A Unified Framework for One-shot Achievability via the Poisson Matching Lemma

Cheuk Ting Li and Venkat Anantharam
EECS, UC Berkeley, Berkeley, CA, USA
Email: ctli@berkeley.edu, ananth@eecs.berkeley.edu

Abstract

We introduce a fundamental lemma called the Poisson matching lemma, and apply it to prove one-shot achievability results for various settings, namely channels with state information at the encoder, lossy source coding with side information at the decoder, joint source-channel coding, broadcast channels, distributed lossy source coding, multiple access channels, channel resolvability and wiretap channels. Our one-shot bounds improve upon the best known one-shot bounds in most of the aforementioned settings (except multiple access channels, channel resolvability and wiretap channels, where we recover bounds comparable to the best known bounds), with shorter proofs in some settings even when compared to the conventional asymptotic approach using typicality. The Poisson matching lemma replaces both the packing and covering lemmas, greatly simplifying the error analysis. This paper extends the work of Li and El Gamal on Poisson functional representation, which mainly considered variable-length source coding settings, whereas this paper studies fixed-length settings, and is not limited to source coding, showing that the Poisson functional representation is a viable alternative to typicality for most problems in network information theory.

I. Introduction

The Poisson functional representation was introduced by Li and El Gamal [1] to prove the strong functional representation lemma: for any pair of random variables (X, Y), there exists a random variable Z independent of X such that Y is a function of (X, Z), and $H(Y|Z) \le I(X;Y) + \log(I(X;Y)+1) + 4$. The lemma is applied to show various one-shot variable-length lossy source coding results, and a simple proof of the asymptotic achievability in the Gelfand-Pinsker theorem [2].

In this paper, we introduce the Poisson matching lemma, which gives a bound on the probability of mismatch between the Poisson functional representations applied on different distributions, and use it to prove one-shot achievability results for various settings, namely channels with state information at the encoder, lossy source coding with side information at the decoder, joint source-channel coding, broadcast channels, distributed lossy source coding, multiple access channels, channel resolvability and wiretap channels. The Poisson matching lemma can replace both the packing and covering lemmas (and generalizations such as the mutual covering lemma) in asymptotic typicality-based proofs. The one-shot bounds in this paper subsume the corresponding asymptotic achievability results by straightforward applications of the law of large numbers.

Various non-asymptotic alternatives to typicality have been proposed, e.g. one-shot packing and covering lemmas [3], [4], stochastic likelihood coder [5], likelihood encoder [6] and random binning [7]. However, these non-asymptotic approaches generally require more complex proofs than their asymptotic counterparts, whereas proofs using the Poisson matching lemma can be even simpler than asymptotic proofs.

Our approach is better than the conventional asymptotic approach using typicality (and previous one-shot results, e.g. [3], [5]), in the following ways:

1) We can give one-shot bounds stronger than the best known one-shot bounds in many settings discussed in this paper, with the exception of channel coding, multiple access channels, channel resolvability and wiretap channels, which are included for demonstration purposes, where we recover bounds comparable to the best known bounds.
2) Our proofs work for random variables in general Polish spaces.
3) To the best of our knowledge, for the achievability in the Gelfand-Pinsker theorem [2] (for channels with state information at the encoder) and the Wyner-Ziv theorem [8], [9] (for lossy source coding with side information at the decoder), our proofs are significantly shorter than all previous proofs (another short proof of the achievability in the Gelfand-Pinsker theorem is given in [1], though it is asymptotic). Using our approach, we can also greatly shorten the proof of the achievability of the dispersion in joint source-channel coding [10].
4) Our proofs only use the Poisson matching lemma introduced in this paper, which replaces both the packing and covering lemmas in proofs using typicality. The Poisson matching lemma can also be used to prove a soft covering lemma. Hence the Poisson matching lemma can be the only tool needed to prove a wide range of results in network information theory.
5) Our analyses usually involve fewer (or no) uses of sub-codebooks and binning. As a result, we can reduce the number of error events and give sharper second-order bounds. For example:
   a) Conventional proofs of the Gelfand-Pinsker theorem involve one sub-codebook, giving an additional error event, whereas we do not use any sub-codebook.
   b) Conventional proofs of the Wyner-Ziv theorem and the Berger-Tung inner bound [11], [12] (for distributed lossy source coding) use binning, giving additional error events, whereas we do not require binning.

c) Conventional proofs of Marton's inner bound [13] (for broadcast channels) involve two sub-codebooks, whereas we use only one.

6) In our approach, the encoders and decoders are characterized using a common framework (the Poisson functional representation), which is noteworthy since the roles of an encoder and a decoder in an operational setting are very different, and their constructions usually have little in common in conventional approaches.

Notation

Throughout this paper, we assume that $\log$ is to base 2 and the entropy H is in bits. We write $\exp_a(b)$ for a^b.

The set of positive integers is denoted as $\mathbb{N} = \{1, 2, \ldots\}$. We use the notation: $X_a^b := (X_a, \ldots, X_b)$, $X^n := X_1^n$ and $[a:b] := [a, b] \cap \mathbb{Z}$. The conditional information density is denoted as

$$\iota_{X;Y|Z}(x; y|z) := \log \frac{dP_{XY|Z=z}}{d(P_{X|Z=z} \times P_{Y|Z=z})}(x, y).$$

We consider $\iota_{X;Y|Z}(x; y|z)$ to be defined only if $P_{XY|Z=z} \ll P_{X|Z=z} \times P_{Y|Z=z}$.

For discrete X, we write the probability mass function as p_X. For continuous X, we write the probability density function as f_X. For a general random variable X in a measurable space, we write its distribution as P_X. The uniform distribution over a finite set S is denoted as $\mathrm{Unif}(S)$. The joint distribution of $X_1, \ldots, X_n \overset{iid}{\sim} P_X$ is written as $P_X^{\otimes n}$. The degenerate distribution $\mathbf{P}\{X = a\} = 1$ is denoted as δ_a. The conditional independence of X and Z given Y is denoted as $X \leftrightarrow Y \leftrightarrow Z$.

The Q-function and its inverse are denoted as $\mathcal{Q}(x)$ and $\mathcal{Q}^{-1}(\epsilon)$ respectively. For $V \in \mathbb{R}^{n \times n}$ positive semidefinite, define $\mathcal{Q}^{-1}(V, \epsilon) = \{x \in \mathbb{R}^n : \mathbf{P}\{X \le x\} \ge 1 - \epsilon\}$ where $X \sim N(0, V)$ and $X \le x$ denotes entrywise comparison.

We assume that every random variable mentioned in this paper lies in a Polish space with its Borel σ-algebra, and all functions mentioned (e.g. distortion measures, the function $x(u, s)$ in Theorem 2) are measurable. The Lebesgue measure over $\mathbb{R}$ is denoted as λ. The Lebesgue measure restricted to the set $S \subseteq \mathbb{R}$ is denoted as λ_S. For two measures μ, ν over $\mathcal{X}$ (a Polish space with its Borel σ-algebra) such that ν is absolutely continuous with respect to μ (denoted as $\nu \ll \mu$), the Radon-Nikodym derivative is written as

$$\frac{d\nu}{d\mu} : \mathcal{X} \to [0, \infty).$$

If $\nu_1, \nu_2 \ll \mu$ (but $\nu_1 \ll \nu_2$ may not hold), we write

$$\frac{d\nu_1}{d\nu_2}(x) = \frac{d\nu_1}{d\mu}(x) \left(\frac{d\nu_2}{d\mu}(x) \right)^{-1} \in [0, \infty], \tag{1}$$

which is 0 if $(d\nu_1/d\mu)(x) = 0$, and is ∞ if $(d\nu_1/d\mu)(x) > 0$ and $(d\nu_2/d\mu)(x) = 0$.

The total variation distance between two distributions P, Q over $\mathcal{X}$ is denoted as $\|P - Q\|_{\mathrm{TV}} = \sup_{A \subseteq \mathcal{X} \text{ measurable}} |P(A) - Q(A)|$.

II. Poisson Matching Lemma

We first state the definition of Poisson functional representation in [1], with a different notation that allows the proofs to be written in a simpler and more intuitive manner.

Definition 1 (Poisson functional representation)**.** Let $\{\bar{U}_i, T_i\}_{i \in \mathbb{N}}$ be the points of a Poisson process with intensity measure $\mu \times \lambda_{\mathbb{R}_{\ge 0}}$ on $\mathcal{U} \times \mathbb{R}_{\ge 0}$ (where $\mathcal{U}$ is a Polish space with its Borel σ-algebra, and μ is σ-finite). For $P \ll \mu$ a probability measure over $\mathcal{U}$, define

$$\tilde{U}_P\left(\{\bar{U}_i, T_i\}_{i \in \mathbb{N}}\right) := \bar{U}_{K_P(\{\bar{U}_i, T_i\}_{i \in \mathbb{N}})},$$

where

$$K_P\left(\{\bar{U}_i, T_i\}_{i \in \mathbb{N}}\right) := \operatorname*{arg\,min}_{i:\, \frac{dP}{d\mu}(\bar{U}_i) > 0} T_i \left(\frac{dP}{d\mu}(\bar{U}_i) \right)^{-1},$$

with arbitrary tie-breaking (a tie occurs with probability 0). We omit $\{\bar{U}_i, T_i\}_{i \in \mathbb{N}}$ and only write $\tilde{U}_P$ if the Poisson process is clear from the context. If the Poisson process is $\{\bar{X}_i, T_i\}_{i \in \mathbb{N}}$ instead of $\{\bar{U}_i, T_i\}_{i \in \mathbb{N}}$, then the Poisson functional representation is likewise denoted as $\tilde{X}_P$. If $\bar{U}_i = (\bar{X}_i, \bar{Y}_i)$ is multivariate, and P is a distribution over $\mathcal{X} \times \mathcal{Y}$, the Poisson functional representation is denoted as $(\tilde{X}, \tilde{Y})_P$. We write its components as $(\tilde{X}, \tilde{Y})_P = (\tilde{X}_P, \tilde{Y}_P)$.

Note that while $dP/d\mu$ is only uniquely defined up to a μ-null set, changing the value of $dP/d\mu$ on a μ-null set will only affect the values of $\tilde{U}_P$ on a null set with respect to the distribution of $\{\bar{U}_i, T_i\}_{i \in \mathbb{N}}$, since the probability that there exists $\bar{U}_i$ on that μ-null set is zero. Therefore $\tilde{U}_P$ is uniquely defined up to a null set.

By the mapping theorem [14], [15] (also see Appendix A of [1]), we have $\tilde{U}_P \sim P$. This is termed Poisson functional representation in [1] since it can be regarded as a construction for the functional representation lemma [16]. Consider the distribution $P_{U,X}$. Let $\{\bar{U}_i, T_i\}_{i\in\mathbb{N}}$ be the points of a Poisson process with intensity measure $P_U \times \lambda_{\mathbb{R}_{\ge 0}}$, $X \sim P_X$ independent of the process, and $U := \tilde{U}_{P_{U|X}(\cdot|X)}$. Then $(U, X) \sim P_{U,X}$. Hence we can express U as a function of X and $\{\bar{U}_i, T_i\}$ (which is independent of X). This fact will be used repeatedly throughout the proofs in this paper.

For two different distributions P and Q, $\tilde{U}_P$ and $\tilde{U}_Q$ are coupled in such a way that $\tilde{U}_P = \tilde{U}_Q$ occurs with a probability that can be bounded in terms of dP/dQ. We now present the core lemma of this paper. The proof is given in Appendix A.

Lemma 1 (Poisson matching lemma)**.** *Let $\{\bar{U}_i, T_i\}_{i\in\mathbb{N}}$ be the points of a Poisson process with intensity measure $\mu \times \lambda_{\mathbb{R}_{\ge 0}}$, and P, Q be probability measures on $\mathcal{U}$ with $P, Q \ll \mu$. Then we have the following almost surely:*

$$\mathbf{P}\left\{ \tilde{U}_Q \neq \tilde{U}_P \,\middle|\, \tilde{U}_P \right\} \le 1 - \left(1 + \frac{dP}{dQ}(\tilde{U}_P)\right)^{-1}, \tag{2}$$

where we write $(dP/dQ)(u) = (dP/d\mu)(u)/((dQ/d\mu)(u))$ as in (1) *(we do not require $P \ll Q$). The right hand side of* (2) *is considered to be 1 if $(dP/d\mu)(\tilde{U}_P) > 0$ and $(dQ/d\mu)(\tilde{U}_P) = 0$.*

The exact expression for the left hand side of (2) is in (16).

We usually do not apply the Poisson matching lemma on fixed P, Q, but rather on conditional distributions. The following conditional version of the Poisson matching lemma follows directly from applying the lemma on $(P, Q) \leftarrow (P_{U|X}(\cdot|X), Q_{U|Y}(\cdot|Y))$. The proof is given in Appendix B for the sake of completeness.

Lemma 2 (Conditional Poisson matching lemma)**.** *Fix a distribution $P_{X,U,Y}$ and a probability kernel $Q_{U|Y}$ (that is not necessarily $P_{U|Y}$) satisfying $P_{U|X}(\cdot|X), Q_{U|Y}(\cdot|Y) \ll \mu$ almost surely. Let $X \sim P_X$, and $\{\bar{U}_i, T_i\}_{i\in\mathbb{N}}$ be the points of a Poisson process with intensity measure $\mu \times \lambda_{\mathbb{R}_{\ge 0}}$ independent of X. Let $U = \tilde{U}_{P_{U|X}(\cdot|X)}$ and $Y|(X, U, \{\bar{U}_i, T_i\}_i) \sim P_{Y|X,U}(\cdot|X, U)$ (note that $(X, U, Y) \sim P_{X,U,Y}$ and $Y \leftrightarrow (X, U) \leftrightarrow \{\bar{U}_i, T_i\}_i$). Then we have the following almost surely:*

$$\mathbf{P}\left\{ \tilde{U}_{Q_{U|Y}(\cdot|Y)} \neq U \,\middle|\, X, U, Y \right\} \le 1 - \left(1 + \frac{dP_{U|X}(\cdot|X)}{dQ_{U|Y}(\cdot|Y)}(U)\right)^{-1}.$$

The condition that $P_{U|X}(\cdot|X), Q_{U|Y}(\cdot|Y) \ll \mu$ almost surely is satisfied, for example, when $\mu = P_U$, $Q_{U|Y} = P_{U|Y}$, $P_{UX} \ll P_U \times P_X$ and $P_{UY} \ll P_U \times P_Y$. Note that since $X \perp\!\!\!\perp \{\bar{U}_i, T_i\}_i$, we have $\tilde{U}_{P_{U|X}(\cdot|X)}|X \sim P_{U|X}$, whereas Y may not be independent of $\{\bar{U}_i, T_i\}_i$, so $\tilde{U}_{Q_{U|Y}(\cdot|Y)}$ may not follow the conditional distribution $Q_{U|Y}$.

III. One-shot Channel Coding

To demonstrate the application of the Poisson matching lemma, we apply it to recover a bound for one-shot channel coding in [5] (with a slight penalty of having L instead of $\mathsf{L} - 1$). Upon observing $M \sim \mathrm{Unif}[1 : \mathsf{L}]$, the encoder produces X, which is sent through the channel $P_{Y|X}$. The decoder observes Y and recovers $\hat{M}$ with error probability $P_e = \mathbf{P}\{M \neq \hat{M}\}$.

Proposition 1. *Fix any P_X. There exists a code for the channel $P_{Y|X}$, with message $M \sim \mathrm{Unif}[1 : \mathsf{L}]$, with average error probability*

$$P_e \le \mathbf{E}\left[1 - \left(1 + \mathsf{L}2^{-\iota_{X;Y}(X;Y)}\right)^{-1}\right]$$

if $P_{XY} \ll P_X \times P_Y$.

Proof: Let $\{(\bar{X}_i, \bar{M}_i), T_i\}_{i\in\mathbb{N}}$ be the points of a Poisson process with intensity measure $P_X \times P_M \times \lambda_{\mathbb{R}_{\ge 0}}$ (where P_M is $\mathrm{Unif}[1 : \mathsf{L}]$) independent of M. The encoding function is $m \mapsto \tilde{X}_{P_X \times \delta_m}$ (i.e., $X = \tilde{X}_{P_X \times \delta_M}$), and the decoding function is $y \mapsto \tilde{M}_{P_{X|Y}(\cdot|y) \times P_M}$ (i.e., $\hat{M} = \tilde{M}_{P_{X|Y}(\cdot|Y) \times P_M}$). Note that the encoding and decoding functions also depend on the common randomness $\{(\bar{X}_i, \bar{M}_i), T_i\}_{i\in\mathbb{N}}$, which will be fixed later. We have $(M, X, Y) \sim P_M \times P_X P_{Y|X}$.

$$\begin{aligned}
&\mathbf{P}\left\{ M \neq \tilde{M}_{P_{X|Y}(\cdot|Y) \times P_M} \right\} \\
&\le \mathbf{P}\left\{ (X, M) \neq (\tilde{X}, \tilde{M})_{P_{X|Y}(\cdot|Y) \times P_M} \right\} \\
&= \mathbf{E}\left[\mathbf{P}\left\{ (X, M) \neq (\tilde{X}, \tilde{M})_{P_{X|Y}(\cdot|Y) \times P_M} \,\middle|\, M, X, Y \right\} \right] \\
&\overset{(a)}{\le} \mathbf{E}\left[1 - \left(1 + \frac{dP_X \times \delta_M}{dP_{X|Y}(\cdot|Y) \times P_M}(X, M)\right)^{-1} \right] \\
&= \mathbf{E}\left[1 - (1 + \mathsf{L}2^{-\iota_{X;Y}(X;Y)})^{-1} \right],
\end{aligned}$$

where (a) is by the conditional Poisson matching lemma (Lemma 2) on $(X, U, Y, Q_{U|Y}) \leftarrow (M, (X, M), Y, P_{X|Y} \times P_M)$ (note that $P_{X,M|M} = P_X \times \delta_M$). Therefore there exists a fixed $\{(\bar{x}_i, \bar{m}_i), t_i\}_{i\in\mathbb{N}}$ such that conditioned on $\{(\bar{X}_i, \bar{M}_i), T_i\}_{i\in\mathbb{N}} = \{(\bar{x}_i, \bar{m}_i), t_i\}_{i\in\mathbb{N}}$, the average probability of error is bounded by $\mathbf{E}\left[1 - (1 + \mathsf{L}2^{-\iota_{X;Y}(X;Y)})^{-1}\right]$. ∎

Compared to the scheme in [5], we use the Poisson process $\{(\bar{X}_i, \bar{M}_i), T_i\}$ to create a codebook, instead of the conventional i.i.d. random codebook in [5]. While the codewords for different m's are still i.i.d., we attach a bias T_i to each codeword. Our scheme does not use a stochastic decoder as in [5], but rather a biased maximum likelihood decoder $\tilde{M}_{P_{X|Y}(\cdot|y)\times P_M} = \bar{M}_K$ where $K = \arg\max_i T_i^{-1}(dP_{X|Y}(\cdot|y)/dP_X)(\bar{X}_i)$. In the following sections, we will demonstrate how our approach can lead to simpler proofs and sharper bounds compared to [5].

