossy source coding result.

Index Terms

Finite blocklength, oblivious relay channel, lossy source coding, channel simulation, network information theory.

I. INTRODUCTION

In the *oblivious relay channel* [1]–[3] (also referred to as the *information bottleneck channel* [4], [5]), the encoder encodes a message M into a sequence $X^n = (X_1, \dots, X_n)$ via random coding and sends it through a memoryless channel $P_{Y|X}$. An oblivious relay receives Y^n and transmits a description W to the decoder, which attempts to decode M . See Figure 1 for an illustration. The relay is *oblivious* in the sense that it does not know the random codebook used by the encoder and the decoder. As shown by [1], the minimum asymptotic description rate (as the blocklength $n \rightarrow \infty$) needed to support a message transmission rate C is given by the *information bottleneck* [6], also known as the *relevance-compression function* [7]

$$\text{IB}_{X \rightarrow Y}(C) := \min_{P_{U|Y}: I(X;U) \geq C} I(Y;U), \quad (1)$$

where we assume $X \rightarrow Y \rightarrow U$ forms a Markov chain.

In this paper (which is the complete version of [8]),¹ to study the trade-off between description rate and message rate when the blocklength n is limited, we show a second-order achievability result for the information bottleneck channel in terms of a natural second-order version of the information bottleneck, which we call the *var-information bottleneck*:

$$\text{VIB}_{X \rightarrow Y}(C) := \text{Var}[\iota_{Y;U}(Y;U) - \lambda^* \iota_{X;U}(X;U)], \quad (2)$$

where $\iota_{Y;U}(y;u) = \log \frac{P_{U|Y}(u|y)}{P_U(u)}$ is computed by the optimal $P_{U|Y}$ in (1), $\iota_{X;U}$ is similar, and $\lambda^* := \frac{d}{dC} \text{IB}_{X \rightarrow Y}(C)$.

For fixed-length description, we show that a rate

$$\text{IB}(C) + \sqrt{\frac{1}{n} \text{VIB}(C) Q^{-1}(\epsilon)} + O\left(\frac{\log n}{n}\right) \quad (3)$$

suffices when the blocklength is n and the error probability is ϵ , where $Q^{-1}(\cdot)$ is the inverse of the Q -function. This is shown by using the Poisson matching lemma [9], and by relating the information bottleneck channel to noisy lossy source coding, where we can utilize the second-order results in [10].

We also study a setting where the description sent by the relay can be variable-length and encoded by a prefix-free code. In this case, the second-order achievability result is instead given in terms of the *conditional-var-information bottleneck*:

$$\text{CVIB}_{X \rightarrow Y}(C) := \mathbb{E}[\text{Var}[\lambda^* \iota_{X;U}(X;U) | Y, U]]. \quad (4)$$

Note that CVIB is generally smaller than VIB since VIB is the variance of $\iota(Y;U) - \lambda^* \iota(X;U)$, whereas CVIB is its (expected) conditional variance given Y, U . We show that for variable-length description, it suffices to use a description rate

$$(1 - \epsilon) \left(\text{IB}(C) + \sqrt{\frac{\ln n}{n} \text{CVIB}(C)} \right) + O\left(\frac{1}{\sqrt{n}}\right). \quad (5)$$

Comparing (3) and (5), we see that variable-length and fixed-length have vastly different finite blocklength behavior.

Yanxiao Liu and Cheuk Ting Li are with the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong, China.

This paper was presented in part at the 2025 IEEE International Symposium on Information Theory (ISIT).

This work was partially supported by two grants from the Research Grants Council of the Hong Kong Special Administrative Region, China [Project No.s: CUHK 24205621 (ECS), CUHK 14209823 (GRF)].

¹Compared to the conference paper [8], this full version includes the complete proofs of all the theorems, as well as updated descriptions in the literature review and the main sections. New figures have also been provided.

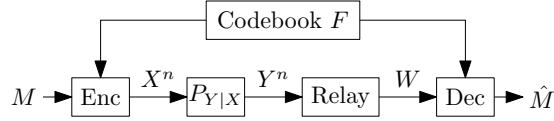


Fig. 1. Information bottleneck channel (or oblivious relaying).

These results are proved using techniques in noisy lossy source coding [11], where we have a 2-discrete memoryless source X^n, Y^n , the encoder observes Y^n and sends a description to the decoder, which recovers Z^n and aims at having a small probability of excess distortion $\mathbb{P}(d(X^n, Z^n) > D) \leq \epsilon$. Given by [11], the optimal asymptotic description rate is $R(D) := \min_{P_{Z|Y}: \mathbb{E}[d(X, Z)] \leq D} I(Y; Z)$. A second-order characterization for the optimal length of the fixed-length description

$$nR(D) + \sqrt{n\tilde{V}(D)}Q^{-1}(\epsilon) + O(\log n),$$

was shown in [10], where $\tilde{V}(D) := \text{Var}[\iota_{Y, Z^*}(Y; Z^*) + \lambda^* d(X, Z^*)]$, $P_{Z^*|Y}$ attains the minimum in $R(D)$, and $\lambda^* := -R'(D)$. In this paper, we show that for a variable-length description, we can achieve an expected length

$$(1 - \epsilon) \left(nR(D) + \sqrt{(n \ln n) \widetilde{CV}(D)} \right) + O(\sqrt{n}),$$

where $\widetilde{CV}(D) := (\lambda^*)^2 \mathbb{E}[\text{Var}[d(X, Z^*) | Y, Z^*]]$. This is proved by techniques in [10] and Poisson functional representation [9], [12].

II. RELATED WORKS

A. Information Bottleneck Channel and Oblivious Relaying

The setting of oblivious relay processing [1]–[3] was motivated by the architecture of modern communication networks (cloud radio access networks [13]), where access points are connected to a central server via rate-limited, error-free links. The scenario involving multiple oblivious relays was investigated by [1] (also see a 2-cascaded-relays setting [14]), while [2] explored cases where encoders can switch among different codebooks and relay nodes have access to certain scheduling information. Driven by the lack of knowledge about codebooks, mismatched decoding at the decoder and mismatched compression at the relay were examined in [4]. For the information bottleneck channel, error exponents have been studied in [5] and [15], where the former considers the channel under constant-composition codes through random generation of compress-forward codebooks, while the latter utilizes the type covering lemma [16] with refined analyses using the method of types. Furthermore, the results of oblivious relaying are closely tied to the information bottleneck problem [17], which has been extensively studied over the past two decades due to its strong connections to machine learning [18], [19]. The solution to the information bottleneck problem (1) aligns with the capacity of the oblivious relay channel and also with the noisy lossy source coding problem [11], [20] under a logarithmic distortion function [21].

B. Noiseless and Noisy Lossy Source Coding

In the conventional noiseless lossy source coding setting, the encoder observes the source directly (i.e., $X^n = Y^n$ in the noisy lossy source coding setting). The optimal length for a fixed-length code was characterized to the second-order in [22], [23], based on the rate-dispersion function [24]. For variable-length codes, the study of d -semifaithful codes concerns the setting where the distortion is bounded by D almost surely [24]–[28]. The optimal expected length is $nR(D) + O(\log n)$ [26], [27]. Pointwise bounds on the length have been studied in [24], where the length is shown to be lower-bounded by $nR(D) + \sqrt{n}G_n + O(\log n)$ where G_n approaches a Gaussian random variable. Variable-length codes allowing a positive probability of excess distortion ϵ have been studied in [29], [30]. It was shown in [30] that the optimal expected length approaches $n(1 - \epsilon)R(D)$ from below.