Using the generalized Poisson matching lemma that will be introduced in Section VII, we can prove the following bound. The proof is in Appendix C.

Theorem 1. *Fix any P_X. There exists a code for the channel $P_{Y|X}$, with message $M \sim \mathrm{Unif}[1 : \mathsf{L}]$, with average error probability*

$$P_e \le \mathbf{E}\left[1 - \left(1 - \min\left\{2^{-\iota_{X;Y}(X;Y)},\, 1\right\}\right)^{(\mathsf{L}+1)/2}\right]$$

if $P_{XY} \ll P_X \times P_Y$.

Compare this to the dependence testing bound [17]:

$$P_e \le \mathbf{E}\left[\min\left\{\frac{\mathsf{L}-1}{2} \cdot 2^{-\iota_{X;Y}(X;Y)},\, 1\right\}\right].$$

Theorem 1 is at least as strong (with a slight penalty of having $(\mathsf{L}+1)/2$ instead of $(\mathsf{L}-1)/2$) since

$$\mathbf{E}\left[1 - \left(1 - \min\left\{2^{-\iota_{X;Y}(X;Y)},\, 1\right\}\right)^{(\mathsf{L}+1)/2}\right] \le \mathbf{E}\left[\min\left\{\frac{\mathsf{L}+1}{2} \cdot 2^{-\iota_{X;Y}(X;Y)},\, 1\right\}\right].$$

Remark 1. Apart from the dependence testing bound [17], there are other one-shot bounds for channel coding such as the random-coding union (RCU) bound and the $\kappa\beta$ bound in [17], which are tighter in certain situations (e.g. the RCU bound is suitable for error exponent analysis). The technique introduced in this paper is suitable for first and second order analysis, but does not seem to give tight error exponent bounds.

IV. One-shot Coding for Channels with State Information at the Encoder

The one-shot coding setting for a channel with state information at the encoder is described as follows. Upon observing $M \sim \mathrm{Unif}[1 : \mathsf{L}]$ and $S \sim P_S$, the encoder produces X, which is sent through the channel $P_{Y|X,S}$ with state S. The decoder observes Y and recovers $\hat{M}$ with error probability $P_e = \mathbf{P}\{M \neq \hat{M}\}$.

We show a one-shot version of the Gelfand-Pinsker theorem [2]. This is the first one-shot bound attaining the best known second order result in [18] (which considers a finite-blocklength, not one-shot scenario). Our bound is stronger than the one-shot bounds in [3], [5], [19] (in the second order), and significantly simpler to state and prove than all the aforementioned results. Unlike previous approaches, our proof does not require sub-codebooks.

Theorem 2. *Fix any $P_{U|S}$ and function $x : \mathcal{U} \times \mathcal{S} \to \mathcal{X}$. There exists a code for the channel $P_{Y|X,S}$ with state distribution P_S with message $M \sim \mathrm{Unif}[1 : \mathsf{L}]$, with error probability*

$$P_e \le \mathbf{E}\left[1 - (1 + \mathsf{L}2^{\iota_{U;S}(U;S) - \iota_{U;Y}(U;Y)})^{-1}\right]$$

if $P_{US} \ll P_U \times P_S$ and $P_{UY} \ll P_U \times P_Y$, where $(S, U, X, Y) \sim P_S P_{U|S} \delta_{x(U,S)} P_{Y|X,S}$.

Proof: Let $\{(\bar{U}_i, \bar{M}_i), T_i\}_{i\in\mathbb{N}}$ be the points of a Poisson process with intensity measure $P_U \times P_M \times \lambda_{\mathbb{R}_{\ge 0}}$ independent of M, S. The encoding function is $(m, s) \mapsto x(\tilde{U}_{P_{U|S}(\cdot|s)\times\delta_m}, s)$ (let $U = \tilde{U}_{P_{U|S}(\cdot|S)\times\delta_M}$, $X = x(U, S)$), and the decoding function is $y \mapsto \tilde{M}_{P_{U|Y}(\cdot|y)\times P_M}$ (i.e., $\hat{M} = \tilde{M}_{P_{U|Y}(\cdot|Y)\times P_M}$). Note that $(M, S, U, X, Y) \sim P_M \times P_S P_{U|S} \delta_{x(U,S)} P_{Y|X,S}$. We have

$$\begin{aligned}
&\mathbf{P}\{M \neq \tilde{M}_{P_{U|Y}(\cdot|Y)\times P_M}\} \\
&\le \mathbf{P}\{(U, M) \neq (\tilde{U}, \tilde{M})_{P_{U|Y}(\cdot|Y)\times P_M}\} \\
&= \mathbf{E}\left[\mathbf{P}\left\{(U, M) \neq (\tilde{U}, \tilde{M})_{P_{U|Y}(\cdot|Y)\times P_M} \,\middle|\, M, S, U, Y\right\}\right]
\end{aligned}$$

$$\overset{(a)}{\leq} \mathbf{E}\left[1-\left(1+\frac{dP_{U|S}(\cdot|S)\times\delta_M}{dP_{U|Y}(\cdot|Y)\times P_M}(U,M)\right)^{-1}\right]$$
$$= \mathbf{E}\left[1-(1+\mathsf{L}2^{\iota_{U;S}(U;S)-\iota_{U;Y}(U;Y)})^{-1}\right].$$

where (a) is by the conditional Poisson matching lemma on $((M,S),\,(U,M),\,Y,\,P_{U|Y}\times P_M)$ (note that $P_{U,M|M,S}=P_{U|S}\times\delta_M$). Therefore there exists a fixed $\{(\bar{u}_i,\bar{m}_i),t_i\}_{i\in\mathbb{N}}$ attaining the desired bound. ∎

Compared to Theorem 3 in [3]:

$$P_e \leq \mathbf{P}\{\iota_{U;S}(U;S)>\log\mathsf{J}-\gamma\}+\mathbf{P}\{\iota_{U;Y}(U;Y)\leq\log\mathsf{LJ}+\gamma\}+2^{-\gamma}+e^{-2^{\gamma}}$$

for any $\gamma>0$, $\mathsf{J}\in\mathbb{N}$, our result is strictly stronger since

$$\begin{aligned}
&\mathbf{E}\left[1-\left(1+\mathsf{L}2^{\iota_{U;S}(U;S)-\iota_{U;Y}(U;Y)}\right)^{-1}\right]\\
&\leq \mathbf{P}\{\iota_{U;S}(U;S)>\log\mathsf{J}-\gamma\}+\mathbf{P}\{\iota_{U;Y}(U;Y)\leq\log\mathsf{LJ}+\gamma\}\\
&\quad+\mathbf{E}\left[1-\left(1+\mathsf{L}2^{\iota_{U;S}(U;S)-\iota_{U;Y}(U;Y)}\right)^{-1}\,\middle|\,\iota_{U;S}(U;S)\leq\log\mathsf{J}-\gamma,\,\iota_{U;Y}(U;Y)>\log\mathsf{LJ}+\gamma\right]\\
&\leq \mathbf{P}\{\iota_{U;S}(U;S)>\log\mathsf{J}-\gamma\}+\mathbf{P}\{\iota_{U;Y}(U;Y)\leq\log\mathsf{LJ}+\gamma\}+2^{-2\gamma}\\
&< \mathbf{P}\{\iota_{U;S}(U;S)>\log\mathsf{J}-\gamma\}+\mathbf{P}\{\iota_{U;Y}(U;Y)\leq\log\mathsf{LJ}+\gamma\}+2^{-\gamma}+e^{-2^{\gamma}}.
\end{aligned}$$

This is due to the fact that the Poisson matching lemma simultaneously replaces both the covering and the packing lemma, resulting in only one error event.

Next, we prove a second-order result. Fix $\epsilon>0$. Let $C:=I(U;Y)-I(U;S)$, $V:=\mathrm{Var}[\iota_{U;S}(U;S)-\iota_{U;Y}(U;Y)]$. We apply Theorem 2 on n uses of the memoryless channel with i.i.d. state sequence $S^n=(S_1,\ldots,S_n)$, and

$$\mathsf{L}:=\left\lfloor\exp_2\left(nC-\sqrt{nV}\mathcal{Q}^{-1}\left(\epsilon-\frac{\alpha}{\sqrt{n}}\right)-\frac{1}{2}\log n\right)\right\rfloor,$$

where α is a constant that depends on $P_{S,U,Y}$. For $n>\alpha^2\epsilon^{-2}$, by the Berry-Esseen theorem [20], [21], [22], we have

$$\begin{aligned}
P_e&\leq\mathbf{E}\left[\min\left\{2^{\log\mathsf{L}+\iota_{U^n;S^n}(U^n;S^n)-\iota_{U^n;Y^n}(U^n;Y^n)},\,1\right\}\right]\\
&\leq\frac{1}{\sqrt{n}}+\mathbf{P}\left\{2^{\log\mathsf{L}+\iota_{U^n;S^n}(U^n;S^n)-\iota_{U^n;Y^n}(U^n;Y^n)}>\frac{1}{\sqrt{n}}\right\}\\
&\leq\frac{1}{\sqrt{n}}+\mathbf{P}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\iota_{U;Y}(U_i;Y_i)-\iota_{U;S}(U_i;S_i)-C)<-\sqrt{V}\mathcal{Q}^{-1}\left(\epsilon-\frac{\alpha}{\sqrt{n}}\right)\right\}\\
&\leq\frac{1}{\sqrt{n}}+\epsilon-\frac{\alpha}{\sqrt{n}}+\frac{\alpha-1}{\sqrt{n}}\\
&\leq\epsilon
\end{aligned}$$

if we let $\alpha-1$ be the constant given by the Berry-Esseen theorem. This coincides with the best known second order result in [18], which is stronger than the second order results implied by [3], [5], [19]. We bound $\iota_{U;S}(U;S)-\iota_{U;Y}(U;Y)$ as a single quantity, instead of bounding the two terms separately as in [3], [5], [19], resulting in a sharper second order bound.

V. One-shot Lossy Source Coding with Side Information at the Decoder

The one-shot lossy source coding setting with side information at the decoder is described as follows. Upon observing $X\sim P_X$, the encoder produces $M\in[1:\mathsf{L}]$. The decoder observes M and $Y\sim P_{Y|X}$ and recovers $\hat{Z}\in\mathcal{Z}$ with probability of excess distortion $P_e=\mathbf{P}\{\mathsf{d}(X,\hat{Z})>\mathsf{D}\}$, where $\mathsf{d}:\mathcal{X}\times\mathcal{Z}\to\mathbb{R}_{\geq 0}$ is a distortion measure.

We show a one-shot version of the Wyner-Ziv theorem [8], [9]. Our bound is stronger than those in [3], [19], and significantly simpler to state and prove. Unlike previous approaches, our proof does not require binning.

Theorem 3. *Fix any $P_{U|X}$ and function $z:\mathcal{U}\times\mathcal{Y}\to\mathcal{Z}$. There exists a code for lossy source coding with source distribution P_X, side information at the decoder given by $P_{Y|X}$, and message size L, with probability of excess distortion*

$$P_e\leq\mathbf{E}\left[1-\mathbf{1}\{\mathsf{d}(X,Z)\leq\mathsf{D}\}(1+\mathsf{L}^{-1}2^{\iota_{U;X}(U;X)-\iota_{U;Y}(U;Y)})^{-1}\right]$$

if $P_{UX}\ll P_U\times P_X$ and $P_{UY}\ll P_U\times P_Y$, where $(X,Y,U,Z)\sim P_XP_{Y|X}P_{U|X}\delta_{z(U,Y)}$.

Proof: Let $\{(\bar{U}_i,\bar{M}_i),T_i\}_{i\in\mathbb{N}}$ be the points of a Poisson process with intensity measure $P_U\times P_M\times\lambda_{\mathbb{R}_{\geq 0}}$ independent of X, where P_M is $\mathrm{Unif}[1:\mathsf{L}]$. The encoding function is $x\mapsto\tilde{M}_{P_{U|X}(\cdot|x)\times P_M}$ (i.e., $M=\tilde{M}_{P_{U|X}(\cdot|X)\times P_M}$), and the decoding function is $(m,y)\mapsto z(\tilde{U}_{P_{U|Y}(\cdot|y)\times\delta_m},y)$ (let $\hat{U}=\tilde{U}_{P_{U|Y}(\cdot|Y)\times\delta_M}$, $\hat{Z}=z(\hat{U},Y)$). Also define $U=\tilde{U}_{P_{U|X}(\cdot|X)\times P_M}$, $Z=z(U,Y)$. Note that $(M,X,Y,U,Z)\sim P_M\times P_X P_{Y|X}P_{U|X}\delta_{z(U,Y)}$. We have

$$\begin{aligned}
&\mathbf{P}\{\mathsf{d}(X,\hat{Z})>\mathsf{D}\}\\
&\le 1-\mathbf{P}\{\mathsf{d}(X,Z)\le\mathsf{D}\,\mathrm{and}\,U=\hat{U}\}\\
&\le\mathbf{E}\left[1-\mathbf{1}\{\mathsf{d}(X,Z)\le\mathsf{D}\}\mathbf{P}\{(U,M)=(\tilde{U},\tilde{M})_{P_{U|Y}(\cdot|Y)\times\delta_M}\,|\,M,X,Y,U\}\right]\\
&\stackrel{(a)}{\le}\mathbf{E}\left[1-\mathbf{1}\{\mathsf{d}(X,Z)\le\mathsf{D}\}\left(1+\frac{dP_{U|X}(\cdot|X)\times P_M}{dP_{U|Y}(\cdot|Y)\times\delta_M}(U,M)\right)^{-1}\right]\\
&\le\mathbf{E}\left[1-\mathbf{1}\{\mathsf{d}(X,Z)\le\mathsf{D}\}(1+\mathsf{L}^{-1}2^{\iota_{U;X}(U;X)-\iota_{U;Y}(U;Y)})^{-1}\right].
\end{aligned}$$

where (a) is by the conditional Poisson matching lemma on $(X,\,(U,M),\,(M,Y),\,P_{U|Y}\times\delta_M)$ (note that $P_{U,M|X}=P_{U|X}\times P_M$). Therefore there exists a fixed $\{(\bar{u}_i,\bar{m}_i),t_i\}_{i\in\mathbb{N}}$ attaining the desired bound. ■

This reduces to lossy source coding (without side information) when $Y=\emptyset$. Note that the encoder is designed in the same way with or without side information. An encoder for lossy source coding is sufficient to achieve the bound in Theorem 3 even when side information is present. Binning is not required at the encoder.

Similar to the case in Section IV, it can be checked that our bound is stronger than that in Theorem 2 in [3]. Compared to Corollary 9 in [19]:

$$P_e\le\mathbf{P}\{\iota_{U;X}(U;X)>\gamma_{\mathrm{c}}\text{ or }\iota_{U;Y}(U;Y)<\gamma_{\mathrm{p}}\text{ or }\mathsf{d}(X,Z)>\mathsf{D}\}+\frac{\mathsf{J}}{2^{\gamma_{\mathrm{p}}}\mathsf{L}}+\frac{1}{2}\sqrt{\frac{2^{\gamma_{\mathrm{c}}}}{\mathsf{J}}}\tag{3}$$

for any $\gamma_{\mathrm{p}},\gamma_{\mathrm{c}}>0$, $\mathsf{J}\in\mathbb{N}$, our result is stronger since

$$\begin{aligned}
&\mathbf{E}\left[1-\mathbf{1}\{\mathsf{d}(X,Z)\le\mathsf{D}\}(1+\mathsf{L}^{-1}2^{\iota_{U;X}(U;X)-\iota_{U;Y}(U;Y)})^{-1}\right]\\
&\le\mathbf{P}\{\iota_{U;X}(U;X)>\gamma_{\mathrm{c}}\text{ or }\iota_{U;Y}(U;Y)<\gamma_{\mathrm{p}}\text{ or }\mathsf{d}(X,Z)>\mathsf{D}\}+\mathsf{L}^{-1}2^{\gamma_{\mathrm{c}}-\gamma_{\mathrm{p}}}\\
&\le\mathbf{P}\{\iota_{U;X}(U;X)>\gamma_{\mathrm{c}}\text{ or }\iota_{U;Y}(U;Y)<\gamma_{\mathrm{p}}\text{ or }\mathsf{d}(X,Z)>\mathsf{D}\}+\frac{\mathsf{J}}{2^{\gamma_{\mathrm{p}}}\mathsf{L}}+\frac{1}{2}\sqrt{\frac{2^{\gamma_{\mathrm{c}}}}{\mathsf{J}}},
\end{aligned}$$

where the last inequality is due to

$$a+b\ge(a+b)^3=27\left(\frac{a+2(b/2)}{3}\right)^3\ge 27a(b/2)^2\ge 4ab^2$$

by the AM-GM inequality for $a,b\ge 0$, $a+b\le 1$ (since the right hand side of (3) ≤ 1 for it to be meaningful). We bound $\iota_{U;X}(U;X)-\iota_{U;Y}(U;Y)$ as a single quantity, instead of bounding the two terms separately, resulting in a sharper bound.

VI. One-shot Joint Source-Channel Coding

The one-shot joint source-channel coding setting is described as follows. Upon observing the source symbol $W\sim P_W$, the encoder produces $X\in\mathcal{X}$, which is sent through the channel $P_{Y|X}$. The decoder observes Y and recovers $\hat{Z}\in\mathcal{Z}$ with probability of excess distortion $P_e=\mathbf{P}\{\mathsf{d}(W,\hat{Z})>\mathsf{D}\}$, where $\mathsf{d}:\mathcal{W}\times\mathcal{Z}\to\mathbb{R}_{\ge 0}$ is a distortion measure.

We show a one-shot joint source-channel coding result that achieves the optimal dispersion in [10].

Theorem 4. *Fix any P_X and P_Z. There exists a code for the source distribution P_W and channel $P_{Y|X}$, with probability of excess distortion*

$$P_e\le\mathbf{E}\left[\left(1+P_Z(\mathcal{B}_{\mathsf{D}}(W))2^{\iota_{X;Y}(X;Y)}\right)^{-1}\right]$$

if $P_{XY}\ll P_X\times P_Y$, where $(W,X,Y)\sim P_W\times P_X P_{Y|X}$, and $\mathcal{B}_{\mathsf{D}}(w):=\{z:\,\mathsf{d}(w,z)\le\mathsf{D}\}$.

Proof: Let $\{(\bar{X}_i,\bar{Z}_i),T_i\}_{i\in\mathbb{N}}$ be the points of a Poisson process with intensity measure $P_X\times P_Z\times\lambda_{\mathbb{R}_{\ge 0}}$ independent of W. Let $\rho(w):=P_Z(\mathcal{B}_{\mathsf{D}}(w))$. Let $P_{\check{Z}|W}$ be defined as

$$P_{\check{Z}|W}(A|w):=\begin{cases}P_Z(A\cap\mathcal{B}_{\mathsf{D}}(w))/\rho(w) & \text{if }\rho(w)>0\\ P_Z(A) & \text{if }\rho(w)=0.\end{cases}$$

The encoding function is $w\mapsto\tilde{X}_{P_X\times P_{\check{Z}|W}(\cdot|w)}$ (i.e., $X=\tilde{X}_{P_X\times P_{\check{Z}|W}(\cdot|W)}$). The decoding function is $y\mapsto\tilde{Z}_{P_{X|Y}(\cdot|y)\times P_Z}$ (i.e., $\hat{Z}=\tilde{Z}_{P_{X|Y}(\cdot|Y)\times P_Z}$). Also define $\check{Z}=\tilde{Z}_{P_X\times P_{\check{Z}|W}(\cdot|W)}$. We have $(X,Y,W,\check{Z})\sim P_X P_{Y|X}\times P_W P_{\check{Z}|W}$.

$$\mathbf{P}\{\mathsf{d}(W,\hat{Z})>\mathsf{D}\}$$

$$
\begin{aligned}
&\leq \mathbf{P}\{\rho(W)=0\} + \mathbf{P}\{\rho(W)>0 \text{ and } \check{Z}\neq\hat{Z}\} \\
&\leq \mathbf{P}\{\rho(W)=0\} + \mathbf{E}\left[\mathbf{1}\{\rho(W)>0\}\mathbf{P}\{(X,\check{Z})\neq(\tilde{X},\tilde{Z})_{P_{X|Y}(\cdot|Y)\times P_Z}\,|\,X,Y,W,\check{Z}\}\right] \\
&\stackrel{(a)}{\leq} \mathbf{P}\{\rho(W)=0\} + \mathbf{E}\left[\mathbf{1}\{\rho(W)>0\}\left(1-\left(1+\frac{dP_X\times P_{\check{Z}|W}(\cdot|W)}{dP_{X|Y}(\cdot|Y)\times P_Z}(X,\check{Z})\right)^{-1}\right)\right] \\
&= \mathbf{P}\{\rho(W)=0\} + \mathbf{E}\left[\mathbf{1}\{\rho(W)>0\}\left(1-\left(1+(\rho(W))^{-1}2^{-\iota_{X;Y}(X;Y)}\right)^{-1}\right)\right] \\
&= \mathbf{E}\left[\left(1+\rho(W)2^{\iota_{X;Y}(X;Y)}\right)^{-1}\right],
\end{aligned}
$$

where (a) is by the conditional Poisson matching lemma on $(W,\,(X,\check{Z}),\,Y,\,P_{X|Y}\times P_Z)$ (note that $P_{X,\check{Z}|W}=P_X\times P_{\check{Z}|W}$). Therefore there exists a fixed $\{(\bar{x}_i,\bar{z}_i),t_i\}_{i\in\mathbb{N}}$ attaining the desired bound. ■

Compare this to Theorem 7 in [10]:

$$
P_e \leq \mathbf{E}\left[\min\left\{J2^{-\iota_{X;Y}(X;Y)},\,1\right\}\right] + \mathbf{E}\left[(1-P_Z(\mathcal{B}_{\mathsf{D}}(W)))^J\right] \tag{4}
$$

for any $P_{J|W}$, $J\in\mathbb{N}$. While neither of the bounds implies the other, our bound is at least within a factor of 2 from (4), since

$$
\begin{aligned}
&\mathbf{E}\left[\left(1+P_Z(\mathcal{B}_{\mathsf{D}}(W))2^{\iota_{X;Y}(X;Y)}\right)^{-1}\right] \\
&\leq \mathbf{E}\left[\left(1+(2J)^{-1}2^{\iota_{X;Y}(X;Y)}\right)^{-1}\right] + \mathbf{P}\left\{(2J)^{-1}\geq P_Z(\mathcal{B}_{\mathsf{D}}(W))\right\} \\
&\leq \mathbf{E}\left[\min\left\{2J2^{-\iota_{X;Y}(X;Y)},\,1\right\}\right] + 2\mathbf{E}\left[\max\{1-JP_Z(\mathcal{B}_{\mathsf{D}}(W)),\,0\}\right] \\
&\leq 2\mathbf{E}\left[\min\left\{J2^{-\iota_{X;Y}(X;Y)},\,1\right\}\right] + 2\mathbf{E}\left[(1-P_Z(\mathcal{B}_{\mathsf{D}}(W)))^J\right].
\end{aligned}
$$

However, (4) does not imply a bound that is within a constant factor from our bound. Theorem 8 in [10] is obtained by substituting $J=\lfloor\gamma/P_Z(\mathcal{B}_{\mathsf{D}}(W))\rfloor$ in (4):

$$
P_e \leq \mathbf{E}\left[\min\left\{\gamma P_Z(\mathcal{B}_{\mathsf{D}}(W))^{-1}2^{-\iota_{X;Y}(X;Y)},\,1\right\}\right] + e^{1-\gamma},
$$

which is strictly weaker than our bound with an unbounded multiplicative gap γ (that tends to ∞ when the bound tends to 0). Hence our bound is stronger than Theorem 7 and 8 in [10] (ignoring constant multiplicative gaps). Also our proof is significantly shorter than that of Theorem 7 in [10].

Please refer to Appendix D for the proof that Theorem 4 achieves the optimal dispersion.

VII. Poisson Matching Lemma Beyond the First Index

The Poisson functional representation concerns the point with the smallest $T_i((dP/d\mu)(\bar{U}_i))^{-1}$. We can generalize it to obtain a sequence ordered in ascending order of $T_i((dP/d\mu)(\bar{U}_i))^{-1}$.

Definition 2 (Mapped Poisson process)**.** Let $\{\bar{U}_i,T_i\}_{i\in\mathbb{N}}$ be the points of a Poisson process with intensity measure $\mu\times\lambda_{\mathbb{R}_{\geq 0}}$ on $\mathcal{U}\times\mathbb{R}_{\geq 0}$ (where $\mathcal{U}$ is a Polish space with its Borel σ-algebra, and μ is σ-finite). For $P\ll\mu$ a probability measure over $\mathcal{U}$, let $i_{P,1},i_{P,2},\ldots\in\mathbb{N}$ be a sequence of distinct integers such that $\bigcup_{j=1}^{\infty}\{i_{P,j}\}=\{i:(dP/d\mu)(\bar{U}_i)>0\}$ and $\{T_{i_{P,j}}((dP/d\mu)(\bar{U}_{i_{P,j}}))^{-1}\}_{j\in\mathbb{N}}$ is sorted in ascending order with arbitrary tie-breaking (a tie occurs with probability 0). For $j\in\mathbb{N}$, $u\in\mathcal{U}$, define the *mapped Poisson process with respect to* P as

$$
\left\{\tilde{U}_P\left(\{\bar{U}_i,T_i\}_{i\in\mathbb{N}},\,j\right),\,\tilde{T}_P\left(\{\bar{U}_i,T_i\}_{i\in\mathbb{N}},\,j\right)\right\}_{j\in\mathbb{N}}, \tag{5}
$$

where

$$
\tilde{T}_P\left(\{\bar{U}_i,T_i\}_{i\in\mathbb{N}},\,j\right) := T_{i_{P,j}}\left(\frac{dP}{d\mu}(\bar{U}_{i_{P,j}})\right)^{-1},
$$

$$
\tilde{U}_P\left(\{\bar{U}_i,T_i\}_{i\in\mathbb{N}},\,j\right) := \bar{U}_{i_{P,j}}.
$$

For $P,Q\ll\mu$ probability measures over $\mathcal{U}$, define $i_{P,1},i_{P,2},\ldots\in\mathbb{N}$ and $i_{Q,1},i_{Q,2},\ldots\in\mathbb{N}$ as above. Define

$$
\Upsilon_{P\|Q}\left(\{\bar{U}_i,T_i\}_{i\in\mathbb{N}},\,j\right) := \min\{k\in\mathbb{N}:\,i_{Q,k}=i_{P,j}\},
$$

where the minimum is ∞ if such k does not exist. We omit $\{\bar{U}_i,T_i\}_{i\in\mathbb{N}}$ and only write $\tilde{U}_P(j)$, $\tilde{T}_P(j)$, $\Upsilon_{P\|Q}(j)$ if the Poisson process is clear from the context. Note that, with probability 1, we have either $\tilde{U}_Q(\Upsilon_{P\|Q}(j))=\tilde{U}_P(j)$ or $\Upsilon_{P\|Q}(j)=\infty$.

Also, for any $j,k \in \mathbb{N}$, $\Upsilon_{P\|Q}(j)=k \Leftrightarrow \Upsilon_{Q\|P}(k)=j$. Loosely speaking, $\Upsilon_{P\|Q}(j)$ can be regarded as "$\tilde{U}_Q^{-1}(\tilde{U}_P(j))$" (if there are no atoms in μ), i.e., finding the j-th point in the mapped Poisson process w.r.t. P, then finding its index in the mapped Poisson process w.r.t. Q.

While $dP/d\mu$ is only uniquely defined up to a μ-null set, changing the value of $dP/d\mu$ on a μ-null set will only affect the values of $\{\tilde{U}_P(j),\, \tilde{T}_P(j)\}_{j\in\mathbb{N}}$ on a null set with respect to the distribution of $\{\bar{U}_i, T_i\}_{i\in\mathbb{N}}$, since the probability that there exists $\bar{U}_i$ in that μ-null set is zero. Therefore $\{\tilde{U}_P(j),\, \tilde{T}_P(j)\}_{j\in\mathbb{N}}$ is uniquely defined up to a null set. The same is true for $\Upsilon_{P\|Q}(j)$.

By the mapping theorem [14], [15] (also see Appendix A of [1]),

$$\{\bar{U}_{i_{P,j}},\, T_{i_{P,j}}((dP/d\mu)(U_{i_{P,j}}))^{-1}\}_{j\in\mathbb{N}} = \{\tilde{U}_P(j), \tilde{T}_P(j)\}_{j\in\mathbb{N}}$$

is a Poisson process with intensity measure $P \times \lambda_{\mathbb{R}_{\ge 0}}$. Hence

$$\tilde{U}_P(1), \tilde{U}_P(2), \ldots \overset{iid}{\sim} P.$$

We present a generalized Poisson matching lemma concerning the indices beyond the first. The proof is given in Appendix A.

Lemma 3 (Generalized Poisson matching lemma)**.** *Let $\{\bar{U}_i, T_i\}_{i\in\mathbb{N}}$ be the points of a Poisson process with intensity measure $\mu \times \lambda_{\mathbb{R}_{\ge 0}}$ on $\mathcal{U} \times \mathbb{R}_{\ge 0}$, and P, Q be probability measures over $\mathcal{U}$ with $P, Q \ll \mu$. Fix any $j \in \mathbb{N}$. Then we have the following almost surely:*

$$\mathbf{E}\left[\Upsilon_{P\|Q}(j) \,\middle|\, \tilde{U}_P(j)\right] \le j\frac{dP}{dQ}(\tilde{U}_P(j)) + 1,$$

where we write $(dP/dQ)(u) = (dP/d\mu)(u)/((dQ/d\mu)(u))$ as in (1) *(we do not require $P \ll Q$). As a result, we have the following almost surely: for all $k \in \mathbb{N}$,*

$$\begin{aligned}\mathbf{P}\left\{\tilde{U}_P(j) \notin \{\tilde{U}_Q(i)\}_{i\in[1:k]} \,\middle|\, \tilde{U}_P(j)\right\} &\le \mathbf{P}\left\{\Upsilon_{P\|Q}(j) > k \,\middle|\, \tilde{U}_P(j)\right\} \\ &\le \min\left\{\frac{j}{k}\frac{dP}{dQ}(\tilde{U}_P(j)),\, 1\right\}.\end{aligned}$$

For $k = 1$, this can be slightly strengthened to

$$\mathbf{P}\left\{\Upsilon_{P\|Q}(j) > 1 \,\middle|\, \tilde{U}_P(j)\right\} \le 1 - \left(1 - \min\left\{\frac{dP}{dQ}(\tilde{U}_P(j)),\, 1\right\}\right)^j.$$

For $j = 1$, this can be slightly strengthened to: for all $k \in \mathbb{N}$,

$$\begin{aligned}\mathbf{P}\left\{\Upsilon_{P\|Q}(1) > k \,\middle|\, \tilde{U}_P(1)\right\} &\le \left(1 - \left(1 + \frac{dP}{dQ}(\tilde{U}_P(1))\right)^{-1}\right)^k \\ &\le 1 - \left(1 + k^{-1}\frac{dP}{dQ}(\tilde{U}_P(1))\right)^{-1}.\end{aligned}$$

The exact distribution of $\Upsilon_{P\|Q}(j)$ is given in (15).

Similar to Lemma 2, we can state a conditional version of the generalized Poisson matching lemma. The proof follows the same logic as Lemma 2 and is omitted.

Lemma 4 (Conditional generalized Poisson matching lemma)**.** *Fix a distribution $P_{X,J,U,Y}$ and a probability kernel $Q_{U|Y}$, satisfying $J \in \mathbb{N}$ and $P_{U|X,J}(\cdot|X,J), Q_{U|Y}(\cdot|Y) \ll \mu$ almost surely. Let $(X,J) \sim P_{X,J}$, and $\{\bar{U}_i, T_i\}_{i\in\mathbb{N}}$ be the points of a Poisson process with intensity measure $\mu\times\lambda_{\mathbb{R}_{\ge 0}}$ independent of (X,J). Let $U = \tilde{U}_{P_{U|X,J}(\cdot|X,J)}(J)$ and $Y|(X,J,U,\{\bar{U}_i,T_i\}_i) \sim P_{Y|X,J,U}(\cdot|X,J,U)$ (note that $(X,J,U,Y) \sim P_{X,J,U,Y}$ and $Y \leftrightarrow (X,J,U) \leftrightarrow \{\bar{U}_i, T_i\}_i$). Then we have the following almost surely:*

$$\mathbf{E}\left[\Upsilon_{P_{U|X,J}(\cdot|X,J)\|Q_{U|Y}(\cdot|Y)}(J) \,\middle|\, X,J,U,Y\right] \le J\frac{dP_{U|X,J}(\cdot|X,J)}{dQ_{U|Y}(\cdot|Y)}(U) + 1,$$

and for all $k \in \mathbb{N}$,

$$\mathbf{P}\left\{\Upsilon_{P_{U|X,J}(\cdot|X,J)\|Q_{U|Y}(\cdot|Y)}(J) > k \,\middle|\, X,J,U,Y\right\} \le \min\left\{\frac{J}{k}\frac{dP_{U|X,J}(\cdot|X,J)}{dQ_{U|Y}(\cdot|Y)}(U),\, 1\right\},$$

and

$$\mathbf{P}\left\{\Upsilon_{P_{U|X,J}(\cdot|X,J)\|Q_{U|Y}(\cdot|Y)}(J) > 1 \,\middle|\, X,J,U,Y\right\} \le 1 - \left(1 - \min\left\{\frac{dP_{U|X,J}(\cdot|X,J)}{dQ_{U|Y}(\cdot|Y)}(U),\, 1\right\}\right)^J.$$

If $J = 1$ almost surely, then we also have the following almost surely: for all $k \in \mathbb{N}$,

$$\mathbf{P}\left\{ \Upsilon_{P_{U|X}(\cdot|X)\|Q_{U|Y}(\cdot|Y)}(1) > k \,\middle|\, X, U, Y \right\} \le \left(1 - \left(1 + \frac{dP_{U|X}(\cdot|X)}{dQ_{U|Y}(\cdot|Y)}(U)\right)^{-1}\right)^k$$
$$\le 1 - \left(1 + k^{-1}\frac{dP_{U|X}(\cdot|X)}{dQ_{U|Y}(\cdot|Y)}(U)\right)^{-1}.$$

Remark 2. We can use the generalized Poisson matching lemma to extend Proposition 1 to the list decoding setting with fixed list size J. The decoder outputs the list $\{\tilde{M}_{P_{X|Y}(\cdot|Y)\times P_M}(j)\}_{j\in[1:\mathsf{J}]}$. The error event becomes $(X, M) \notin \{(\tilde{X}, \tilde{M})_{P_{X|Y}(\cdot|Y)\times P_M}(j)\}_{j\in[1:\mathsf{J}]}$. The probability of error is bounded by $\mathbf{E}\left[(1-(1+\mathsf{L}2^{-\iota_{X;Y}(X;Y)})^{-1})^{\mathsf{J}}\right]$.