The noisy lossy source coding problem (where the encoder only has a noisy observation Y^n of the source X^n), also referred to as remote lossy source coding, was first introduced by Dobrushin and Tsybakov [11]. It was shown to be asymptotically equivalent to the rate-distortion problem with a surrogate distortion between the output of the noisy channel and the decoder's output, an idea that was further explored in [20]. It reduces to the information bottleneck problem [6] under the logarithmic distortion measure [21]. For its strong converse exponent, see [31] and references therein. In [32], the problem was extended to incorporate an f -separable distortion measure. Refer to [33] for further extensions with privacy constraints.

Finite-blocklength achievability and converse bounds have been investigated in [10], where it was demonstrated that the dispersion function from [34] can be adapted to obtain nonasymptotic results for the noisy lossy source coding setting. A converse bound for variable-length noisy lossy source coding was derived in [35]. For additional nonasymptotic studies on this problem, see [36], [37] and references therein.

C. Channel Simulation and Poisson Functional Representation

Channel simulation aims to determine the minimal communication required over a noiseless channel to simulate a given channel $P_{Y|X}$. Various settings of channel simulation have been explored [38]–[40]. One-shot exact channel simulation with unlimited common randomness has been shown to require an average communication cost $I(X; Y) + O(\log I(X; Y))$, by using rejection sampling [41], [42], Poisson functional representation [12], or a combination of both [43]. Together with other related techniques, the Poisson functional representation [12] has been applied to various fields, including network information theory [9], [44], differential privacy [45], minimax learning [46] and neural compression [47]. Readers are referred to [48] for a comprehensive review on the channel simulation problem. Moreover, for the information bottleneck channel, a scheme similar to channel simulation has been employed in [5] to derive the achievable error exponents: by utilizing random coding and uniform index selection, the authors demonstrated that the relay's compression scheme can be asymptotically viewed as a discrete memoryless channel $P_{U|Y}$.

Notations: All random variables are discrete with finite support unless otherwise stated. Entropy is in bits and \log is to the base 2. Write $[n] := \{1, \dots, n\}$, $X^n = (X_1, \dots, X_n)$. Write $D(P\|Q)$ for the relative entropy, and $\iota_{X;Y}(x; y) = \log \frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)}$ for the information density. Write $\{0, 1\}^* := \cup_{k=0}^{\infty} \{0, 1\}^k$ for the set of bit sequences with any length. For a set \mathcal{X} , $|\mathcal{X}|$ denotes the cardinality. For a sequence $W \in \{0, 1\}^*$, $|W|$ denotes its length. $Q(t) := \mathbb{P}(Z \geq t)$ where $Z \sim N(0, 1)$, and $Q^{-1}(\epsilon)$ is its inverse function. “Almost surely” is often abbreviated as “a.s.”.

III. NOISY LOSSY SOURCE CODING

We first study noisy lossy source coding [11], which will be utilized in our analyses on the information bottleneck channel.

In the one-shot noisy lossy source coding setting, we have a pair of random variables $(X, Y) \sim P_{X,Y}$, where $X \in \mathcal{X}$ is the source, and $Y \in \mathcal{Y}$ is the observation. The encoder observes Y and produces a description $W = f(S, Y)$, where $S \sim P_S$ (independent of Y) is the encoder's local randomness.² The decoder observes W and recovers $Z = g(W) \in \mathcal{Z}$. The goal is to have a small probability of excess distortion $P_e := \mathbb{P}(d(X, Z) > D)$, where $d : \mathcal{X} \times \mathcal{Z} \rightarrow \mathbb{R}$ is the distortion measure, and $D \in \mathbb{R}$. The setting is explained in Figure 2 as follows.

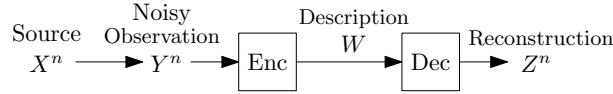


Fig. 2. Noisy Lossy Source Coding.

For the fixed-length setting where we require $W \in [L]$, this problem has been studied in [10] by Kostina and Verdú, which gives the following result.

Theorem 1 ([10]): For any $P_{\bar{Z}}$ and $\gamma > 0$, there exists a fixed-length code with

$$P_e \leq \mathbb{P}(\psi_{\bar{Z}}(Y, D, T) \geq \log \gamma) + e^{-L/\gamma},$$

where $T \sim \text{Unif}(0, 1)$ is independent of Y , and

$$\psi_{\bar{Z}}(y, D, t) := \inf_{P_{\bar{Z}}: \mathbb{P}(d(X, Z) > D | Z, Y=y) \leq t \text{ a.s.}} D(P_{\bar{Z}} \| P_{\bar{Z}}). \quad (6)$$

(We assume $Z \perp\!\!\!\perp (X, Y)$ above.)

In this paper, we discuss a variable-length setting where $W \in \mathcal{C}$ lies in a prefix-free codebook $\mathcal{C} \subseteq \{0, 1\}^*$, and we are interested in minimizing the expected length $\mathbb{E}[|W|]$. We show that the expected length can also be bounded in terms of $\psi_{\bar{Z}}$ in (6), using the technique in [10] and the strong functional representation lemma [12], [49].

Theorem 2: For any $P_{\bar{Z}}$, $\epsilon' > 0$, and function $\beta : \mathcal{Y} \rightarrow [0, 1]$, there is a variable-length code with $P_e \leq \mathbb{E}[\beta(Y)] + \epsilon'$ and

$$\mathbb{E}[|W|] \leq \ell \left(\mathbb{E}[(1 - \beta(Y))\psi_{\bar{Z}}(Y, D, \epsilon')] \right),$$

assuming the expectation above is finite,³ where $\ell(t) := t + \log(t + 2) + 4$.

Proof: Let $\phi(y, z, D) := \mathbb{P}(d(X, z) > D | Y = y)$. We use the Poisson functional representation [9], [12]. Let $0 < T_1 < T_2, \dots$ be a Poisson process and $\bar{Z}_1, \bar{Z}_2, \dots \stackrel{iid}{\sim} P_{\bar{Z}}$. Consider a channel $P_{\hat{Z}|Y}$, where conditional on $Y = y$, \hat{Z} has the same distribution as $\bar{Z} \sim P_{\bar{Z}}$ conditional on $\phi(y, \bar{Z}, D) \leq \epsilon'$.

$$K := \operatorname{argmin}_k \frac{T_k}{P_{\hat{Z}|Y}(\bar{Z}_k | Y) / P_{\bar{Z}}(\bar{Z}_k)}.$$

²The local randomness S is not useful for fixed-length settings, but can be useful in variable-length cases to randomize between two encoding functions.

³In particularity, we must have “ $\psi_{\bar{Z}}(Y, D, \epsilon') < \infty$ or $\beta(Y) = 1$ ” a.s..