VIII. One-shot Coding for Broadcast Channels and Mutual Covering

The one-shot coding setting for the broadcast channel with common message is described as follows. Upon observing three independent messages $M_j \sim \mathrm{Unif}[1 : \mathsf{L}_j]$, $j = 0, 1, 2$, the encoder produces X, which is sent through the broadcast channel $P_{Y_1,Y_2|X}$. Decoder j observes Y_j and recovers $\hat{M}_{0j}$ and $\hat{M}_j$ ($j = 1, 2$). The error probability is $P_e = \mathbf{P}\{(M_0, M_0, M_1, M_2) \neq (\hat{M}_{01}, \hat{M}_{02}, \hat{M}_1, \hat{M}_2)\}$.

We show a one-shot version of the inner bound in [23, Theorem 5] (which is shown to be equivalent to [24, Theorem 1] in [25]). The proof is given in Appendix F.

Theorem 5. *Fix any P_{U_0,U_1,U_2} and function $x : \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \to \mathcal{X}$. For any $\mathsf{J}, \mathsf{K}_1, \mathsf{K}_2 \in \mathbb{N}$, there exists a code for the broadcast channel $P_{Y_1,Y_2|X}$ for independent messages $M_j \sim \mathrm{Unif}[1 : \mathsf{L}_j]$, $j = 0, 1, 2$, with the error probability bounded by*

$$\begin{aligned}P_e \le \mathbf{E}\Bigg[\min\Big\{&\tilde{\mathsf{L}}_0\tilde{\mathsf{L}}_1\mathsf{J}A2^{-\iota_{U_0,U_1;Y_1}(U_0,U_1;Y_1)} + \tilde{\mathsf{L}}_1\mathsf{J}A2^{-\iota_{U_1;Y_1|U_0}(U_1;Y_1|U_0)}\\
&+ \tilde{\mathsf{L}}_0\tilde{\mathsf{L}}_2\mathsf{J}^{-1}B2^{\iota_{U_1;U_2|U_0}(U_1;U_2|U_0)-\iota_{U_0,U_2;Y_2}(U_0,U_2;Y_2)} + \tilde{\mathsf{L}}_0\tilde{\mathsf{L}}_2(1-\mathsf{J}^{-1})B2^{-\iota_{U_0,U_2;Y_2}(U_0,U_2;Y_2)}\\
&+ \tilde{\mathsf{L}}_2\mathsf{J}^{-1}B2^{\iota_{U_1;U_2|U_0}(U_1;U_2|U_0)-\iota_{U_2,Y_2|U_0}(U_2;Y_2|U_0)} + \tilde{\mathsf{L}}_2(1-\mathsf{J}^{-1})B2^{-\iota_{U_2,Y_2|U_0}(U_2;Y_2|U_0)},\, 1\Big\}\Bigg]\end{aligned}$$

if all the information density terms are defined almost surely, where

$$\begin{aligned}\tilde{\mathsf{L}}_0 &:= \mathsf{L}_0\mathsf{K}_1\mathsf{K}_2,\\
\tilde{\mathsf{L}}_a &:= \lceil \mathsf{L}_a/\mathsf{K}_a\rceil \text{ for } a = 1, 2,\\
A &:= (\log(\tilde{\mathsf{L}}_1^{-1}\mathsf{J}^{-1}2^{\iota_{U_1;Y_1|U_0}(U_1;Y_1|U_0)} + 1) + 1)^2,\\
B &:= \Big(\log((\tilde{\mathsf{L}}_2\mathsf{J}^{-1}2^{\iota_{U_1;U_2|U_0}(U_1;U_2|U_0)-\iota_{U_2,Y_2|U_0}(U_2;Y_2|U_0)}\\
&\qquad + \tilde{\mathsf{L}}_2(1-\mathsf{J}^{-1})2^{-\iota_{U_2,Y_2|U_0}(U_2;Y_2|U_0)})^{-1} + 1) + 1\Big)^2.\end{aligned}$$

As a result, for $\gamma > 0$,

$$\begin{aligned}P_e \le \mathbf{P}\Big\{&\log\tilde{\mathsf{L}}_1\mathsf{J} > \iota_{U_1;Y_1|U_0}(U_1;Y_1|U_0) - \gamma \;\text{ or }\; \log\tilde{\mathsf{L}}_2 > \iota_{U_2,Y_2|U_0}(U_2;Y_2|U_0) - \gamma\\
&\text{or }\; \log\tilde{\mathsf{L}}_2\mathsf{J}^{-1} > \iota_{U_2,Y_2|U_0}(U_2;Y_2|U_0) - \iota_{U_1;U_2|U_0}(U_1;U_2|U_0) - \gamma\\
&\text{or }\; \log\tilde{\mathsf{L}}_0\tilde{\mathsf{L}}_1\mathsf{J} > \iota_{U_0,U_1;Y_1}(U_0,U_1;Y_1) - \gamma \;\text{ or }\; \log\tilde{\mathsf{L}}_0\tilde{\mathsf{L}}_2 > \iota_{U_0,U_2;Y_2}(U_0,U_2;Y_2) - \gamma\\
&\text{or }\; \log\tilde{\mathsf{L}}_0\tilde{\mathsf{L}}_2\mathsf{J}^{-1} > \iota_{U_0,U_2;Y_2}(U_0,U_2;Y_2) - \iota_{U_1;U_2|U_0}(U_1;U_2|U_0) - \gamma\Big\}\\
&+ 2^{-\gamma}\left(8\mathbf{E}\left[(\iota_{U_1;Y_1|U_0}(U_1;Y_1|U_0))^2 + (\iota_{U_2,Y_2|U_0}(U_2;Y_2|U_0))^2\right] + 12\gamma^2 + 84\right).\end{aligned} \tag{6}$$

The logarithmic terms A and B (or the last term in (6)) result in an $O(n^{-1}\log n)$ penalty on the rate in the finite blocklength regime, and do not affect the second order result. Ignoring the last term in (6), the error event in (6) is a strict subset of those in [5, eqn (32)] and [4, eqn (49)]. This is because the error event in [5] is a superset of (6) by Fourier-Motzkin elimination on J_2 in the error event in [5], but the reverse is not true since Fourier-Motzkin elimination only guarantees the existence of a random variable for J_2 (that depends on the information density terms) satisfying the bounds, but J_2 must be a constant since it is a parameter of the code construction in [5].

Theorem 5 gives the following second order bound. Consider n independent channel uses. Let $\mathsf{L}_a = 2^{nR_a}$ for $a = 0, 1, 2$. By the multi-dimensional Berry-Esseen theorem [26] (using the notation in [5]), we have $P_e \le \epsilon$ if there exists $\bar{R}, \hat{R}_1, \hat{R}_2 \ge 0$ such that

$$\begin{bmatrix} \tilde{R}_1 + \bar{R} \\ \tilde{R}_2 \\ \tilde{R}_2 - \bar{R} \\ \tilde{R}_0 + \tilde{R}_1 + \bar{R} \\ \tilde{R}_0 + \tilde{R}_2 \\ \tilde{R}_0 + \tilde{R}_2 - \bar{R} \end{bmatrix} \in \mathbf{E}[I] - \frac{1}{\sqrt{n}} \mathcal{Q}^{-1}\left(\mathrm{Cov}[I],\, \epsilon - \frac{\beta}{\sqrt{n}}\right) - \frac{\beta \log n}{n}$$

if $n > \beta^2 \epsilon^{-2}$, where β is a constant that depends on P_{U_0,U_1,U_2,Y_1,Y_2}, and $\tilde{R}_0 = R_0 + \hat{R}_1 + \hat{R}_2$, $\tilde{R}_a = R_a - \hat{R}_a$ for $a = 1, 2$, and

$$I = \begin{bmatrix} \iota_{U_1;Y_1|U_0}(U_1; Y_1|U_0) \\ \iota_{U_2,Y_2|U_0}(U_2; Y_2|U_0) \\ \iota_{U_2,Y_2|U_0}(U_2; Y_2|U_0) - \iota_{U_1;U_2|U_0}(U_1; U_2|U_0) \\ \iota_{U_0,U_1;Y_1}(U_0, U_1; Y_1) \\ \iota_{U_0,U_2;Y_2}(U_0, U_2; Y_2) \\ \iota_{U_0,U_2;Y_2}(U_0, U_2; Y_2) - \iota_{U_1;U_2|U_0}(U_1; U_2|U_0) \end{bmatrix}.$$

To demonstrate the use of the generalized Poisson matching lemma in place of the mutual covering lemma, we prove a one-shot version of Marton's inner bound without common message [13] (i.e., $\mathsf{L}_0 = 1$). Our bound is stronger than that in [3] in the sense that our bound implies [3] (with a slight penalty of having $2^{1-\gamma} + 2^{-2\gamma}$ instead of $2^{1-\gamma} + e^{-2^\gamma}$), but [3] does not imply our bound. We also note that a finite-blocklength bound is given in [7]. Nevertheless, the analysis in [7] only works for discrete auxiliary random variables U_1, U_2, and does not appear to yield a one-shot bound due to the use of typical sequences.

In the conventional mutual covering approach in [5], [4], sub-codebooks for both U_1 and U_2 are generated, whereas in our approach we generate a sub-codebook only for U_1, and the codebook of U_2 adapts to the sub-codebook automatically, eliminating the need for a sub-codebook for U_2.

Theorem 6. *Fix any P_{U_1,U_2} and function $x : \mathcal{U}_1 \times \mathcal{U}_2 \to \mathcal{X}$. For any $\mathsf{J} \in \mathbb{N}$, there exists a code for the broadcast channel $P_{Y_1,Y_2|X}$ for independent private messages $M_j \sim \mathrm{Unif}[1 : \mathsf{L}_j]$, $j = 1, 2$, with the error probability bounded by*

$$P_e \le \mathbf{E}\left[\min\left\{\mathsf{L}_1 \mathsf{J} 2^{-\iota_{U_1;Y_1}(U_1;Y_1)} + \mathsf{L}_2(1 - \mathsf{J}^{-1}) 2^{-\iota_{U_2;Y_2}(U_2;Y_2)} + \mathsf{L}_2 \mathsf{J}^{-1} 2^{\iota_{U_1;U_2}(U_1;U_2) - \iota_{U_2;Y_2}(U_2;Y_2)},\, 1\right\}\right]$$

if all the information density terms are defined, where $(U_1, U_2, X, Y_1, Y_2) \sim P_{U_1 U_2} \delta_{x(U_1,U_2)} P_{Y_1,Y_2|X}$.

Proof: Let $\{(\bar{U}_{1,i}, \bar{M}_{1,i}), T_{1,i}\}_{i\in\mathbb{N}}$, $\{(\bar{U}_{2,i}, \bar{M}_{2,i}), T_{2,i}\}_{i\in\mathbb{N}}$ be two independent Poisson processes with intensity measures $P_{U_1} \times P_{M_1} \times \lambda_{\mathbb{R}_{\ge 0}}$ and $P_{U_2} \times P_{M_2} \times \lambda_{\mathbb{R}_{\ge 0}}$ respectively, independent of M_1, M_2.

The encoder would generate X such that

$$(M_1, M_2, K, \{\check{U}_{1j}\}_{j\in[1:\mathsf{J}]}, U_1, U_2, X) \sim P_{M_1} \times P_{M_2} \times P_K P_{U_1}^{\otimes \mathsf{J}} \delta_{\check{U}_{1K}} P_{U_2|U_1} \delta_{x(U_1,U_2)}, \tag{7}$$

where $P_K = \mathrm{Unif}[1 : \mathsf{J}]$, and $\{\check{U}_{1j}\}_{j\in[1:\mathsf{J}]} \in \mathcal{U}_1^{\mathsf{J}}$ is an intermediate list (which can be regarded as a sub-codebook). The term $P_{U_1}^{\otimes \mathsf{J}} \delta_{\check{U}_{1K}}$ in (7) means that $\{\check{U}_{1j}\}_j$ are i.i.d. P_{U_1}, and $U_1 = \check{U}_{1K}$. To accomplish this, the encoder computes $\check{U}_{1j} = (\tilde{U}_1)_{P_{U_1} \times \delta_{M_1}}(j)$ for $j = 1, \ldots, \mathsf{J}$ (which Poisson process we are referring to can be deduced from whether we are discussing U_1 or U_2), $U_2 = (\tilde{U}_2)_{\mathsf{J}^{-1} \sum_{j=1}^{\mathsf{J}} P_{U_2|U_1}(\cdot|\check{U}_{1j}) \times \delta_{M_2}}$, and $(K, U_1)|(\{\check{U}_{1j}\}_j, U_2) \sim P_{K,U_1|\{\check{U}_{1j}\}_j,U_2}$ (where $P_{K,U_1|\{\check{U}_{1j}\}_j,U_2}$ is derived from (7)), and outputs $X = x(U_1, U_2)$. It can be verified that (7) is satisfied.

The decoding functions are $\hat{M}_1 = (\tilde{M}_1)_{P_{U_1|Y_1}(\cdot|Y_1) \times P_{M_1}}$, $\hat{M}_2 = (\tilde{M}_2)_{P_{U_2|Y_2}(\cdot|Y_2) \times P_{M_2}}$. We have the following almost surely:

$$\begin{aligned} &\mathbf{P}\left\{(\tilde{U}_1, \tilde{M}_1)_{P_{U_1|Y_1}(\cdot|Y_1) \times P_{M_1}} \neq (U_1, M_1) \,\middle|\, U_1, U_2, Y_1, Y_2, M_1, K\right\} \\ &\stackrel{(a)}{=} \mathbf{P}\left\{(\tilde{U}_1, \tilde{M}_1)_{P_{U_1|Y_1}(\cdot|Y_1) \times P_{M_1}} \neq (U_1, M_1) \,\middle|\, U_1, Y_1, M_1, K\right\} \\ &\stackrel{(b)}{\le} K \frac{dP_{U_1} \times \delta_{M_1}}{dP_{U_1|Y_1}(\cdot|Y_1) \times P_{M_1}}(U_1, M_1) \end{aligned}$$

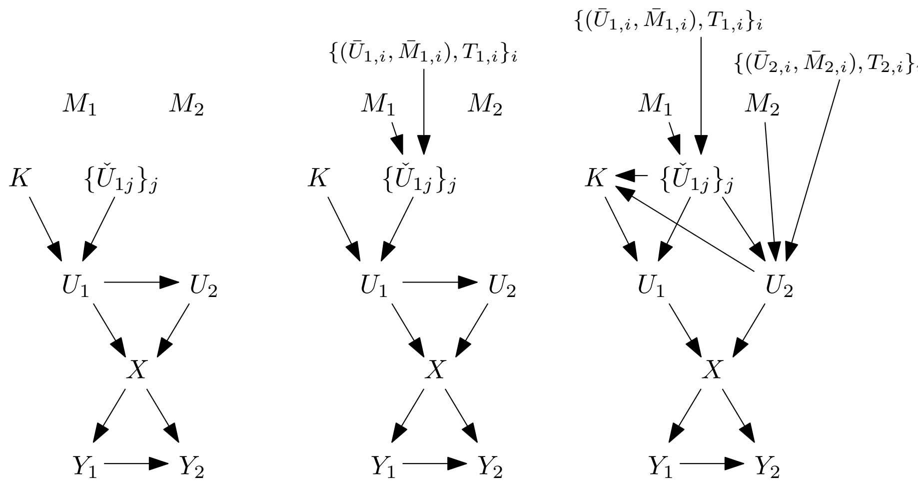

Figure 1. Left: The Bayesian network described in (7). Middle: The Bayesian network deduced from (7) and $\check{U}_{1j} = (\tilde{U}_1)_{P_{U_1} \times \delta_{M_1}}(j)$. Right: The Bayesian network describing the encoding scheme. Note that all three are valid Bayesian networks, and the desired conditional independence relations can be deduced using d-separation.

$$\leq \mathsf{L}_1 \mathsf{J} 2^{-\iota_{U_1;Y_1}(U_1;Y_1)},$$

where (a) is by $(U_2, Y_2) \leftrightarrow (U_1, Y_1, M_1, K) \leftrightarrow \{(\bar{U}_{1,i}, \bar{M}_{1,i}), T_{1,i}\}_i$ (see Figure 1 middle), and (b) is by the conditional generalized Poisson matching lemma on $(X, J, U, Y, Q_{U|Y}) \leftarrow (M_1,\, K,\, (U_1, M_1),\, Y_1,\, P_{U_1|Y_1} \times P_{M_1})$, since $P_{U_1, M_1 | M_1, K} = P_{U_1} \times \delta_{M_1}$, $(M_1, K) \perp\!\!\!\perp \{(\bar{U}_{1,i}, \bar{M}_{1,i}), T_{1,i}\}_i$, and $Y_1 \leftrightarrow (U_1, M_1, K) \leftrightarrow \{(\bar{U}_{1,i}, \bar{M}_{1,i}), T_{1,i}\}_i$, which can be deduced from (7) and $\check{U}_{1j} = (\tilde{U}_1)_{P_{U_1} \times \delta_{M_1}}(j)$ (see Figure 1 middle).

Also, almost surely,

$$\begin{aligned}
&\mathbf{P}\left\{ (\tilde{U}_2, \tilde{M}_2)_{P_{U_2|Y_2}(\cdot|Y_2) \times P_{M_2}} \neq (U_2, M_2) \,\middle|\, U_1, U_2, Y_1, Y_2, M_2 \right\} \\
&\overset{(a)}{=} \mathbf{P}\left\{ (\tilde{U}_2, \tilde{M}_2)_{P_{U_2|Y_2}(\cdot|Y_2) \times P_{M_2}} \neq (U_2, M_2) \,\middle|\, U_1, U_2, Y_2, M_2 \right\} \\
&\overset{(b)}{\leq} \mathbf{E}\left[\frac{d(\mathsf{J}^{-1} \sum_{j=1}^{\mathsf{J}} P_{U_2|U_1}(\cdot|\check{U}_{1j})) \times \delta_{M_2}}{dP_{U_2|Y_2}(\cdot|Y_2) \times P_{M_2}}(U_2, M_2) \,\middle|\, U_1, U_2, Y_2, M_2 \right] \\
&= \mathbf{E}\left[\mathsf{L}_2 \mathsf{J}^{-1} \sum_{j=1}^{\mathsf{J}} 2^{\iota_{U_1;U_2}(\check{U}_{1j};U_2) - \iota_{U_2;Y_2}(U_2;Y_2)} \,\middle|\, U_1, U_2, Y_2, M_2 \right] \\
&= \mathbf{E}\left[\mathsf{L}_2 \mathsf{J}^{-1} 2^{-\iota_{U_2;Y_2}(U_2;Y_2)} \left(2^{\iota_{U_1;U_2}(U_1;U_2)} + \sum_{j \in [1:\mathsf{J}] \setminus K} 2^{\iota_{U_1;U_2}(\check{U}_{1j};U_2)} \right) \,\middle|\, U_1, U_2, Y_2, M_2 \right] \\
&= \mathbf{E}\left[\mathsf{L}_2 \mathsf{J}^{-1} 2^{-\iota_{U_2;Y_2}(U_2;Y_2)} \left(2^{\iota_{U_1;U_2}(U_1;U_2)} + \sum_{j=1}^{\mathsf{J}-1} 2^{\iota_{U_1;U_2}(\check{U}_{1,j+\mathbf{1}\{j \geq K\}};U_2)} \right) \,\middle|\, U_1, U_2, Y_2, M_2 \right] \\
&\overset{(c)}{\leq} \mathsf{L}_2 \mathsf{J}^{-1} 2^{-\iota_{U_2;Y_2}(U_2;Y_2)} (2^{\iota_{U_1;U_2}(U_1;U_2)} + \mathsf{J} - 1),
\end{aligned}$$

where (a) is by $Y_1 \leftrightarrow (U_1, U_2, Y_2, M_2) \leftrightarrow \{(\bar{U}_{2,i}, \bar{M}_{2,i}), T_{2,i}\}_i$ (see Figure 1 right), (b) is by the conditional Poisson matching lemma on $((\{\check{U}_{1j}\}_j, M_2),\, (U_2, M_2),\, Y_2,\, P_{U_2|Y_2} \times P_{M_2})$, and (c) is because $\{\check{U}_{1,j+\mathbf{1}\{j \geq K\}}\}_{j \in [1:\mathsf{J}-1]}$ (the $\check{U}_{1j}$'s not selected as U_1) are independent of (U_1, U_2, Y_2, M_2), $\mathbf{E}[2^{\iota_{U_1;U_2}(\check{U}_{1,j+\mathbf{1}\{j \geq K\}};U_2)} \,|\, U_2] = 1$, and Jensen's inequality. Hence,

$$\begin{aligned}
&\mathbf{P}\{(M_1, M_2) \neq (\hat{M}_1, \hat{M}_2)\} \\
&= \mathbf{E}\left[\mathbf{P}\left\{ (M_1, M_2) \neq (\hat{M}_1, \hat{M}_2) \,\middle|\, U_1, U_2, Y_1, Y_2 \right\} \right] \\
&\leq \mathbf{E}\left[\min\left\{ \mathbf{P}\left\{ M_1 \neq \hat{M}_1 \,\middle|\, U_1, U_2, Y_1, Y_2 \right\} + \mathbf{P}\left\{ M_2 \neq \hat{M}_2 \,\middle|\, U_1, U_2, Y_1, Y_2 \right\},\, 1 \right\} \right] \\
&\leq \mathbf{E}\left[\min\left\{ \mathsf{L}_1 \mathsf{J} 2^{-\iota_{U_1;Y_1}(U_1;Y_1)} + \mathsf{L}_2 \mathsf{J}^{-1} 2^{-\iota_{U_2;Y_2}(U_2;Y_2)} (2^{\iota_{U_1;U_2}(U_1;U_2)} + \mathsf{J} - 1),\, 1 \right\} \right].
\end{aligned}$$

Therefore there exist fixed realizations of the Poisson processes attaining the desired bound. ■

IX. One-shot Distributed Lossy Source Coding

The one-shot distributed lossy source coding setting is described as follows. Let $(X_1, X_2) \sim P_{X_1,X_2}$. Upon observing X_j, encoder j produces $M_j \in [1 : \mathsf{L}_j]$, $j = 1, 2$. The decoder observes M_1, M_2 and recovers $\hat{Z}_1 \in \mathcal{Z}_1$, $\hat{Z}_2 \in \mathcal{Z}_2$ with probability of excess distortion $P_e = \mathbf{P}\{\mathsf{d}_1(X_1, \hat{Z}_1) > \mathsf{D}_1 \text{ or } \mathsf{d}_2(X_2, \hat{Z}_2) > \mathsf{D}_2\}$, where $\mathsf{d}_j : \mathcal{X}_j \times \mathcal{Z}_j \to \mathbb{R}_{\ge 0}$ is a distortion measure for $j = 1, 2$.

We show a one-shot version of the Berger-Tung inner bound [11], [12].

Theorem 7. *Fix any $P_{U_1|X_1}$, $P_{U_2|X_2}$ and functions $z_j : \mathcal{U}_1 \times \mathcal{U}_2 \to \mathcal{Z}_j$, $j = 1, 2$. There exists a code for distributed lossy source coding with sources P_{X_1}, P_{X_2} and message sizes $\mathsf{L}_1, \mathsf{L}_2$, with probability of excess distortion*

$$P_e \le \mathbf{E}\bigg[\min\bigg\{\mathbf{1}\{\mathsf{d}_1(X_1, Z_1) > \mathsf{D}_1 \text{ or } \mathsf{d}_2(X_2, Z_2) > \mathsf{D}_2\} + \mathsf{L}_1^{-1} 2^{\iota_{U_1;X_1|U_2}(U_1;X_1|U_2)} + \Big(\mathsf{L}_1^{-1}\mathsf{L}_2^{-1} 2^{\iota_{U_1,U_2;X_1,X_2}(U_1,U_2;X_1,X_2)} + \mathsf{L}_2^{-1} 2^{\iota_{U_2;X_2|U_1}(U_2;X_2|U_1)}\Big)\Big(\log(\mathsf{L}_2 2^{-\iota_{U_2;X_2|U_1}(U_2;X_2|U_1)} + 1) + 1\Big)^2,\, 1\bigg\}\bigg] \quad (8)$$

if all the information density terms are defined, where $(X_1, X_2, U_1, U_2, Z_1, Z_2) \sim P_{X_1,X_2} P_{U_1|X_1} P_{U_2|X_2} \delta_{z_1(U_1,U_2)} \delta_{z_2(U_1,U_2)}$. As a result, for $\gamma > 0$,

$$P_e \le \mathbf{P}\Big\{\mathsf{d}_1(X_1, Z_1) > \mathsf{D}_1 \text{ or } \mathsf{d}_2(X_2, Z_2) > \mathsf{D}_2 \text{ or } \log \mathsf{L}_1 < \iota_{U_1;X_1|U_2}(U_1; X_1|U_2) + \gamma \text{ or } \log \mathsf{L}_2 < \iota_{U_2;X_2|U_1}(U_2; X_2|U_1) + \gamma \text{ or } \log \mathsf{L}_1\mathsf{L}_2 < \iota_{U_1,U_2;X_1,X_2}(U_1, U_2; X_1, X_2) + \gamma\Big\} + 2^{-\gamma}\left(4\mathbf{E}[(\iota_{U_1;U_2}(U_1; U_2))^2] + 4\gamma^2 + 29\right). \quad (9)$$

The logarithmic term in (8) (or the last term in (9)) results in an $O(n^{-1}\log n)$ penalty on the rate in the finite blocklength regime, and does not affect the second order result. Ignoring the last term in (9), the error event in (9) is a strict subset of that in [5, eqn (47)]. This is because the error event in [5] is a superset of (9) by Fourier-Motzkin elimination on $\mathsf{J}_1, \mathsf{J}_2$ in the error event in [5], but the reverse is not true since Fourier-Motzkin elimination only guarantees the existence of random variables for $\mathsf{J}_1, \mathsf{J}_2$ (that depend on the information density terms) satisfying the bounds, but $\mathsf{J}_1, \mathsf{J}_2$ must be constants since they are parameters of the code construction in [5].

We now prove the result. Unlike previous approaches, our proof does not require binning. The encoders are the same as those for point-to-point lossy source coding.

Proof: Let $\{(\bar{U}_{1,i}, \bar{M}_{1,i}), T_{1,i}\}_{i\in\mathbb{N}}$, $\{(\bar{U}_{2,i}, \bar{M}_{2,i}), T_{2,i}\}_{i\in\mathbb{N}}$ be two independent Poisson processes with intensity measures $P_{U_1} \times P_{M_1} \times \lambda_{\mathbb{R}_{\ge 0}}$ and $P_{U_2} \times P_{M_2} \times \lambda_{\mathbb{R}_{\ge 0}}$ respectively, independent of X_1, X_2. The encoding functions are $M_j = (\tilde{M}_j)_{P_{U_j|X_j}(\cdot|X_j)\times P_{M_j}}$, $j = 1, 2$ (which Poisson process we are referring to can be deduced from whether we are discussing M_1 or M_2). Also define $U_j = (\tilde{U}_j)_{P_{U_j|X_j}(\cdot|X_j)\times P_{M_j}}$, $Z_j = z_j(U_1, U_2)$, $j = 1, 2$. For the decoding function, let $\check{U}_{1k} = (\tilde{U}_1)_{P_{U_1}\times\delta_{M_1}}(k)$ for $k \in \mathbb{N}$, $\hat{U}_2 = (\tilde{U}_2)_{\sum_{k=1}^{\infty}\phi(k)P_{U_2|U_1}(\cdot|\check{U}_{1k})\times\delta_{M_2}}$ where $\phi(k) \propto k^{-1}(\log(k + 2))^{-2}$ with $\sum_{k=1}^{\infty}\phi(k) = 1$, and $\hat{U}_1 = (\tilde{U}_1)_{P_{U_1|U_2}(\cdot|\hat{U}_2)\times\delta_{M_1}}$, $\hat{Z}_j = z_j(\hat{U}_1, \hat{U}_2)$, $j = 1, 2$. Note that $(M_1, M_2, X_1, X_2, U_1, U_2, Z_1, Z_2) \sim P_{M_1} \times P_{M_2} \times P_{X_1,X_2} P_{U_1|X_1} P_{U_2|X_2} \delta_{z_1(U_1,U_2)} \delta_{z_2(U_1,U_2)}$.

Let $K = \Upsilon_{P_{U_1|X_1}(\cdot|X_1)\times P_{M_1}\|P_{U_1}\times\delta_{M_1}}(1)$ (using the Poisson process $\{(\bar{U}_{1,i}, \bar{M}_{1,i}), T_{1,i}\}_{i\in\mathbb{N}}$). By the conditional generalized Poisson matching lemma on $(X_1,\, 1,\, (U_1, M_1),\, M_1,\, P_{U_1} \times \delta_{M_1})$ (note that $P_{U_1,M_1|X_1} = P_{U_1|X_1} \times P_{M_1}$), almost surely,

$$\begin{aligned}\mathbf{E}\left[K \mid X_1, U_1, M_1\right] &\le \frac{dP_{U_1|X_1}(\cdot|X_1) \times P_{M_1}}{dP_{U_1} \times \delta_{M_1}}(U_1, M_1) + 1 \\ &= \mathsf{L}_1^{-1} 2^{\iota_{U_1;X_1}(U_1;X_1)} + 1. \end{aligned} \quad (10)$$

Since $\{\check{U}_{1k}\}_k$ is a function of $\{(\bar{U}_{1,i}, \bar{M}_{1,i}), T_{1,i}\}_i$ and M_1, we have $\{\check{U}_{1k}\}_k \leftrightarrow (X_1, X_2, U_1, U_2, M_2) \leftrightarrow \{(\bar{U}_{2,i}, \bar{M}_{2,i}), T_{2,i}\}_i$. By the conditional Poisson matching lemma on $(X_2,\, (U_2, M_2),\, (\{\check{U}_{1k}\}_k, M_2),\, \sum_{k=1}^{\infty}\phi(k)P_{U_2|U_1}(\cdot|\check{U}_{1k}) \times \delta_{M_2})$ (note that $P_{U_2,M_2|X_2} = P_{U_2|X_2} \times P_{M_2}$), almost surely,

$$\begin{aligned}&\mathbf{P}\Big\{(\tilde{U}_2, \tilde{M}_2)_{\sum_{k=1}^{\infty}\phi(k)P_{U_2|U_1}(\cdot|\check{U}_{1k})\times\delta_{M_2}} \ne (U_2, M_2)\,\Big|\, X_1, X_2, U_1, U_2, M_2\Big\} \\ &\le \mathbf{E}\left[\min\left\{\frac{dP_{U_2|X_2}(\cdot|X_2) \times P_{M_2}}{d(\sum_{k=1}^{\infty}\phi(k)P_{U_2|U_1}(\cdot|\check{U}_{1k})) \times \delta_{M_2}}(U_2, M_2),\, 1\right\}\,\middle|\, X_1, X_2, U_1, U_2, M_2\right] \\ &\le \mathbf{E}\left[\min\left\{\mathsf{L}_2^{-1}\frac{dP_{U_2|X_2}(\cdot|X_2)}{\phi(K)dP_{U_2|U_1}(\cdot|U_1)}(U_2),\, 1\right\}\,\middle|\, X_1, X_2, U_1, U_2, M_2\right] \\ &= \mathbf{E}\left[\min\{\mathsf{L}_2^{-1}(\phi(K))^{-1}2^{\iota_{U_2;X_2|U_1}(U_2;X_2|U_1)},\, 1\}\,\middle|\, X_1, X_2, U_1, U_2, M_2\right]\end{aligned}$$

$$
\begin{aligned}
&\overset{(a)}{\leq} \mathbf{E}\left[K\mathsf{L}_2^{-1}2^{\iota_{U_2;X_2|U_1}(U_2;X_2|U_1)}\left(\log(\mathsf{L}_2 2^{-\iota_{U_2;X_2|U_1}(U_2;X_2|U_1)}+1)+1\right)^2 \,\middle|\, X_1,X_2,U_1,U_2,M_2\right]\\
&\overset{(b)}{\leq} \left(\mathsf{L}_1^{-1}2^{\iota_{U_1;X_1}(U_1;X_1)}+1\right)\mathsf{L}_2^{-1}2^{\iota_{U_2;X_2|U_1}(U_2;X_2|U_1)}\left(\log(\mathsf{L}_2 2^{-\iota_{U_2;X_2|U_1}(U_2;X_2|U_1)}+1)+1\right)^2\\
&= \left(\mathsf{L}_1^{-1}\mathsf{L}_2^{-1}2^{\iota_{U_1,U_2;X_1,X_2}(U_1,U_2;X_1,X_2)}+\mathsf{L}_2^{-1}2^{\iota_{U_2;X_2|U_1}(U_2;X_2|U_1)}\right)\left(\log(\mathsf{L}_2 2^{-\iota_{U_2;X_2|U_1}(U_2;X_2|U_1)}+1)+1\right)^2,
\end{aligned}
$$

where (a) is by Proposition 6, and (b) is by $K\leftrightarrow(U_1,X_1)\leftrightarrow(X_2,U_2,M_2)$, (10) and Jensen's inequality. By the conditional Poisson matching lemma on $(X_1,\,(U_1,M_1),\,(U_2,M_1),\,P_{U_1|U_2}\times\delta_{M_1})$ (note that $P_{U_1,M_1|X_1}=P_{U_1|X_1}\times P_{M_1}$), and $X_2\leftrightarrow(X_1,U_1,U_2,M_1)\leftrightarrow\{(\bar{U}_{1,i},\bar{M}_{1,i}),T_{1,i}\}_i$, almost surely,

$$
\begin{aligned}
&\mathbf{P}\left\{(\tilde{U}_1,\tilde{M}_1)_{P_{U_1|U_2}(\cdot|U_2)\times\delta_{M_1}}\neq(U_1,M_1)\,\middle|\,X_1,X_2,U_1,U_2,M_1\right\}\\
&\leq \frac{dP_{U_1|X_1}(\cdot|X_1)\times P_{M_1}}{dP_{U_1|U_2}(\cdot|U_2)\times\delta_{M_1}}(U_1,M_1)\\
&= \mathsf{L}_1^{-1}2^{\iota_{U_1;X_1|U_2}(U_1;X_1|U_2)}.
\end{aligned}
$$

We have

$$
\begin{aligned}
&\mathbf{P}\{\mathsf{d}_1(X_1,\hat{Z}_1)>\mathsf{D}_1 \text{ or } \mathsf{d}_2(X_2,\hat{Z}_2)>\mathsf{D}_2\}\\
&\leq \mathbf{E}\Bigg[\mathbf{P}\bigg\{\mathsf{d}_1(X_1,Z_1)>\mathsf{D}_1 \text{ or } \mathsf{d}_2(X_2,Z_2)>\mathsf{D}_2 \text{ or } \hat{U}_2\neq U_2\\
&\quad \text{or } (\hat{U}_2=U_2 \text{ and } \hat{U}_1\neq U_1)\,\bigg|\,X_1,X_2,U_1,U_2\bigg\}\Bigg]\\
&\leq \mathbf{E}\Bigg[\min\bigg\{\mathbf{1}\{\mathsf{d}_1(X_1,Z_1)>\mathsf{D}_1 \text{ or } \mathsf{d}_2(X_2,Z_2)>\mathsf{D}_2\}+\mathsf{L}_1^{-1}2^{\iota_{U_1;X_1|U_2}(U_1;X_1|U_2)}\\
&\quad +\left(\mathsf{L}_1^{-1}\mathsf{L}_2^{-1}2^{\iota_{U_1,U_2;X_1,X_2}(U_1,U_2;X_1,X_2)}+\mathsf{L}_2^{-1}2^{\iota_{U_2;X_2|U_1}(U_2;X_2|U_1)}\right)\left(\log(\mathsf{L}_2 2^{-\iota_{U_2;X_2|U_1}(U_2;X_2|U_1)}+1)+1\right)^2,\,1\bigg\}\Bigg]
\end{aligned}
$$

Therefore there exist fixed values of the Poisson processes attaining the desired bound.