By [49] and [48, Lemma 12], $\bar{Z}_K|\{Y = y\} \sim P_{\bar{Z}|Y}(\cdot|y)$ and

$$\begin{aligned} & \mathbb{E}[\log K|Y = y] \\ & \leq D(P_{\bar{Z}|Y}(\cdot|y)\|P_{\bar{Z}}) + 1 \\ & = -\log \mathbb{P}(\phi(y, \bar{Z}, D) \leq \epsilon') + 1 \\ & \leq \psi_{\bar{Z}}(y, D, \epsilon') + 1, \end{aligned}$$

where the last inequality is because if P_Z satisfies that $\phi(y, Z, D) \leq \epsilon$ almost surely, then

$$\begin{aligned} & D(P_Z\|P_{\bar{Z}}) \\ & \geq D(P_{\mathbf{1}\{\phi(y, Z, D) \leq \epsilon\}}\|P_{\mathbf{1}\{\phi(y, \bar{Z}, D) \leq \epsilon\}}) \\ & = -\log \mathbb{P}(\phi(y, \bar{Z}, D) \leq \epsilon) \end{aligned}$$

(this step appeared in [10]). Construct a randomized coding scheme as follows: the encoder observes Y , outputs $\tilde{K} = K$ with probability $1 - \beta(Y)$, or outputs $\tilde{K} = 1$ with probability $\beta(Y)$, and then encodes \tilde{K} using an optimal prefix-free code into W . The decoder decodes \tilde{K} from W and outputs $Z = \bar{Z}_{\tilde{K}}$. Since $\bar{Z}_K|\{Y = y\} \sim P_{\bar{Z}|Y}(\cdot|y)$,

$$\begin{aligned} & \mathbb{P}(d(X, Z) > D) \\ & \leq \mathbb{P}(\tilde{K} \neq K) + \mathbb{P}(d(X, \bar{Z}_K) > D) \\ & \leq \mathbb{E}[\beta(Y)] + \mathbb{E}[\phi(Y, \bar{Z}_K, D)] \\ & \leq \mathbb{E}[\beta(Y)] + \epsilon'. \end{aligned}$$

We have

$$\begin{aligned} & \mathbb{E}[\log \tilde{K}] \\ & = \mathbb{E}[(1 - \beta(Y))\mathbb{E}[\log K|Y]] \\ & \leq \mathbb{E}[(1 - \beta(Y))\psi_{\bar{Z}}(Y, D, \epsilon')] + 1. \end{aligned}$$

By the maximum entropy distribution for fixed $\mathbb{E}[\log \tilde{K}]$ [12], [48], we have

$$H(\tilde{K}) \leq \ell(\mathbb{E}[(1 - \beta(Y))\psi_{\bar{Z}}(Y, D, \epsilon')]) - 2.$$

We can then use Huffman code [50] to derive $\mathbb{E}[|W|] \leq H(\tilde{K}) + 1$.

The remaining problem is that these encoding and decoding functions depend on the common randomness $G := (\bar{Z}_i, T_i)_i$. To resolve this, we use the strategy in [12, Theorem 2] which incurs a 1-bit penalty on $\mathbb{E}[|W|]$. See Appendix A. ■

We also study the block setting where $X = X^n$, $Y = Y^n$, $Z = Z^n$ are sequences, $(X_i, Y_i) \stackrel{iid}{\sim} P_{X,Y}$, and $d(x^n, z^n) = n^{-1} \sum_{i=1}^n d(x_i, z_i)$. For the fixed-length setting, as $n \rightarrow \infty$, the optimal asymptotic description rate is [11]

$$R(D) := \min_{P_{Z|Y}: \mathbb{E}[d(X, Z)] \leq D} I(Y; Z).$$

The following refined bound for the fixed-length setting was shown in [10, Theorem 5].

Theorem 3 ([10]): Under the regularity conditions in [10],⁴ the smallest L such that there exists a fixed-length code with blocklength n and $P_e \leq \epsilon$ is

$$nR(D) + \sqrt{n\tilde{V}(D)Q^{-1}(\epsilon)} + O(\log n),$$

where $\tilde{V}(D) := \text{Var}[\iota_{Y;Z^*}(Y; Z^*) + \lambda^* d(X, Z^*)]$, $P_{Z^*|Y}$ attains the minimum in $R(D)$, and $\lambda^* := -R'(D)$.

In this paper, for the variable-length setting, we prove the following bound. The proof is in Appendix B.

Theorem 4: Under the regularity conditions in Theorem 3, for $\epsilon > 0$, if $n \geq n_0$ (where n_0 depends on $P_{X,Y}, d, D, \epsilon$), there exists a variable-length code with $P_e \leq \epsilon$,⁵ and

$$\mathbb{E}[|W|] \leq (1 - \epsilon) \left(nR(D) + \sqrt{(n \ln n) \widetilde{CV}(D)} \right) + O(\sqrt{n}),$$

where $\widetilde{CV}(D) := (\lambda^*)^2 \mathbb{E}[\text{Var}[d(X, Z^*) | Y, Z^*]]$, $P_{Z^*|Y}$ attains the minimum in $R(D)$,⁶ and $\lambda^* := -R'(D)$. The constant in $O(\sqrt{n})$ can depend on $P_{X,Y}, d, D, \epsilon$.

⁴We need $\min\{D' : R(D') < \infty\} < D < \min_z \mathbb{E}[d(X, z)]$, $R(D)$ is twice continuously differentiable as a function of P_Y (assuming $P_{X,Y} = P_{X|Y}P_Y$), and perturbing P_Y within a neighborhood of the original P_Y will not affect the support of Z^* , where $P_{Z^*|Y}$ attains the minimum in $R(D)$.

⁵This can be strengthened to $\mathbb{P}(d(X^n, Z^n) > D - n^{-1} \log n) \leq \epsilon - 1/\sqrt{n}$.

⁶If there are multiple $P_{Z^*|Y}$ attaining the minimum, choose the one that gives the smallest $\mathbb{E}[\text{Var}[d(X, Z^*) | Y, Z^*]]$.

Note that $\widetilde{\text{CV}}(\text{D}) \leq \tilde{\text{V}}(\text{D})$ since $\tilde{\text{V}}(\text{D})$ is the variance of $\iota(Y; Z^*) + \lambda^* d(X, Z^*)$, and $\widetilde{\text{CV}}(\text{D})$ is its conditional variance given Y, Z^* . We observe that the variable-length case in Theorem 4 exhibits a different behavior compared to the fixed-length case in Theorem 3. The asymptotic rate is $(1 - \epsilon)R(\text{D})$ instead of $R(\text{D})$, which is similar to the phenomenon observed in [29], [30] for lossless and lossy source coding with error. Intuitively, we can discard a portion ϵ of the sequences Y^n by assigning the same short codeword to them, which induces an error probability ϵ while reducing the expected length by $\approx \epsilon R(\text{D})$. Nevertheless, unlike the result for variable-length noiseless (i.e., $X = Y$) lossy source coding in [30] which shows that $\mathbb{E}[|W|] = (1 - \epsilon)nR(\text{D}) - \zeta\sqrt{n} + O(\log n)$ (the constant ζ is given in terms of ϵ and the rate-dispersion function) where the rate $\mathbb{E}[|W|]/n$ approaches $(1 - \epsilon)R(\text{D})$ from below, the noisy lossy source coding result in Theorem 4 gives a rate which approaches $(1 - \epsilon)R(\text{D})$ from *above*. The reason is that we have to take into account of the variance of $d(X^n, Z^n)$ which increases with n , whereas in the noiseless case $d(X^n, Z^n)$ is fixed by $X^n = Y^n$ and Z^n .

IV. INFORMATION BOTTLENECK CHANNEL

In this section, we construct schemes for both the fixed-length and the variable-length cases of the information bottleneck channel, by utilizing results on noisy lossy source coding.

We first define the one-shot information bottleneck channel. An encoder observes the message $M \sim \text{Unif}([L])$ and a shared random codebook $F \sim P_F$,⁷ and sends $X = f(F, M) \in \mathcal{X}$ (where \mathcal{X} is a finite set) through a channel $P_{Y|X}$ which outputs $Y \in \mathcal{Y}$ (where \mathcal{Y} is also finite). We require the codebook to be i.i.d., i.e., $f(F, 1), \dots, f(F, L)$ are i.i.d. following an input distribution P_X . An oblivious relay observes Y (but not F) and sends a description $W = f_r(S, Y)$ noiselessly to the decoder, where $S \sim P_S$ is the relay's local randomness. The decoder observes F and W , and then recovers $\hat{M} = g(F, W) \in [L]$. The goal is to minimize the error probability $P_e := \mathbb{P}(M \neq \hat{M})$. The setting has been presented in Figure 1.

About the description W , two settings will be studied:

- Fixed-length description: $W \in [K]$, and we want to minimize $K \in \mathbb{N}$;
- Variable-length description: $W \in \mathcal{C}_W$ is in a prefix-free codebook $\mathcal{C}_W \subseteq \{0, 1\}^*$, and we want to minimize the expected length $\mathbb{E}[|W|]$.

Moreover, we also study the block case, where the encoder sends a sequence $X^n \in \mathcal{X}$ that is passed through a memoryless channel $P_{Y|X}^n$ (n copies of $P_{Y|X}$), so the relay observes a sequence Y^n , i.e., we substitute $X = X^n, Y = Y^n$ and $P_{Y|X} = P_{Y|X}^n$ in the one-shot setting. Let $B_F^*(n, C, \epsilon)$ be the smallest possible relay description rate $n^{-1} \log K$ among fixed-length schemes with message rate $n^{-1} \log L \geq C$ and $P_e \leq \epsilon$. Define $B_V^*(n, C, \epsilon)$ for variable-length schemes similarly.

For the asymptotic setting where $n \rightarrow \infty$, the capacity for fixed-length description has been characterized in [1] as

$$\limsup_{\epsilon \rightarrow 0} \limsup_{n \rightarrow \infty} B_F^*(n, C, \epsilon) = \text{IB}_{X \rightarrow Y}(C).$$

[1] shows that the same asymptotic limit also holds for B_V^* . In the following subsections, using the techniques in noisy lossy source coding, we present nonasymptotic achievability results for the variable-length and fixed-length settings. Interestingly, the nonasymptotic results for them are rather different.

A. Variable-length Description

We start with the setting where the relay sends a variable-length description W in a prefix-free codebook. For the one-shot variable-length setting, one straightforward scheme is to consider the $P_{U|Y}$ that achieves the minimum in the information bottleneck (1), and have the relay perform a one-shot exact channel simulation scheme (e.g., [12], [41]–[43]) to simulate $P_{U|Y}$. Channel simulation can be performed with an expected description length bounded by $I(Y; U) + \log(I(Y; U) + 2) + 3$ [12], [49]. The decoder can then recover U , treat $P_{U|X}$ as a usual noisy channel, and perform decoding of the channel code. We explain the idea in Figure 3.

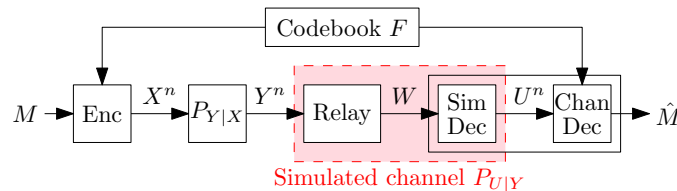


Fig. 3. Information bottleneck channel utilizing channel simulation.

The following result is a consequence of the channel simulation result in [12], [49] and the one-shot channel coding result in [9] (one may also use any of the bounds in [51]).

⁷ F is a random variable that represents a random choice out of the $|\mathcal{X}|^L$ different mappings from $[L]$ to \mathcal{X} .

Theorem 5: Fix any P_X , $P_{U|Y}$ and $\epsilon' \geq 0$. There is a one-shot variable-length scheme with

$$P_e \leq \mathbb{E} \left[1 - \left(1 - \min \{ 2^{-\iota_{X;U}(X;U)}, 1 \} \right)^{(L+1)/2} \right] + \epsilon', \quad (7)$$

and $\mathbb{E}[|W|] \leq \ell((1 - \epsilon')I(Y;U))$, where $\ell(t) := t + \log(t + 2) + 4$.

Proof: Applying the channel simulation result in [12], [49] to $P_{U|Y}$, we have an encoding function $K = f_a(S_a, Y)$ where $K \in \mathbb{N}$ (to be used by the relay) and a decoding function $U = g_a(S_a, K)$ (to be used by the decoder, where S_a is a common randomness) such that U follows the conditional distribution $P_{U|Y}$ given Y , and $\mathbb{E}[\log K] \leq I(Y;U) + 1$. Let $\tilde{K} = K$ with probability $1 - \epsilon'$, and $\tilde{K} = 1$ with probability ϵ' . As in the proof of Theorem 2, \tilde{K} can be encoded into W using a prefix-free code with

$$\begin{aligned} \mathbb{E}[|W|] &\leq \ell(\mathbb{E}[\log \tilde{K}] - 1) - 1 \\ &\leq \ell((1 - \epsilon')(\mathbb{E}[\log K] - 1)) - 1 \\ &\leq \ell((1 - \epsilon')I(Y;U)) - 1. \end{aligned}$$

Using the one-shot channel coding result in [9, Theorem 1] on $P_{U|X}$, we have an encoding function $X = f_b(F, M)$ (to be used by the encoder) and a decoding function $\hat{M} = g_b(F, U)$ (to be used by the decoder, where F is an i.i.d. random codebook) such that $\mathbb{P}(M \neq \hat{M}) \leq \mathbb{E}[1 - (1 - \min\{2^{-\iota_{X;U}(X;U)}, 1\})^{(L+1)/2}]$.

The scheme for the information bottleneck channel is as follows: the encoder observes M and outputs $X = f_b(F, M)$; the relay observes Y , computes $K = f_a(S_a, Y)$, generates \tilde{K} and encodes it to W ; the decoder observes W , recovers \tilde{K} , $U = g_a(S_a, \tilde{K})$, $\hat{M} = g_b(F, U)$. Since $\mathbb{P}(K \neq \tilde{K}) \leq \epsilon'$, using \tilde{K} instead of K increases P_e by at most ϵ' .