For (9), if the event in (9) does not occur, by Proposition 6 with $\alpha=\gamma-1$, $\tilde{\alpha}=\gamma$, $\beta=\iota_{U_1;U_2}(U_1;U_2)-\gamma$,

$$
\begin{aligned}
&\mathsf{L}_1^{-1}2^{\iota_{U_1;X_1|U_2}(U_1;X_1|U_2)}\\
&\quad+\left(\mathsf{L}_1^{-1}\mathsf{L}_2^{-1}2^{\iota_{U_1,U_2;X_1,X_2}(U_1,U_2;X_1,X_2)}+\mathsf{L}_2^{-1}2^{\iota_{U_2;X_2|U_1}(U_2;X_2|U_1)}\right)\left(\log(\mathsf{L}_2 2^{-\iota_{U_2;X_2|U_1}(U_2;X_2|U_1)}+1)+1\right)^2\\
&\leq 2^{-\gamma}+2^{1-\gamma}\left(2(\iota_{U_1;U_2}(U_1;U_2))^2+2\gamma^2+14\right)\\
&= 2^{-\gamma}\left(4(\iota_{U_1;U_2}(U_1;U_2))^2+4\gamma^2+29\right).
\end{aligned}
$$

■

Remark 3. The reason for the logarithmic term is that we want to translate a bound on $\mathbf{E}[K]$ (given by the generalized Poisson matching lemma) into a bound on $\mathbf{E}[(\phi(K))^{-1}]$ for some distribution ϕ over $\mathbb{N}$. Ideally, we wish $(\phi(k))^{-1}\propto k$, but this is impossible since the harmonic series diverges. Therefore we use a slow converging series $\phi(k)\propto k^{-1}(\log(k+2))^{-2}$ instead, resulting in a logarithmic penalty.

If we use $\mathsf{J}^{-1}\mathbf{1}\{k\leq\mathsf{J}\}$ instead of $\phi(k)$ in the proof, we can obtain the following bound for any $\mathsf{J}\in\mathbb{N}$:

$$
\begin{aligned}
P_e \leq \mathbf{E}\Bigg[\min\bigg\{&\mathbf{1}\{\mathsf{d}_1(X_1,Z_1)>\mathsf{D}_1 \text{ or } \mathsf{d}_2(X_2,Z_2)>\mathsf{D}_2\}\\
&+\mathsf{L}_1^{-1}\mathsf{J}^{-1}2^{\iota_{U_1;X_1}(U_1;X_1)}+\mathsf{L}_2^{-1}\mathsf{J}2^{\iota_{U_2;X_2|U_1}(U_2;X_2|U_1)}+\mathsf{L}_1^{-1}2^{\iota_{U_1;X_1|U_2}(U_1;X_1|U_2)},\,1\bigg\}\Bigg].
\end{aligned}
$$

Compared to Theorem 7, this does not contain the logarithmic term, but requires optimizing over J, and may give a worse second order result.

Another choice is to use $g(k)\propto k^{-1}\mathbf{1}\{k\leq\mathsf{J}\}$ instead of $\phi(k)$. We can obtain the following bound for any $\mathsf{J}\in\mathbb{N}$:

$$
\begin{aligned}
P_e \leq \mathbf{E}\Bigg[\min\bigg\{&\mathbf{1}\{\mathsf{d}_1(X_1,Z_1)>\mathsf{D}_1 \text{ or } \mathsf{d}_2(X_2,Z_2)>\mathsf{D}_2\}+\mathsf{L}_1^{-1}\mathsf{J}^{-1}2^{\iota_{U_1;X_1}(U_1;X_1)}\\
&+\mathsf{L}_1^{-1}\mathsf{L}_2^{-1}(\ln\mathsf{J}+1)2^{\iota_{U_1,U_2;X_1,X_2}(U_1,U_2;X_1,X_2)}+\mathsf{L}_2^{-1}(\ln\mathsf{J}+1)2^{\iota_{U_2;X_2|U_1}(U_2;X_2|U_1)}+\mathsf{L}_1^{-1}2^{\iota_{U_1;X_1|U_2}(U_1;X_1|U_2)},\,1\bigg\}\Bigg].
\end{aligned}
$$

which gives the same second order result as Theorem 7. Nevertheless, we prefer using $\phi(k)$ which eliminates the need for a parameter J at the decoder.

X. One-shot Coding for Multiple Access Channels

The one-shot coding setting for the multiple access channel is described as follows. Upon observing $M_j \sim \mathrm{Unif}[1:\mathsf{L}_j]$ (M_1, M_2 independent), encoder j produces X_j, $j=1,2$. The decoder observes the output Y of the channel $P_{Y|X_1,X_2}$ and recovers $(\hat{M}_1, \hat{M}_2)$. The error probability is $P_e = \mathbf{P}\{(M_1,M_2) \neq (\hat{M}_1,\hat{M}_2)\}$.

We present a one-shot achievability result for the capacity region in [27], [28], [29]. While this result is slightly weaker than that in [3], we include it to illustrate the use of the generalized Poisson matching lemma in simultaneous decoding. Note that the logarithmic term results in an $O(n^{-1}\log n)$ penalty on the rate in the finite blocklength regime, and does not affect the second order result.

Theorem 8. *Fix any P_{X_1}, P_{X_2}. There exists a code for the multiple access channel $P_{Y|X_1,X_2}$ for messages $M_j \sim \mathrm{Unif}[1:\mathsf{L}_j]$, $j=1,2$, with the error probability bounded by*

$$P_e \le \mathbf{E}\Big[\min\Big\{\Big(\mathsf{L}_1\mathsf{L}_2 2^{-\iota_{X_1,X_2;Y}(X_1,X_2;Y)} + \mathsf{L}_2 2^{-\iota_{X_2;X_1,Y}(X_2;X_1,Y)}\Big)\Big(\log(\mathsf{L}_2^{-1}2^{\iota_{X_2;X_1,Y}(X_2;X_1,Y)}+1)+1\Big)^2 + \mathsf{L}_1 2^{-\iota_{X_1;X_2,Y}(X_1;X_2,Y)},\, 1\Big\}\Big]$$

if $P_{X_1X_2Y} \ll P_{X_1}\times P_{X_2}\times P_Y$, where $(X_1,X_2,Y)\sim P_{X_1}P_{X_2}P_{Y|X_1,X_2}$. As a result, for $\gamma>0$,

$$P_e \le \mathbf{P}\Big\{\log \mathsf{L}_1 > \iota_{X_1;X_2,Y}(X_1;X_2,Y)-\gamma \;\text{ or }\; \log\mathsf{L}_2 > \iota_{X_2;X_1,Y}(X_2;X_1,Y)-\gamma \;\text{ or }\; \log\mathsf{L}_1\mathsf{L}_2 > \iota_{X_1,X_2;Y}(X_1,X_2;Y)-\gamma\Big\} + 2^{-\gamma}\left(4\mathbf{E}[(\iota_{X_1;X_2|Y}(X_1;X_2|Y))^2] + 4\gamma^2 + 29\right). \tag{11}$$

Proof: Let $\{(\bar{X}_{1,i},\bar{M}_{1,i}),T_{1,i}\}_{i\in\mathbb{N}}$, $\{(\bar{X}_{2,i},\bar{M}_{2,i}),T_{2,i}\}_{i\in\mathbb{N}}$ be two independent Poisson processes with intensity measures $P_{X_1}\times P_{M_1}\times\lambda_{\mathbb{R}_{\ge0}}$ and $P_{X_2}\times P_{M_2}\times\lambda_{\mathbb{R}_{\ge0}}$ respectively, independent of M_1, M_2. The encoding functions are $X_1 = (\tilde{X}_1)_{P_{X_1}\times\delta_{M_1}}$, $X_2 = (\tilde{X}_2)_{P_{X_2}\times\delta_{M_2}}$ (which Poisson process we are referring to can be deduced from whether we are discussing X_1 or X_2). For the decoding function, let $\check{X}_{1k} = (\tilde{X}_1)_{P_{X_1|Y}(\cdot|Y)\times P_{M_1}}(k)$ for $k\in\mathbb{N}$, $(\hat{X}_2,\hat{M}_2) = (\tilde{X}_2,\tilde{M}_2)_{\sum_{k=1}^\infty \phi(k)P_{X_2|X_1,Y}(\cdot|\check{X}_{1k},Y)\times P_{M_2}}$ where $\phi(k)\propto k^{-1}(\log(k+2))^{-2}$ with $\sum_{k=1}^\infty\phi(k)=1$, and $\hat{M}_1 = (\tilde{M}_1)_{P_{X_1|X_2,Y}(\cdot|\hat{X}_2,Y)\times P_{M_1}}$.

Let $K = \Upsilon_{P_{X_1}\times\delta_{M_1}\|P_{X_1|Y}(\cdot|Y)\times P_{M_1}}(1)$ (using the Poisson process $\{(\bar{X}_{1,i},\bar{M}_{1,i}),T_{1,i}\}_{i\in\mathbb{N}}$). By the conditional generalized Poisson matching lemma on $(M_1,\, 1,\, (X_1,M_1),\, Y,\, P_{X_1|Y}\times P_{M_1})$ (note that $P_{X_1,M_1|M_1} = P_{X_1}\times\delta_{M_1}$), almost surely,

$$\begin{aligned}\mathbf{E}[K \mid X_1,Y,M_1] &\le \frac{dP_{X_1}\times\delta_{M_1}}{dP_{X_1|Y}(\cdot|Y)\times P_{M_1}}(X_1,M_1)+1\\ &= \mathsf{L}_1 2^{-\iota_{X_1;Y}(X_1;Y)}+1.\end{aligned}\tag{12}$$

Since $\{\check{X}_{1k}\}_k$ is a function of $\{(\bar{X}_{1,i},\bar{M}_{1,i}),T_{1,i}\}_i$ and Y, we have $\{\check{X}_{1k}\}_k \leftrightarrow (X_1,X_2,Y,M_2)\leftrightarrow\{(\bar{X}_{2,i},\bar{M}_{2,i}),T_{2,i}\}_i$. By the conditional Poisson matching lemma on $(M_2,\, (X_2,M_2),\, (\{\check{X}_{1k}\}_k, Y),\, \sum_{k=1}^\infty\phi(k)P_{X_2|X_1,Y}(\cdot|\check{X}_{1k},Y)\times P_{M_2})$ (note that $P_{X_2,M_2|M_2} = P_{X_2}\times\delta_{M_2}$), almost surely,

$$\begin{aligned}&\mathbf{P}\Big\{(\tilde{X}_2,\tilde{M}_2)_{\sum_{k=1}^\infty\phi(k)P_{X_2|X_1,Y}(\cdot|\check{X}_{1k},Y)\times P_{M_2}} \neq (X_2,M_2)\,\Big|\, X_1,X_2,Y,M_2\Big\}\\
&\le \mathbf{E}\left[\min\left\{\frac{dP_{X_2}\times\delta_{M_2}}{d(\sum_{k=1}^\infty\phi(k)P_{X_2|X_1,Y}(\cdot|\check{X}_{1k},Y))\times P_{M_2}}(X_2,M_2),\,1\right\}\,\middle|\, X_1,X_2,Y,M_2\right]\\
&\le \mathbf{E}\left[\min\left\{\mathsf{L}_2\frac{dP_{X_2}}{\phi(K)dP_{X_2|X_1,Y}(\cdot|X_1,Y)}(X_2),\,1\right\}\,\middle|\, X_1,X_2,Y,M_2\right]\\
&= \mathbf{E}\left[\min\{\mathsf{L}_2(\phi(K))^{-1}2^{-\iota_{X_2;X_1,Y}(X_2;X_1,Y)},\,1\}\,\middle|\, X_1,X_2,Y,M_2\right]\\
&\overset{(a)}{\le} \mathbf{E}\left[K\mathsf{L}_2 2^{-\iota_{X_2;X_1,Y}(X_2;X_1,Y)}\left(\log(\mathsf{L}_2^{-1}2^{\iota_{X_2;X_1,Y}(X_2;X_1,Y)}+1)+1\right)^2\,\middle|\, X_1,X_2,Y,M_2\right]\\
&\overset{(b)}{\le} \left(\mathsf{L}_1 2^{-\iota_{X_1;Y}(X_1;Y)}+1\right)\mathsf{L}_2 2^{-\iota_{X_2;X_1,Y}(X_2;X_1,Y)}\left(\log(\mathsf{L}_2^{-1}2^{\iota_{X_2;X_1,Y}(X_2;X_1,Y)}+1)+1\right)^2\\
&= \left(\mathsf{L}_1\mathsf{L}_2 2^{-\iota_{X_1,X_2;Y}(X_1,X_2;Y)} + \mathsf{L}_2 2^{-\iota_{X_2;X_1,Y}(X_2;X_1,Y)}\right)\left(\log(\mathsf{L}_2^{-1}2^{\iota_{X_2;X_1,Y}(X_2;X_1,Y)}+1)+1\right)^2,\end{aligned}$$

where (a) is by Proposition 6, and (b) is by $K\leftrightarrow(X_1,Y)\leftrightarrow X_2$, (12) and Jensen's inequality. By the conditional Poisson matching lemma on $(M_1,\, (X_1,M_1),\, (X_2,Y),\, P_{X_1|X_2,Y}\times P_{M_1})$ (note that $P_{X_1,M_1|M_1} = P_{X_1}\times\delta_{M_1}$), almost surely,

$$\mathbf{P}\Big\{(\tilde{X}_1,\tilde{M}_1)_{P_{X_1|X_2,Y}(\cdot|X_2,Y)\times P_{M_1}} \neq (X_1,M_1)\,\Big|\, X_1,X_2,Y,M_1\Big\}$$

$$\leq \frac{dP_{X_1} \times \delta_{M_1}}{dP_{X_1|X_2,Y}(\cdot|X_2,Y) \times P_{M_1}}(X_1, M_1)$$
$$= \mathsf{L}_1 2^{-\iota_{X_1;X_2,Y}(X_1;X_2,Y)}.$$

Therefore there exist fixed values of the Poisson processes attaining the desired bound.

For (11), if the event in (11) does not occur, by Proposition 6 with $\alpha = \gamma - 1$, $\tilde{\alpha} = \gamma$, $\beta = \iota_{X_1;X_2|Y}(X_1;X_2|Y) - \gamma$,

$$\left(\mathsf{L}_1\mathsf{L}_2 2^{-\iota_{X_1,X_2;Y}(X_1,X_2;Y)} + \mathsf{L}_2 2^{-\iota_{X_2;X_1,Y}(X_2;X_1,Y)}\right)\left(\log(\mathsf{L}_2^{-1} 2^{\iota_{X_2;X_1,Y}(X_2;X_1,Y)} + 1) + 1\right)^2 + \mathsf{L}_1 2^{-\iota_{X_1;X_2,Y}(X_1;X_2,Y)}$$
$$\leq 2^{1-\gamma}\left(2(\iota_{X_1;X_2|Y}(X_1;X_2|Y))^2 + 2\gamma^2 + 14\right) + 2^{-\gamma}$$
$$= 2^{-\gamma}\left(4(\iota_{X_1;X_2|Y}(X_1;X_2|Y))^2 + 4\gamma^2 + 29\right).$$

■

Remark 4. If we use $\mathsf{J}^{-1}\mathbf{1}\{k \leq \mathsf{J}\}$ instead of $\phi(k)$ in the proof, we can obtain the following bound for any $\mathsf{J} \in \mathbb{N}$:

$$P_e \leq \mathbf{E}\left[\min\left\{\mathsf{L}_1\mathsf{J}^{-1}2^{-\iota_{X_1;Y}(X_1;Y)} + \mathsf{L}_2\mathsf{J}2^{-\iota_{X_2;X_1,Y}(X_2;X_1,Y)} + \mathsf{L}_1 2^{-\iota_{X_1;X_2,Y}(X_1;X_2,Y)},\, 1\right\}\right].$$

Compared to Theorem 8, this does not contain the logarithmic term, but requires optimizing over J, and may give a worse second order result.

Another choice is to use $g(k) \propto k^{-1}\mathbf{1}\{k \leq \mathsf{J}\}$ instead of $\phi(k)$. We can obtain the following bound for any $\mathsf{J} \in \mathbb{N}$:

$$P_e \leq \mathbf{E}\Big[\min\Big\{\mathsf{L}_1\mathsf{L}_2(\ln \mathsf{J} + 1)2^{-\iota_{X_1,X_2;Y}(X_1,X_2;Y)} + \mathsf{L}_2(\ln \mathsf{J} + 1)2^{-\iota_{X_2;X_1,Y}(X_2;X_1,Y)}$$
$$+ \mathsf{L}_1 2^{-\iota_{X_1;X_2,Y}(X_1;X_2,Y)} + \mathsf{L}_1\mathsf{J}^{-1}2^{-\iota_{X_1;Y}(X_1;Y)},\, 1\Big\}\Big],$$

which gives the same second order result as Theorem 8. Nevertheless, we prefer using $\phi(k)$ which eliminates the need for a parameter J at the decoder.

XI. One-shot Channel Resolvability and Soft Covering

The one-shot channel resolvability setting [30] is described as follows. Fix a channel $P_{Y|X}$ and input distribution P_X. Upon observing an integer $M \sim \mathrm{Unif}[1:\mathsf{L}]$, the encoder applies a deterministic mapping $g : [1:\mathsf{L}] \to \mathcal{X}$ on M to produce $\hat{X} = g(M)$, which is sent through the channel $P_{Y|X}$ and gives the output $\hat{Y}$. The goal is to minimize the total variation distance between $P_{\hat{Y}}$ and P_Y (Y-marginal of $P_X P_{Y|X}$), i.e., $\epsilon := \|\mathsf{L}^{-1}\sum_{m=1}^{\mathsf{L}} P_{Y|X}(\cdot|g(m)) - P_Y(\cdot)\|_{\mathrm{TV}}$.

We show a one-shot channel resolvability result using the the Poisson matching lemma. This result can also be regarded as a one-shot soft covering lemma [31].

Proposition 2. *Given channel $P_{Y|X}$ and input distribution P_X with $P_{XY} \ll P_X \times P_Y$. Let $\{\check{X}_m\}_{m\in[1:\mathsf{L}]} \stackrel{iid}{\sim} P_X$, then for any $\mathsf{J} \in \mathbb{N}$,*

$$\mathbf{E}\left[\left\|\mathsf{L}^{-1}\sum_{m=1}^{\mathsf{L}} P_{Y|X}(\cdot|\check{X}_m) - P_Y(\cdot)\right\|_{\mathrm{TV}}\right]$$
$$\leq \mathbf{E}\left[(1 + 2^{-\iota_{X;Y}(X;Y)})^{-\mathsf{J}}\right] + \frac{1}{2}\sqrt{\mathsf{J}\mathsf{L}^{-1}}. \tag{13}$$

As a result, for any $0 < \gamma \leq \log \mathsf{L}$,

$$\mathbf{E}\left[\left\|\mathsf{L}^{-1}\sum_{m=1}^{\mathsf{L}} P_{Y|X}(\cdot|\check{X}_m) - P_Y(\cdot)\right\|_{\mathrm{TV}}\right]$$
$$\leq \mathbf{P}\{\iota_{X;Y}(X;Y) > \log \mathsf{L} - \gamma\} + 2^{-\gamma/2}\left(1 + \frac{1}{2}\sqrt{\gamma}\right) + \frac{1}{2}\sqrt{\mathsf{L}^{-1}}. \tag{14}$$

Hence there exists a code for channel resolvability satisfying the above bounds.

Proof: Let $\mathfrak{P} = \{\bar{Y}_i, T_i\}_{i\in\mathbb{N}}$ be the points of a Poisson process with intensity measure $P_Y \times \lambda_{\mathbb{R}_{\geq 0}}$. Let $M \sim \mathrm{Unif}[1:\mathsf{L}]$, $\{\check{X}_m\}_{m\in[1:\mathsf{L}]} \stackrel{iid}{\sim} P_X$ ($M \perp\!\!\!\perp \{\check{X}_j\}_j \perp\!\!\!\perp \mathfrak{P}$), and $X = \check{X}_M$. Let $Y = \tilde{Y}_{P_{Y|X}(\cdot|X)}$, and $\hat{Y}_j = \tilde{Y}_{P_Y}(j)$ for $j \in \mathbb{N}$. We have

$$\mathbf{E}\left[\|P_{Y|\{\check{X}_m\}_m}(\cdot|\{\check{X}_m\}_m) - P_Y(\cdot)\|_{\mathrm{TV}}\right]$$
$$\stackrel{(a)}{\leq} \mathbf{E}\left[\|P_{Y|\{\check{X}_m\}_m,\mathfrak{P}}(\cdot|\{\check{X}_m\}_m,\mathfrak{P}) - P_{Y|\mathfrak{P}}(\cdot|\mathfrak{P})\|_{\mathrm{TV}}\right]$$

$$
\begin{aligned}
&\overset{(b)}{=} \frac{1}{2}\mathbf{E}\left[\sum_{y\in\{\hat{Y}_j\}_{j\in\mathbb{N}}} \left|P_{Y|\mathfrak{P}}(y|\mathfrak{P}) - P_{Y|\{\check{X}_m\}_m,\mathfrak{P}}(y|\{\check{X}_m\}_m,\mathfrak{P})\right|\right] \\
&\le \frac{1}{2}\mathbf{E}\Bigg[\sum_{y\in\{\hat{Y}_j\}_{j\in[1:\mathsf{J}]}} \left|P_{Y|\mathfrak{P}}(y|\mathfrak{P}) - P_{Y|\{\check{X}_m\}_m,\mathfrak{P}}(y|\{\check{X}_m\}_m,\mathfrak{P})\right| \\
&\qquad + \sum_{y\in\{\hat{Y}_j\}_{j\in\mathbb{N}}\setminus\{\hat{Y}_j\}_{j\in[1:\mathsf{J}]}} \left(P_{Y|\mathfrak{P}}(y|\mathfrak{P}) + P_{Y|\{\check{X}_m\}_m,\mathfrak{P}}(y|\{\check{X}_m\}_m,\mathfrak{P})\right)\Bigg] \\
&= \frac{1}{2}\mathbf{E}\left[\sum_{y\in\{\hat{Y}_j\}_{j\in[1:\mathsf{J}]}} \left|P_{Y|\mathfrak{P}}(y|\mathfrak{P}) - \mathsf{L}^{-1}\sum_{m=1}^{\mathsf{L}} P_{Y|X,\mathfrak{P}}(y|\check{X}_m,\mathfrak{P})\right|\right] \\
&\qquad + \frac{1}{2}\mathbf{E}\left[P_{Y|\mathfrak{P}}(\mathcal{Y}\setminus\{\hat{Y}_j\}_{j\in[1:\mathsf{J}]}|\mathfrak{P}) + P_{Y|\{\check{X}_m\}_m,\mathfrak{P}}(\mathcal{Y}\setminus\{\hat{Y}_j\}_{j\in[1:\mathsf{J}]}\,|\,\{\check{X}_m\}_m,\mathfrak{P})\right] \\
&= \frac{1}{2}\mathbf{E}\left[\sum_{y\in\{\hat{Y}_j\}_{j\in[1:\mathsf{J}]}} \left|P_{Y|\mathfrak{P}}(y|\mathfrak{P}) - \mathsf{L}^{-1}\sum_{m=1}^{\mathsf{L}} P_{Y|X,\mathfrak{P}}(y|\check{X}_m,\mathfrak{P})\right|\right] + \mathbf{P}\left\{Y\notin\{\hat{Y}_j\}_{j\in[1:\mathsf{J}]}\right\},
\end{aligned}
$$

where (a) is by the convexity of the total variation distance, and (b) is because $Y \in \{\hat{Y}_j\}_{j\in\mathbb{N}}$ almost surely (note that the summation $\sum_{y\in\{\hat{Y}_j\}_{j\in\mathbb{N}}}$ ignores multiplicity of elements in $\{\hat{Y}_j\}_{j\in\mathbb{N}}$). For the first term, note that since Y is a function of $(X,\mathfrak{P})$, we have $P_{Y|X,\mathfrak{P}}(y|\check{X}_m,\mathfrak{P}) \in \{0,1\}$, and hence

$$
\left(\sum_{m=1}^{\mathsf{L}} P_{Y|X,\mathfrak{P}}(y|\check{X}_m,\mathfrak{P})\right)\Bigg|\,\mathfrak{P} \sim \mathrm{Bin}(\mathsf{L},\, P_{Y|\mathfrak{P}}(y|\mathfrak{P})).
$$

We have

$$
\begin{aligned}
&\frac{1}{2}\mathbf{E}\left[\sum_{y\in\{\hat{Y}_j\}_{j\in[1:\mathsf{J}]}} \left|P_{Y|\mathfrak{P}}(y|\mathfrak{P}) - \mathsf{L}^{-1}\sum_{m=1}^{\mathsf{L}} P_{Y|X,\mathfrak{P}}(y|\check{X}_m,\mathfrak{P})\right|\right] \\
&= \frac{1}{2}\mathbf{E}\left[\sum_{y\in\{\hat{Y}_j\}_{j\in[1:\mathsf{J}]}} \mathbf{E}\left[\left|P_{Y|\mathfrak{P}}(y|\mathfrak{P}) - \mathsf{L}^{-1}\sum_{m=1}^{\mathsf{L}} P_{Y|X,\mathfrak{P}}(y|\check{X}_m,\mathfrak{P})\right|\,\middle|\,\mathfrak{P}\right]\right] \\
&\le \frac{1}{2}\mathbf{E}\left[\sum_{y\in\{\hat{Y}_j\}_{j\in[1:\mathsf{J}]}} \sqrt{\mathrm{Var}\left[\mathsf{L}^{-1}\sum_{m=1}^{\mathsf{L}} P_{Y|X,\mathfrak{P}}(y|\check{X}_m,\mathfrak{P})\,\middle|\,\mathfrak{P}\right]}\right] \\
&\le \frac{1}{2}\mathbf{E}\left[\sum_{y\in\{\hat{Y}_j\}_{j\in[1:\mathsf{J}]}} \sqrt{\mathsf{L}^{-1}P_{Y|\mathfrak{P}}(y|\mathfrak{P})}\right] \\
&\le \frac{1}{2}\mathbf{E}\left[\sqrt{\mathsf{J}\sum_{y\in\{\hat{Y}_j\}_{j\in[1:\mathsf{J}]}} \mathsf{L}^{-1}P_{Y|\mathfrak{P}}(y|\mathfrak{P})}\right] \\
&\le \frac{1}{2}\sqrt{\mathsf{J}\mathsf{L}^{-1}}.
\end{aligned}
$$

For the second term, by the conditional generalized Poisson matching lemma on $(X,\, 1,\, Y,\, \emptyset,\, P_Y)$,

$$
\begin{aligned}
&\mathbf{P}\{Y\notin\{\hat{Y}_j\}_{j\in[1:\mathsf{J}]}\} \\
&\le \mathbf{E}\left[\left(1-\left(1+\frac{dP_{Y|X}(\cdot|X)}{dP_Y}(Y)\right)^{-1}\right)^{\mathsf{J}}\right] \\
&= \mathbf{E}\left[(1-(1+2^{\iota_{X;Y}(X;Y)})^{-1})^{\mathsf{J}}\right].
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&\mathbf{E}\left[\|P_{Y|\{\check{X}_j\}_j}(\cdot|\{\check{X}_m\}_m) - P_Y(\cdot)\|_{\mathrm{TV}}\right] \\
&\le \mathbf{E}\left[(1-(1+2^{\iota_{X;Y}(X;Y)})^{-1})^{\mathsf{J}}\right] + \frac{1}{2}\sqrt{\mathsf{J}\mathsf{L}^{-1}}
\end{aligned}
$$

$$= \mathbf{E}\left[(1+2^{-\iota_{X;Y}(X;Y)})^{-\mathsf{J}}\right] + \frac{1}{2}\sqrt{\mathsf{J}\mathsf{L}^{-1}}.$$

For (14), substitute $\mathsf{J} = \lceil \gamma 2^{-\gamma}\mathsf{L}\rceil$,

$$\begin{aligned}
&\mathbf{E}\left[(1-(1+2^{\iota_{X;Y}(X;Y)})^{-1})^{\mathsf{J}}\right] + \frac{1}{2}\sqrt{\mathsf{J}\mathsf{L}^{-1}} \\
&\stackrel{(a)}{\le} \mathbf{E}\left[(1-(1+(2\mathsf{L}2^{-\gamma})^{-1}2^{\iota_{X;Y}(X;Y)})^{-1})^{\mathsf{J}(2\mathsf{L}2^{-\gamma})^{-1}}\right] + \frac{1}{2}\sqrt{\mathsf{J}\mathsf{L}^{-1}} \\
&\le \mathbf{P}\left\{\iota_{X;Y}(X;Y) > \log\mathsf{L} - \gamma\right\} + 2^{-\mathsf{J}(2\mathsf{L}2^{-\gamma})^{-1}} + \frac{1}{2}\sqrt{(\gamma 2^{-\gamma}\mathsf{L}+1)\mathsf{L}^{-1}} \\
&\le \mathbf{P}\left\{\iota_{X;Y}(X;Y) > \log\mathsf{L} - \gamma\right\} + 2^{-\gamma/2} + \frac{1}{2}\sqrt{\gamma 2^{-\gamma}} + \frac{1}{2}\sqrt{\mathsf{L}^{-1}} \\
&= \mathbf{P}\left\{\iota_{X;Y}(X;Y) > \log\mathsf{L} - \gamma\right\} + 2^{-\gamma/2}\left(1+\frac{1}{2}\sqrt{\gamma}\right) + \frac{1}{2}\sqrt{\mathsf{L}^{-1}},
\end{aligned}$$

where (a) is because $\gamma \le \log\mathsf{L}$, $2\mathsf{L}2^{-\gamma} > 1$ and $(1-(1+\alpha)^{-1})^{\beta} \le 1-(1+\beta^{-1}\alpha)^{-1}$ for $\alpha \ge 0$, $\beta \ge 1$.

■

Compare this to Theorem 2 in [32] (weakened by substituting $\delta'_{p,W,C} \le C$): for any $\alpha > 0$,

$$\epsilon \le \mathbf{P}\left\{\iota_{X;Y}(X;Y) > \log\alpha\right\} + \frac{1}{2}\sqrt{\alpha\mathsf{L}^{-1}}.$$

If we assume $1 \le \alpha \le \mathsf{L}$ and substitute $\gamma = \log(\mathsf{L}/\alpha)$ in (14), we obtain the following slightly weaker bound (within a logarithmic gap from that in [32]):

$$\epsilon \le \mathbf{P}\left\{\iota_{X;Y}(X;Y) > \log\alpha\right\} + \sqrt{\alpha\mathsf{L}^{-1}}\left(1+\frac{1}{2}\sqrt{\log(\mathsf{L}/\alpha)}\right) + \frac{1}{2}\sqrt{\mathsf{L}^{-1}}.$$

Nevertheless, the bound in [32] does not imply (13), so neither bound is stronger than the other.

The channel resolvability or soft covering bound in Proposition 2 can be applied to prove various secrecy and coordination results, e.g. one-shot coding for wiretap channels [33], one-shot channel synthesis [31], and one-shot distributed source simulation [34]. Hence these results can also be proved using the Poisson matching lemma alone. In the next section, we will prove a one-shot result for wiretap channels.

XII. One-shot Coding for Wiretap Channels

The one-shot version of the wiretap channel setting [33] is described as follows. Upon observing $M \sim \mathrm{Unif}[1:\mathsf{L}]$, the encoder produces X, which is sent through the broadcast channel $P_{Y,Z|X}$. The legitimate decoder observes Y and recovers $\hat{M}$ with error probability $P_e = \mathbf{P}\{M \neq \hat{M}\}$. The eavesdropper observes Z. Secrecy is measured by the total variation distance $\epsilon := \|P_{M,Z} - P_M \times P_Z\|_{\mathrm{TV}}$.

The following bound is a direct result of the generalized Poisson matching lemma and Proposition 2. It is included for demonstration purposes. See [32], [35], [36] for other one-shot bounds (that are not strictly stronger or weaker than ours).

Proposition 3. *Fix any $P_{U,X}$. For any $\nu \ge 0$, $\mathsf{K}, \mathsf{J} \in \mathbb{N}$, there exists a code for the wiretap channel $P_{Y,Z|X}$, with message $M \sim \mathrm{Unif}[1:\mathsf{L}]$, with average error probability P_e and secrecy measure ϵ satisfying*

$$\begin{aligned}
P_e + \nu\epsilon \le\ & \mathbf{E}\left[\min\{\mathsf{L}\mathsf{K}2^{-\iota_{U;Y}(U;Y)},\, 1\}\right] \\
& + \nu\left(2\mathbf{E}\left[(1+2^{-\iota_{U;Z}(U;Z)})^{-\mathsf{J}}\right] + \sqrt{\mathsf{J}\mathsf{K}^{-1}}\right)
\end{aligned}$$

if $P_{UY} \ll P_U \times P_Y$ and $P_{UZ} \ll P_U \times P_Z$.