The remaining problem is that there is a common randomness S_a shared between the relay and the decoder. It can be removed using the same technique as in the proof of Theorem 2, incurring a 1-bit penalty on $\mathbb{E}[|W|]$. ■

A direct application of Theorem 5 to the asymptotic setting $X = X^n$, $Y = Y^n$ yields the asymptotic result $\limsup_{n \rightarrow \infty} B_V^*(n, C, \epsilon) \leq (1 - \epsilon)IB_{X \rightarrow Y}(C)$. To refine the bound, we utilize the noisy lossy source coding result in Theorem 2. Intuitively, for any fixed $p_{U|Y}$ (e.g., the one achieving the minimum in $IB_{X \rightarrow Y}(C) = \min_{P_{U|Y}: I(X;U) \geq C} I(Y;U)$), the relay performs a noisy lossy source coding on Y to allow the decoder to recover \hat{U} , with a distortion function $d(x, \hat{u}) = -\iota_{X;U}(x; \hat{u})$. As long as the distortion is small enough, that is, $\iota_{X;U}(x; \hat{u}) \gg \log L$, the decoder can decode X using \hat{U} via the Poisson matching lemma [9].

Theorem 6: Fix any P_X , $P_{U|Y}$, $C, \epsilon' > 0$, and function $\beta : \mathcal{Y} \rightarrow [0, 1]$. There is a one-shot variable-length scheme with message size L ,

$$\begin{aligned} P_e &\leq \mathbb{E}[\beta(Y)] + 2^{-C}(L + 1)/2 + \epsilon', \\ \mathbb{E}[|W|] &\leq \ell(\mathbb{E}[(1 - \beta(Y))\psi_U(Y, C, \epsilon')]), \end{aligned}$$

assuming the expectation above is finite, where $\ell(t) := t + \log(t + 2) + 4$, and

$$\psi_U(y, C, t) := \inf_{P_{\tilde{U}}: \mathbb{P}(\iota_{X;U}(X; \tilde{U}) < C | \tilde{U}, Y=y) \leq t \text{ a.s.}} D(P_{\tilde{U}} \| P_U). \quad (8)$$

(We assume $\tilde{U} \perp\!\!\!\perp (X, Y)$ above.)

Proof: Fix any P_X , $P_{U|Y}$. Applying Theorem 2 on the distortion function $d(x, \hat{u}) = -\iota_{X;U}(x; \hat{u})$ and distortion level $D = -C$, we have a code for noisy lossy source coding, with encoding function $W = f_r(S, Y)$ (used by the relay) and decoding function $\hat{U} = g_r(W)$ (used by the decoder) so that

$$\mathbb{P}(d(X; \hat{U}) > D) \quad (9)$$

$$\begin{aligned} &= \mathbb{P}(\iota_{X;U}(X; \hat{U}) < C) \\ &\leq \mathbb{E}[\beta(Y)] + \epsilon', \end{aligned} \quad (10)$$

and $\mathbb{E}[|W|] \leq \ell(\mathbb{E}[(1 - \beta(Y))\psi_U(Y, C, \epsilon')])$. It is left to design the encoder and the decoder.

We utilize the Poisson functional representation [9], [12]. Let $\bar{X}_1, \bar{X}_2, \dots \stackrel{iid}{\sim} P_X$, and $0 < T_1 < T_2, \dots$ be a Poisson process. The encoder observes $M \in [L]$ and sends $X = \bar{X}_M$. The decoder observes $\hat{U} = g_r(W)$ and recovers

$$\hat{M} := \operatorname{argmin}_{k \in [L]} \frac{T_k}{P_{X|U}(\bar{X}_k | \hat{U}) / P_X(\bar{X}_k)}.$$

By the generalized Poisson matching lemma [9, Lemma 3],

$$\begin{aligned}
& \mathbb{P}(M \neq \hat{M} | M = m) \\
&= \mathbb{E} \left[\min \left\{ m \frac{P_X(X)}{P_{X|U}(X|\hat{U})}, 1 \right\} \right] \\
&\leq \mathbb{E} \left[\min \left\{ m 2^{-\iota_{X;U}(X;\hat{U})}, 1 \right\} \right] \\
&\leq 2^{-C} m + \mathbb{E}[\beta(Y)] + \epsilon',
\end{aligned}$$

where the last inequality is by (10). The result follows from averaging over $M \sim \text{Unif}([L])$. We take the codebook random variable to be $F = (\tilde{X}_m)_{m \in [L]}$.⁸ ■

We then show an achievability result for the block setting in terms of the information bottleneck $\text{IB}(\mathbf{C})$ in (1) and the conditional var-information bottleneck $\text{CVIB}(\mathbf{C})$ in (4), using the noisy lossy source coding result in Theorem 4.

Theorem 7: Fix any P_X , $\epsilon > 0$ and $0 < \mathbf{C} < I(X; Y)$. Under the regularity conditions in the footnote,⁹ we have

$$B_V^*(n, \mathbf{C}, \epsilon) \leq (1 - \epsilon) \left(\text{IB}(\mathbf{C}) + \sqrt{\frac{\ln n}{n} \text{CVIB}(\mathbf{C})} \right) + O\left(\frac{1}{\sqrt{n}}\right),$$

where $\text{IB}(\mathbf{C}) = \text{IB}_{X \rightarrow Y}(\mathbf{C})$ and $\text{CVIB}(\mathbf{C}) = \text{CVIB}_{X \rightarrow Y}(\mathbf{C})$.

Proof: Consider the $P_{U|Y}$ that achieves the minimum in $\text{IB}(\mathbf{C})$ (for tie-breaking, choose the $P_{U|Y}$ that gives the smallest $\mathbb{E}[\text{Var}[\iota_{X;U}(X, U) | Y, U]]$). Define a distortion function $d(x, u) = -\iota_{X;U}(x; u)$, and consider the rate-distortion function $R(\mathbf{D}) = \min_{P_{\tilde{U}|Y}: \mathbb{E}[d(X, \tilde{U})] \leq \mathbf{D}} I(Y; \tilde{U})$ of the noisy lossy source coding problem at $\mathbf{D} = -\mathbf{C}$. Since

$$\begin{aligned}
& I(X; \tilde{U}) - \mathbb{E}[\iota_{X;U}(X, \tilde{U})] \\
&= \mathbb{E} \left[\log \frac{P_{X|\tilde{U}}(X|\tilde{U})}{P_{X|U}(X|\tilde{U})} \right] \\
&= \mathbb{E} \left[D(P_{X|\tilde{U}}(\cdot|\tilde{U}) \| P_{X|U}(\cdot|\tilde{U})) \right] \geq 0,
\end{aligned} \tag{11}$$

$\mathbb{E}[d(X, \tilde{U})] \leq \mathbf{D}$ implies $I(X; \tilde{U}) \geq \mathbf{C}$, and hence $P_{U|Y}$ also achieves the minimum in $R(\mathbf{D})$, and $R(\mathbf{D}) = \text{IB}(\mathbf{C})$. (11) also implies that $R(\mathbf{D} - \delta) \geq \text{IB}(\mathbf{C} + \delta)$ for every $\delta \in \mathbb{R}$, so we must have $-R'(\mathbf{D}) = \text{IB}'(\mathbf{C}) =: \lambda^*$. Theorem 4 gives a noisy lossy coding scheme with

$$\begin{aligned}
& \mathbb{E}[|W|] \\
&\leq (1 - \epsilon) \left(n I(Y; U) \right. \\
&\quad \left. + \lambda^* \sqrt{(n \ln n) \mathbb{E}[\text{Var}[d(X, U) | Y, U]]} \right) + O(\sqrt{n}) \\
&\leq (1 - \epsilon) \left(n \text{IB}(\mathbf{C}) + \sqrt{(n \ln n) \text{CVIB}(\mathbf{C})} \right) + O(\sqrt{n}),
\end{aligned}$$

with a decoded sequence \hat{U}^n satisfying (see footnote 5)

$$\mathbb{P}(\iota_{X^n; U^n}(X^n; \hat{U}^n) < n\mathbf{C} + \log n) \leq \epsilon - 1/\sqrt{n}. \tag{12}$$

We apply the Poisson matching lemma [9] as in the proof of Theorem 6 on X^n, \hat{U}^n and $L = \lceil 2^{n\mathbf{C}} \rceil$, which for $n \geq 4$ gives

$$\begin{aligned}
& \mathbb{P}(M \neq \hat{M} | M = m) \\
&\leq 2^{-(n\mathbf{C} + \log n)} m + \epsilon - 1/\sqrt{n} \\
&\leq 2^{-(n\mathbf{C} + \log n)} (2^{n\mathbf{C}} + 1) + \epsilon - 1/\sqrt{n} \\
&\leq 2/n + \epsilon - 1/\sqrt{n} \leq \epsilon.
\end{aligned}$$

■

⁸ $(T_m)_m$ is only a local randomness at the decoder. If this is not allowed, we can fix a particular $(t_m)_m$ that satisfies the bound on P_e .

⁹We need $\hat{R}(\mathbf{C}) := \min_{P_{\tilde{U}|Y}: \mathbb{E}[\iota_{X;U}(X, \tilde{U})] \geq \mathbf{C}} I(Y; \tilde{U})$ to be twice continuously differentiable as a function of P_Y (assuming $P_{X,Y} = P_{X|Y}P_Y$, and let $P_{U|Y}$ be the minimizer in $\text{IB}(\mathbf{C})$), and perturbing P_Y within a neighborhood of the original P_Y will not affect the support of U^* , where $P_{U^*|Y}$ attains the minimum in $\hat{R}(\mathbf{C})$.

B. Fixed-length Description

We now consider the case where $W \in [K]$ is a fixed-length description. The following achievability result is a corollary of Theorem 1 and the Poisson matching lemma [9]. The proof is the same as that of Theorem 6 (except we use the fixed-length result in Theorem 1 instead of the variable-length result in Theorem 2), and is omitted.

Theorem 8: Fix P_X , $P_{U|Y}$ and $C, \gamma > 0$. There is a one-shot fixed-length scheme with message size L , description size K ,

$$P_e \leq \mathbb{P}(\psi_U(Y, C, T) \geq \log \gamma) + 2^{-C}(L+1)/2 + e^{-K/\gamma},$$

where $T \sim \text{Unif}(0, 1)$, $T \perp\!\!\!\perp Y$, with ψ_U defined in (8).

We then give a second-order result in terms of the var-information bottleneck $\text{VIB}(C)$ in (2).

Theorem 9: Fix any P_X , $\epsilon > 0$ and $0 < C < I(X; Y)$. Under the regularity conditions in Theorem 7, we have

$$\mathbf{B}_F^*(n, C, \epsilon) \leq \text{IB}(C) + \sqrt{\frac{1}{n} \text{VIB}(C) Q^{-1}(\epsilon)} + O\left(\frac{\log n}{n}\right),$$

where $\text{IB}(C) = \text{IB}_{X \rightarrow Y}(C)$ and $\text{VIB}(C) = \text{VIB}_{X \rightarrow Y}(C)$.

Proof: As in Theorem 7, consider $P_{U|Y}$ that achieves the minimum in $\text{IB}(C)$. Let $d(x, u) = -\iota_{X;U}(x; u)$ and $R(D) = \min_{P_{\tilde{U}|Y}: \mathbb{E}[d(X, \tilde{U})] \leq D} I(Y; \tilde{U})$. We have shown that $R(D) = \text{IB}(C)$ and $-R'(D) = \text{IB}'(C) =: \lambda^*$. Theorem 3 gives a noisy lossy coding scheme with decoded sequence \hat{U}^n satisfying

$$\begin{aligned} \log K &\leq n\text{IB}(C) + \sqrt{n\text{Var}[\iota_{Y;U}(Y; U) - \lambda^* \iota_{X;U}(X; U)]} \\ &\quad \cdot Q^{-1}(\epsilon) + O(\log n) \\ &= n\text{IB}(C) + \sqrt{n\text{VIB}(C)} Q^{-1}(\epsilon) + O(\log n). \end{aligned}$$

Inspecting [10, Appendix D] shows that the bound $\mathbb{P}(d(X^n, \hat{U}^n) > D) \leq \epsilon$ in Theorem 3 can be strengthened to $\mathbb{P}(d(X^n, \hat{U}^n) > D - n^{-1} \log n) \leq \epsilon - 1/\sqrt{n}$, giving the same bound as (12). The proof is completed by applying the Poisson matching lemma [9] as in Theorem 7. ■

V. CONCLUDING REMARKS

We have shown nonasymptotic achievability results for the information bottleneck channel with fixed and variable-length descriptions, using techniques in noisy lossy source coding and Poisson functional representation. We have also shown novel bounds for variable-length noisy lossy source coding. For future directions, it is of interest to study converse results and investigate whether Theorems 4, 7 and 9 are tight.

VI. ACKNOWLEDGEMENT

This work was partially supported by two grants from the Research Grants Council of the Hong Kong Special Administrative Region, China [Project No.s: CUHK 24205621 (ECS), CUHK 14209823 (GRF)]. We would like to thank Prof. Ioannis Kontoyiannis for the insightful discussion. We would also like to thank the anonymous reviewers for their constructive and helpful comments on the earlier conference version of the manuscript [8].

REFERENCES

- [1] A. Sanderovich, S. Shamai, Y. Steinberg, and G. Kramer, "Communication via decentralized processing," *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 3008–3023, 2008.
- [2] O. Simeone, E. Erkip, and S. Shamai, "On codebook information for interference relay channels with out-of-band relaying," *IEEE Transactions on Information Theory*, vol. 57, no. 5, pp. 2880–2888, 2011.
- [3] I. E. Aguerri, A. Zaidi, G. Caire, and S. Shamai, "On the capacity of cloud radio access networks with oblivious relaying," *IEEE Transactions on Information Theory*, vol. 65, no. 7, pp. 4575–4596, 2019.
- [4] M. Dikshstein, N. Weinberger, and S. Shamai, "On mismatched oblivious relaying," in *2023 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2023, pp. 1687–1692.
- [5] H. Wu and H. Joudé, "An achievable error exponent for the information bottleneck channel," in *2024 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2024, pp. 1297–1302.
- [6] N. Tishby, F. Pereira, and W. Bialek, "The information bottleneck method," in *Proceedings of the 37th Annual Allerton Conference on Communication, Control, and Computing*. IEEE, 1999, pp. 368–377.
- [7] N. Slonim, "The information bottleneck: Theory and applications," Ph.D. dissertation, Hebrew University, 2002.
- [8] Y. Liu, S. H. Advary, and C. T. Li, "Nonasymptotic oblivious relaying and variable-length noisy lossy source coding," in *2025 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2025.
- [9] C. T. Li and V. Anantharam, "A unified framework for one-shot achievability via the Poisson matching lemma," *IEEE Transactions on Information Theory*, vol. 67, no. 5, pp. 2624–2651, 2021.
- [10] V. Kostina and S. Verdú, "Nonasymptotic noisy lossy source coding," *IEEE Transactions on Information Theory*, vol. 62, no. 11, pp. 6111–6123, 2016.
- [11] R. Dobrushin and B. Tsybakov, "Information transmission with additional noise," *IRE Transactions on Information Theory*, vol. 8, no. 5, pp. 293–304, 1962.

- [12] C. T. Li and A. El Gamal, "Strong functional representation lemma and applications to coding theorems," *IEEE Transactions on Information Theory*, vol. 64, no. 11, pp. 6967–6978, Nov 2018.
- [13] M. Peng, C. Wang, V. Lau, and H. V. Poor, "Fronthaul-constrained cloud radio access networks: Insights and challenges," *IEEE Wireless Communications*, vol. 22, no. 2, pp. 152–160, 2015.
- [14] M. Ensan, H. Joudeh, A. Alvarado, U. Gustavsson, and F. M. Willems, "On cloud radio access networks with cascade oblivious relaying," in *2021 IEEE Information Theory Workshop (ITW)*. IEEE, 2021, pp. 1–6.
- [15] H. Wu and H. Joudeh, "Error exponents for oblivious relaying and connections to source coding with a helper," *arXiv preprint arXiv:2505.00567*, 2025.
- [16] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.
- [17] N. Tishby, F. C. Pereira, and W. Bialek, "The information bottleneck method," *arXiv preprint physics/0004057*, 2000.
- [18] A. Zaidi, I. Estella-Aguerrí, and S. Shamai, "On the information bottleneck problems: Models, connections, applications and information theoretic views," *Entropy*, vol. 22, no. 2, p. 151, 2020.
- [19] Z. Goldfeld and Y. Polyanskiy, "The information bottleneck problem and its applications in machine learning," *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 1, pp. 19–38, 2020.
- [20] H. Witsenhausen, "Indirect rate distortion problems," *IEEE Transactions on Information Theory*, vol. 26, no. 5, pp. 518–521, 1980.
- [21] T. A. Courtade and T. Weissman, "Multiterminal source coding under logarithmic loss," *IEEE Transactions on Information Theory*, vol. 60, no. 1, pp. 740–761, 2013.
- [22] A. Ingber and Y. Kochman, "The dispersion of lossy source coding," in *2011 Data Compression Conference*, March 2011, pp. 53–62.
- [23] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3309–3338, June 2012.
- [24] I. Kontoyiannis, "Pointwise redundancy in lossy data compression and universal lossy data compression," *IEEE Transactions on Information Theory*, vol. 46, no. 1, pp. 136–152, 2000.
- [25] D. S. Ornstein and P. C. Shields, "Universal almost sure data compression," *The Annals of Probability*, pp. 441–452, 1990.
- [26] B. Yu and T. P. Speed, "A rate of convergence result for a universal d-semifair code," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 813–820, 1993.
- [27] Z. Zhang, E.-H. Yang, and V. K. Wei, "The redundancy of source coding with a fidelity criterion. 1. known statistics," *IEEE Transactions on Information Theory*, vol. 43, no. 1, pp. 71–91, 1997.
- [28] I. Kontoyiannis and J. Zhang, "Arbitrary source models and bayesian codebooks in rate-distortion theory," *IEEE Transactions on Information Theory*, vol. 48, no. 8, pp. 2276–2290, 2002.
- [29] H. Koga and H. Yamamoto, "Asymptotic properties on codeword lengths of an optimal FV code for general sources," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1546–1555, 2005.
- [30] V. Kostina, Y. Polyanskiy, and S. Verdú, "Variable-length compression allowing errors," *IEEE Transactions on Information Theory*, vol. 61, no. 8, pp. 4316–4330, 2015.
- [31] H. Wu and H. Joudeh, "Strong converse exponent for remote lossy source coding," *arXiv preprint arXiv:2501.14620*, 2025.
- [32] P. A. Stavrou, Y. Shkel, and M. Kountouris, "Indirect rate distortion functions with f-separable distortion criterion," in *2023 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2023, pp. 1050–1055.
- [33] K. Kittichokechai and G. Caire, "Privacy-constrained remote source coding," in *2016 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2016, pp. 1078–1082.
- [34] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3309–3338, 2012.
- [35] S. Saito, H. Yagi, and T. Matsushima, "New results on variable-length lossy compression allowing positive overflow and excess distortion probabilities," in *2018 International Symposium on Information Theory and Its Applications (ISITA)*. IEEE, 2018, pp. 359–363.
- [36] H. Yang, Y. Shi, S. Shao, and X. Yuan, "Indirect lossy source coding with observed source reconstruction: Nonasymptotic bounds and second-order asymptotics," *IEEE Transactions on Communications*, pp. 1–1, 2024.
- [37] L. Zhou, M. Motani *et al.*, "Finite blocklength lossy source coding for discrete memoryless sources," *Foundations and Trends® in Communications and Information Theory*, vol. 20, no. 3, pp. 157–389, 2023.
- [38] C. H. Bennett, P. W. Shor, J. Smolin, and A. V. Thapliyal, "Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem," *IEEE Transactions on Information Theory*, vol. 48, no. 10, pp. 2637–2655, 2002.
- [39] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, and A. Winter, "The quantum reverse Shannon theorem and resource tradeoffs for simulating quantum channels," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2926–2959, May 2014.
- [40] P. Cuff, "Distributed channel synthesis," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7071–7096, Nov 2013.
- [41] P. Harsha, R. Jain, D. McAllester, and J. Radhakrishnan, "The communication complexity of correlation," *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 438–449, Jan 2010.
- [42] G. Flamich, "Greedy Poisson rejection sampling," *Advances in Neural Information Processing Systems*, vol. 36, 2023.
- [43] L. Theis and N. Yosri, "Algorithms for the communication of samples," in *International Conference on Machine Learning*. PMLR, 2022, pp. 21 308–21 328.
- [44] Y. Liu and C. T. Li, "One-shot coding over general noisy networks," in *2024 IEEE International Symposium on Information Theory (ISIT)*, 2024, pp. 3124–3129.
- [45] Y. Liu, W.-N. Chen, A. Özgür, and C. T. Li, "Universal exact compression of differentially private mechanisms," *Advances in Neural Information Processing Systems*, 2024.
- [46] C. T. Li, X. Wu, A. Özgür, and A. El Gamal, "Minimax learning for distributed inference," *IEEE Transactions on Information Theory*, vol. 66, no. 12, pp. 7929–7938, 2020.
- [47] E. Lei, H. Hassani, and S. S. Bidokhti, "Neural estimation of the rate-distortion function with applications to operational source coding," *IEEE Journal on Selected Areas in Information Theory*, vol. 3, no. 4, pp. 674–686, 2022.
- [48] C. T. Li, "Channel simulation: Theory and applications to lossy compression and differential privacy," *Foundations and Trends® in Communications and Information Theory*, vol. 21, no. 6, pp. 847–1106, 2024. [Online]. Available: <http://dx.doi.org/10.1561/0100000141>
- [49] —, "Pointwise redundancy in one-shot lossy compression via Poisson functional representation," in *arXiv preprint; short version presented at 2024 International Zurich Seminar on Information and Communication*, 2024, pp. 28–29. [Online]. Available: <https://arxiv.org/pdf/2401.14805.pdf>
- [50] D. A. Huffman, "A method for the construction of minimum-redundancy codes," *Proceedings of the IRE*, vol. 40, no. 9, pp. 1098–1101, 1952.
- [51] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

APPENDIX

A. Remainder of the Proof of Theorem 2

We use the strategy in [12, Theorem 2] to remove the common randomness $G := (\bar{Z}_i, T_i)_i$. Consider two values g_0, g_1 of the common randomness and $\lambda_0 \in [0, 1]$ (let $\lambda_1 = 1 - \lambda_0$) such that

$$\begin{aligned} \sum_{i=0}^1 \lambda_i \mathbb{E}[|W| \mid G = g_i] &\leq \mathbb{E}[|W|], \\ \sum_{i=0}^1 \lambda_i \mathbb{P}(d(X, Z) > D \mid G = g_i) &\leq \mathbb{P}(d(X, Z) > D). \end{aligned}$$

This is possible by Carathéodory's theorem. The encoder generates $J \sim \text{Bern}(\lambda_1)$ and transmits it to the decoder, and then performs the aforementioned encoding scheme conditional on $G = g_J$. The decoder receives J, W and performs the decoding scheme conditional on $G = g_J$. Since J takes one bit to transmit, the resultant expected length is

$$\begin{aligned} &\mathbb{E}[|W|] + 1 \\ &\leq H(\tilde{K}) + 2 \\ &\leq \ell(\mathbb{E}[(1 - \beta(Y))\psi_{\bar{Z}}(y, D, \epsilon')]). \end{aligned}$$

B. Proof of Theorem 4

Define $\phi(y^n, z^n, D) := \mathbb{P}(d(X^n, z^n) > D \mid Y^n = y^n)$. Fix $\epsilon_1 > 0$. It was shown in [10, Appendix D] that for all typical $y^n \in \mathcal{T}_n$ (where $\mathcal{T}_n := \{y^n : \|\hat{P}_{y^n} - P_Y\|^2 \leq |\mathcal{Y}|n^{-1} \log n\}$, and \hat{P}_{y^n} is the empirical distribution of y^n),

$$\begin{aligned} &\psi_{Z^{*n}}(y^n, D, \epsilon_1) \\ &= \inf_{P_{Z^n} : \phi(y^n, Z^n, D) \leq \epsilon_1 \text{ a.s.}} D(P_{Z^n} \| P_{Z^*}^n) \\ &\leq \sum_{i=1}^n j(y_i, D) + \lambda^* \sqrt{n V_d} Q^{-1}(\epsilon_1) + O(\log n), \end{aligned} \tag{13}$$

where $j(y, D) := \iota_{Y; Z^*}(y; z) + \lambda^*(\mathbb{E}[d(X, z) \mid Y = y] - D)$ (this holds for P_{Z^*} -almost all z) is the d -tilted information for the corresponding noiseless source coding problem, and $V_d := \mathbb{E}[\text{Var}[d(X, Z^*) \mid Y, Z^*]]$. For footnote 5, note that the arguments in [10, Appendix D] show that (13) continues to hold if $\psi_{Z^{*n}}(y^n, D, \epsilon_1)$ in the left hand side is replaced with $\psi_{Z^{*n}}(y^n, D - n^{-1} \log n, \epsilon_1)$, which does not affect the use of the Berry-Esseen theorem. We have [10]

$$\begin{aligned} &\mathbb{P}(Y^n \notin \mathcal{T}_n) \\ &\leq 2|\mathcal{Y}|/\sqrt{n} =: \epsilon_2 - 1/\sqrt{n}, \end{aligned}$$

where $\epsilon_2 := (2|\mathcal{Y}| + 1)/\sqrt{n}$. Take $\beta(y^n) = 1$ if $y^n \notin \mathcal{T}_n$, and $\beta(y^n) = \epsilon_3$ if $y^n \in \mathcal{T}_n$. By (13), letting $\epsilon = \epsilon_1 + \epsilon_2 + \epsilon_3$ and $R := R(D) = \mathbb{E}[j(Y, D)]$,

$$\begin{aligned} &\mathbb{E}[\psi_{Z^{*n}}(Y^n, \epsilon_1)(1 - \beta(Y^n))] \\ &\leq (1 - \epsilon_3) \left(\mathbb{E} \left[\sum_{i=1}^n j(Y_i, D) \right] + \lambda^* \sqrt{n V_d} Q^{-1}(\epsilon_1) \right) + O(\log n) \\ &= (1 - \epsilon_3) \left(nR + \lambda^* \sqrt{n V_d} Q^{-1}(\epsilon_1) \right) + O(\log n). \end{aligned} \tag{14}$$

Take $\epsilon_1 = 1/(2\sqrt{n})$. We have

$$\begin{aligned} (1 - \epsilon_3)nR &= (1 - \epsilon)nR + \epsilon_1 nR + \epsilon_2 nR \\ &= (1 - \epsilon)nR + (R/2)\sqrt{n} + (2|\mathcal{Y}| + 1)R\sqrt{n} \\ &= (1 - \epsilon)nR + O(\sqrt{n}). \end{aligned} \tag{15}$$

By Chernoff bound $Q(t) \leq e^{-t^2/2}$, we have

$$\begin{aligned} \lambda^* \sqrt{n V_d} Q^{-1}(\epsilon_1) &\leq \lambda^* \sqrt{2n V_d \ln(1/\epsilon_1)} \\ &= \lambda^* \sqrt{V_d n \ln n} + O\left(\sqrt{\frac{n}{\ln n}}\right), \end{aligned}$$

and

$$\begin{aligned}
& (1 - \epsilon_3)\lambda^* \sqrt{nV_d} Q^{-1}(\epsilon_1) \\
& \leq (1 - \epsilon_3)\lambda^* \sqrt{V_d n \ln n} + O\left(\sqrt{\frac{n}{\ln n}}\right) \\
& = (1 - \epsilon)\lambda^* \sqrt{V_d n \ln n} + O\left(\sqrt{\frac{n}{\ln n}}\right).
\end{aligned} \tag{16}$$

Substituting (15) and (16) into (14),

$$\begin{aligned}
& \mathbb{E}[\psi_{Z^{*n}}(Y^n, D, \epsilon_1)(1 - \beta(Y^n))] \\
& \leq (1 - \epsilon) \left(nR + \lambda^* \sqrt{V_d n \ln n} \right) + O(\sqrt{n}).
\end{aligned}$$

Invoking Theorem 2, there exists a variable-length code with

$$\begin{aligned}
\mathbb{E}[|W|] & \leq \ell(\mathbb{E}[\psi_{Z^{*n}}(Y^n, D, \epsilon_1)(1 - \beta(Y^n))]) \\
& \leq (1 - \epsilon) \left(nR + \lambda^* \sqrt{V_d n \ln n} \right) + O(\sqrt{n}) + O(\log n) \\
& = (1 - \epsilon) \left(nR + \lambda^* \sqrt{V_d n \ln n} \right) + O(\sqrt{n}),
\end{aligned}$$

and

$$\begin{aligned}
P_e & \leq \mathbb{E}[\beta(Y^n)] + \epsilon_1 \\
& \leq \mathbb{P}(Y^n \notin \mathcal{T}_n) + \epsilon_3 + \epsilon_1 \\
& \leq \epsilon_2 - 1/\sqrt{n} + \epsilon_3 + \epsilon_1 \\
& = \epsilon - 1/\sqrt{n}.
\end{aligned}$$