Proof: Let $\mathfrak{P} = \{(\bar{U}_i, \bar{M}_i), T_i\}_{i\in\mathbb{N}}$ be the points of a Poisson process with intensity measure $P_U \times P_M \times \lambda_{\mathbb{R}_{\ge 0}}$ independent of M. Let $K \sim \mathrm{Unif}[1:\mathsf{K}]$ independent of $(M, \mathfrak{P})$. The encoder computes $U = \tilde{U}_{P_U \times \delta_M}(K)$ and generates $X|U \sim P_{X|U}$. The decoder recovers $\hat{M} = \tilde{M}_{P_{U|Y}(\cdot|Y)\times P_M}$. We have $(M, K, U, X, Y, Z) \sim P_M \times P_K \times P_{U,X}P_{Y,Z|X}$. By the conditional generalized Poisson matching lemma on $(M,\, K,\, (U, M),\, Y,\, P_{U|Y} \times P_M)$ (note that $P_{U,M|M,K} = P_U \times \delta_M$),

$$\begin{aligned}
&\mathbf{P}\left\{M \neq \hat{M}\right\} \\
&\le \mathbf{E}\left[\mathbf{P}\left\{(U,M) \neq (\tilde{U},\tilde{M})_{P_{U|Y}(\cdot|Y)\times P_M} \,|\, M, K, U, Y\right\}\right] \\
&\le \mathbf{E}\left[\min\left\{\mathsf{K}\frac{dP_U \times \delta_M}{dP_{U|Y}(\cdot|Y) \times P_M}(U, M),\, 1\right\}\right] \\
&= \mathbf{E}\left[\min\{\mathsf{L}\mathsf{K}2^{-\iota_{U;Y}(U;Y)},\, 1\}\right].
\end{aligned}$$

For the secrecy measure,

$$\begin{aligned}
&\mathbf{E}\left[\left\|P_{M,Z|\mathfrak{P}}(\cdot,\cdot|\mathfrak{P}) - P_M(\cdot)\times P_{Z|\mathfrak{P}}(\cdot|\mathfrak{P})\right\|_{\mathrm{TV}}\right] \\
&= \mathbf{E}\left[\left\|P_{Z|M,\mathfrak{P}}(\cdot|M,\mathfrak{P}) - P_{Z|\mathfrak{P}}(\cdot|\mathfrak{P})\right\|_{\mathrm{TV}}\right] \\
&\le \mathbf{E}\left[\left\|P_{Z|M,\mathfrak{P}}(\cdot|M,\mathfrak{P}) - P_Z(\cdot)\right\|_{\mathrm{TV}}\right] + \mathbf{E}\left[\left\|P_{Z|\mathfrak{P}}(\cdot|\mathfrak{P}) - P_Z(\cdot)\right\|_{\mathrm{TV}}\right] \\
&\stackrel{(a)}{\le} 2\mathbf{E}\left[\left\|P_{Z|M,\mathfrak{P}}(\cdot|M,\mathfrak{P}) - P_Z(\cdot)\right\|_{\mathrm{TV}}\right] \\
&= 2\mathbf{E}\left[\left\|\mathsf{K}^{-1}\sum_{k=1}^{\mathsf{K}} P_{Z|U}(\cdot|\tilde{U}_{P_U\times\delta_M}(k)) - P_Z(\cdot)\right\|_{\mathrm{TV}}\right] \\
&\stackrel{(b)}{\le} 2\mathbf{E}\left[(1+2^{-\iota_{U;Z}(U;Z)})^{-\mathsf{J}}\right] + \sqrt{\mathsf{J}\mathsf{K}^{-1}},
\end{aligned}$$

where (a) is by the convexity of total variation distance, and (b) is by Proposition 2 since $\{\tilde{U}_{P_U\times\delta_m}(k)\}_{k\in[1:\mathsf{K}]} \stackrel{iid}{\sim} P_U$ for any m. Therefore there exists a fixed set of points for $\mathfrak{P}$ satisfying the desired bound. ■

XIII. Strong Functional Representation Lemma and Noncausal Sampling

The generalized Poisson matching lemma can be applied to give a slight improvement on the constant in the strong functional representation lemma in [1], and hence improves on the variable-length channel simulation result in [37], and the result on minimax remote prediction with a communication constraint in [38]. It also gives an achievability bound on the moments for the noncausal sampling setting in [39].

Proposition 4. *Let $\{\bar{U}_i, T_i\}_{i\in\mathbb{N}}$ be the points of a Poisson process with intensity measure $\mu\times\lambda_{\mathbb{R}_{\ge 0}}$ over $\mathcal{U}\times\mathbb{R}_{\ge 0}$, and P, Q be probability measures over $\mathcal{U}$ with $P \ll Q \ll \mu$. For any $j\in\mathbb{N}$, $g:\mathbb{R}_{\ge 0}\to\mathbb{R}$ concave nondecreasing, we have*

$$\mathbf{E}\left[g(\Upsilon_{P\|Q}(j)-1)\right] \le \mathbf{E}_{U\sim P}\left[g\left(j\frac{dP}{dQ}(U)\right)\right],$$

i.e., $j(dP/dQ)(U)$ dominates $\Upsilon_{P\|Q}(j)-1$ in the second order. As a result, let

$$\mathfrak{C}[xg'(x)](y) = \inf\left\{\alpha y+\beta:\ xg'(x)\le\alpha x+\beta\ \forall x\ge 0\right\}$$

be the upper concave envelope of $xg'(x)$, then

$$\mathbf{E}\left[g(\Upsilon_{P\|Q}(j))\right] \le \mathbf{E}_{U\sim P}\left[g\left(j\frac{dP}{dQ}(U)\right)\right] + j^{-1}\mathfrak{C}[xg'(x)](j).$$

In particular,

$$\mathbf{E}\left[\log\Upsilon_{P\|Q}(j)\right] \le D(P\|Q) + \log j + j^{-1}\log e,$$

and for $\gamma\in(0,1)$,

$$\begin{aligned}
\mathbf{E}\left[(\Upsilon_{P\|Q}(j))^\gamma\right] &\le j^\gamma\mathbf{E}_{U\sim P}\left[\left(\frac{dP}{dQ}(U)\right)^\gamma\right] + \gamma j^{\gamma-1} \\
&= j^\gamma 2^{\gamma D_{\gamma+1}(P\|Q)} + \gamma j^{\gamma-1},
\end{aligned}$$

where $D_{\gamma+1}(P\|Q) = \gamma^{-1}\log\mathbf{E}_{U\sim P}\left[((dP/dQ)(U))^\gamma\right]$ is the Rényi divergence.

Proof: For $g:\mathbb{R}_{\ge 0}\to\mathbb{R}$ concave nondecreasing, we have

$$\begin{aligned}
&\mathbf{E}\left[g(\Upsilon_{P\|Q}(j)-1)\right] \\
&= \int \mathbf{E}\left[g(\Upsilon_{P\|Q}(j)-1)\,\middle|\,\tilde{U}_P(j)=u\right]P(du) \\
&\stackrel{(a)}{\le} \int g\left(\mathbf{E}\left[\Upsilon_{P\|Q}(j)\,\middle|\,\tilde{U}_P(j)=u\right]-1\right)P(du) \\
&\stackrel{(b)}{\le} \int g\left(j\frac{dP}{dQ}(u)\right)P(du),
\end{aligned}$$

where (a) is by Jensen's inequality, and (b) is by the generalized Poisson matching lemma. For any α,β such that $xg'(x)\le\alpha x+\beta$ for $x\ge 0$,

$$\mathbf{E}\left[g(\Upsilon_{P\|Q}(j))\right]$$

$$
\begin{aligned}
&\leq \int g\left(j\frac{dP}{dQ}(u)+1\right)P(du) \\
&\leq \int g\left(j\frac{dP}{dQ}(u)\right)P(du) + \int g'\left(j\frac{dP}{dQ}(u)\right)P(du) \\
&= \int g\left(j\frac{dP}{dQ}(u)\right)P(du) + j^{-1}\int g'\left(j\frac{dP}{dQ}(u)\right)j\frac{dP}{dQ}(u)Q(du) \\
&\leq \int g\left(j\frac{dP}{dQ}(u)\right)P(du) + j^{-1}\int \left(\alpha j\frac{dP}{dQ}(u)+\beta\right)Q(du) \\
&= \int g\left(j\frac{dP}{dQ}(u)\right)P(du) + j^{-1}(\alpha j+\beta).
\end{aligned}
$$

For $g(x) = \log x$, $xg'(x) = \log e$, and hence

$$
\mathbf{E}\left[\log \Upsilon_{P\|Q}(j)\right] \leq D(P\|Q) + \log j + j^{-1}\log e.
$$

For $g(x) = x^\gamma$, $\gamma \in (0,1)$, $xg'(x) = \gamma x^\gamma$ is concave, and hence

$$
\begin{aligned}
\mathbf{E}\left[(\Upsilon_{P\|Q}(j))^\gamma\right] &\leq \mathbf{E}_{U\sim P}\left[\left(j\frac{dP}{dQ}(U)\right)^\gamma\right] + j^{-1}\gamma j^\gamma \\
&= j^\gamma \mathbf{E}_{U\sim P}\left[\left(\frac{dP}{dQ}(U)\right)^\gamma\right] + \gamma j^{\gamma-1}.
\end{aligned}
$$

■

Consider the setting in the strong functional representation lemma [1]: given (X, Y), we want to find a random variable Z independent of X such that Y is a function of (X, Z), and $H(Y|Z)$ is minimized. Take $Z = \{\bar{Y}_i, T_i\}_{i\in\mathbb{N}}$. Applying Proposition 4 on $P = P_{Y|X}(\cdot|X)$, $Q = P_Y$, we obtain

$$
\begin{aligned}
\mathbf{E}\left[\log \Upsilon_{P_{Y|X}(\cdot|X)\|P_Y}(1)\right] &\leq \mathbf{E}\left[D(P_{Y|X}(\cdot|X)\|P_Y)\right] \\
&= I(X;Y).
\end{aligned}
$$

Using Proposition 4 in [1],

$$
\begin{aligned}
&H(Y|Z) \\
&\leq H(\Upsilon_{P_{Y|X}(\cdot|X)\|P_Y}(1)) \\
&\leq \mathbf{E}\left[\log \Upsilon_{P_{Y|X}(\cdot|X)\|P_Y}(1)\right] + \log\left(\mathbf{E}\left[\log \Upsilon_{P_{Y|X}(\cdot|X)\|P_Y}(1)\right] + 1\right) + 1 \\
&\leq I(X;Y) + \log e + \log\left(I(X;Y) + \log e + 1\right) + 1 \\
&\leq I(X;Y) + \log\left(I(X;Y) + 1\right) + \log e + 1 + \log\left(\log e + 1\right) \\
&\leq I(X;Y) + \log\left(I(X;Y) + 1\right) + 3.732.
\end{aligned}
$$

The constant 3.732 is smaller than that in [1]:

$$
e^{-1}\log e + 2 + \log\left(e^{-1}\log e + 2\right) \approx 3.870.
$$

XIV. Conclusions and Discussion

In this paper, we introduced a simple yet versatile approach to achievability proofs via the Poisson matching lemma. By reducing the uses of sub-codebooks and binning, we improved upon existing one-shot bounds on channels with state information at the encoder, lossy source coding with side information at the decoder, broadcast channels, and distributed lossy source coding. The Poisson matching lemma can replace the packing lemma, covering lemma and soft covering lemma to be the only tool needed to prove a wide range of results in network information theory.

In the proofs, random variables (e.g. the channel input and message in channel coding settings, the source and description in source coding settings, the channel output in channel resolvability) are regarded as points in a Poisson process. The Poisson functional representation is applied to map the Poisson process to give the correct conditional distribution. Viewing every random variable in the operational setting as a Poisson process gives a simple, unified and systematic approach to code constructions.

A possible extension is to generalize the Poisson functional representation to the multivariate case. In the proof of Marton's inner bound for broadcast channels, we had two independent Poisson processes for U_1 and U_2 respectively. We first used the process for U_1 to obtain a list of values for U_1, then used the list to index into the process for U_2. A more symmetric approach where we select (U_1, U_2) together (similar to the conventional mutual covering approach) using a multivariate version of the Poisson functional representation may be possible. Similarly, for distributed lossy source coding and the multiple access channel, it may be possible to decode both sources/messages simultaneously. While it can be argued that the gain we obtained in broadcast channels and distributed lossy source coding over conventional approaches comes from the asymmetry of our construction (our bounds are asymmetric unlike previous bounds), a symmetric treatment that does not result in a looser bound may be developed in the future.

XV. Acknowledgements

The authors acknowledge support from the NSF grants CNS-1527846, CCF-1618145, the NSF Science & Technology Center grant CCF-0939370 (Science of Information), and the William and Flora Hewlett Foundation supported Center for Long Term Cybersecurity at Berkeley.

Appendix

A. Proof of Lemmas 1 and 3

We first prove Lemma 3. For notational simplicity, we use $\{X_i\}_{i\in\mathbb{N}} \sim \mathfrak{P}(\mu)$ to denote that $\{X_i\}_{i\in\mathbb{N}}$ is the set of points of a Poisson process with intensity measure μ (the ordering of the points is ignored). Let $f(u) = (dP/d\mu)(u)$, $g(u) = (dQ/d\mu)(u)$. Let $\{\bar{U}_i, T_i\}_{i\in\mathbb{N}} \sim \mathfrak{P}(\mu \times \lambda_{\mathbb{R}_{\geq 0}})$. Let $\{\check{U}_k, \check{T}_k\}_{k\in\mathbb{N}}$ be the points $(\bar{U}_i, T_i)$ where $f(\bar{U}_i) = 0$. By the mapping theorem [14], [15] on the mapping

$$\psi(u,t) = \begin{cases} (1,\, u,\, t/f(u)) & \text{if } f(u) > 0 \\ (0,\, u,\, t) & \text{if } f(u) = 0, \end{cases}$$

we have $\{\psi(\bar{U}_i, T_i)\}_{i\in\mathbb{N}} \sim \mathfrak{P}(\delta_1 \times P \times \lambda_{\mathbb{R}_{\geq 0}} + \delta_0 \times \mu_{\{f(u)=0\}} \times \lambda_{\mathbb{R}_{\geq 0}})$ (where $\mu_{\{f(u)=0\}}$ denotes μ restricted to the set $\{u : f(u) = 0\}$), and hence $\{\tilde{U}_P(k), \tilde{T}_P(k)\}_{k\in\mathbb{N}} \sim \mathfrak{P}(P \times \lambda_{\mathbb{R}_{\geq 0}})$ (the points in $\{\psi(\bar{U}_i, T_i)\}_{i\in\mathbb{N}}$ with $f(\bar{U}_i) > 0$) is independent of $\{\check{U}_k, \check{T}_k\}_{k\in\mathbb{N}} \sim \mathfrak{P}(\mu_{\{f(u)=0\}} \times \lambda_{\mathbb{R}_{\geq 0}})$ (the points in $\{\psi(\bar{U}_i, T_i)\}_{k\in\mathbb{N}}$ with $f(\bar{U}_i) = 0$).

Condition on $\tilde{U}_P(j) = u$ and $\tilde{T}_P(j) = t$ unless otherwise stated. Assume $f(u) > 0$ (which happens almost surely since $\tilde{U}_P(j) \sim P$) and $g(u) > 0$ (otherwise the inequalities in the lemmas trivially hold). Recall that $\tilde{T}_P(1) \leq \tilde{T}_P(2) \leq \cdots$ by definition. It is straightforward to check that $\{\tilde{U}_P(k), \tilde{T}_P(k)\}_{k>j} \sim \mathfrak{P}(P \times \lambda_{[t,\infty)})$ independent of $\{\tilde{U}_P(k)\}_{k<j} \overset{iid}{\sim} P$ independent of $\{\tilde{T}_P(k)\}_{k<j} \sim \mathrm{Unif}(t\Delta_*^{j-1})$, the uniform distribution over the ordered simplex $t\Delta_*^{j-1} = \{s^{j-1} : 0 \leq s_1 \leq \cdots \leq s_{j-1} \leq t\}$ (i.e., $\{\tilde{U}_P(k), \tilde{T}_P(k)\}_{k<j}$ has the same distribution as $j-1$ i.i.d. points following $P \times \mathrm{Unif}[0,t]$ sorted in ascending order of the second coordinate). We have

$$\begin{aligned}
&\Upsilon_{P\|Q}(j) - 1 \\
&= \left|\left\{k : T_k/g(\bar{U}_k) < tf(u)/g(u)\right\}\right| \\
&= \left|\left\{k : f(\bar{U}_k) = 0 \text{ and } T_k/g(\bar{U}_k) < tf(u)/g(u)\right\}\right| \\
&\quad + \left|\left\{k : \; f(\bar{U}_k) > 0 \text{ and } T_k/g(\bar{U}_k) < tf(u)/g(u)\right\}\right| \\
&= \left|\left\{k : \check{T}_k/g(\check{U}_k) < tf(u)/g(u)\right\}\right| \\
&\quad + \left|\left\{k : \tilde{T}_P(k)f(\tilde{U}_P(k))/g(\tilde{U}_P(k)) < tf(u)/g(u)\right\}\right| \\
&= A_0 + A_1 + B,
\end{aligned}$$

where

$$\begin{aligned}
A_0 &:= \left|\left\{k : \check{T}_k/g(\check{U}_k) < tf(u)/g(u)\right\}\right|, \\
A_1 &:= \left|\left\{k > j : \tilde{T}_P(k)f(\tilde{U}_P(k))/g(\tilde{U}_P(k)) < tf(u)/g(u)\right\}\right|, \\
B &:= \left|\left\{k < j : \tilde{T}_P(k)f(\tilde{U}_P(k))/g(\tilde{U}_P(k)) < tf(u)/g(u)\right\}\right|.
\end{aligned}$$

Due to the aforementioned independence between $\{\check{U}_k, \check{T}_k\}_{k\in\mathbb{N}}$, $\{\tilde{U}_P(k), \tilde{T}_P(k)\}_{k>j}$ and $\{\tilde{U}_P(k), \tilde{T}_P(k)\}_{k<j}$, we have $A_0 \perp\!\!\!\perp A_1 \perp\!\!\!\perp B$. For A_0, since $\{\tilde{U}_P(k), \tilde{T}_P(k)\}_k \perp\!\!\!\perp \{\check{U}_k, \check{T}_k\}_k$, conditioning on $(\tilde{U}_P(j), \tilde{T}_P(j)) = (u,t)$ does not affect the distribution of $\{\check{U}_k, \check{T}_k\}_k$, and hence A_0 follows the Poisson distribution with rate

$$\begin{aligned}
&\left(\mu_{\{f(u)=0\}} \times \lambda\right)\left(\{(v,s) : \; s/g(v) < tf(u)/g(u)\}\right) \\
&= \int \mathbf{1}\{f(v) = 0\} \frac{tg(v)f(u)}{g(u)} \mu(dv).
\end{aligned}$$

For A_1, since $\{\tilde{U}_P(k), \tilde{T}_P(k)\}_{k>j} \sim \mathfrak{P}(P \times \lambda_{[t,\infty)})$, A_1 follows the Poisson distribution with rate

$$\left(P \times \lambda_{[t,\infty)}\right)\left(\{(v,s) : \; sf(v)/g(v) < tf(u)/g(u)\}\right)$$

$$
\begin{aligned}
&= \int \max\left\{\frac{tg(v)f(u)}{f(v)g(u)} - t,\, 0\right\} f(v)\mu(dv) \\
&= t\int \mathbf{1}\{f(v) > 0\} \max\left\{\frac{g(v)f(u)}{g(u)} - f(v),\, 0\right\} \mu(dv).
\end{aligned}
$$

Hence $A := A_0 + A_1$ follows the Poisson distribution with rate

$$
\begin{aligned}
&t\int \left(\mathbf{1}\{f(v) = 0\}\frac{g(v)f(u)}{g(u)} + \mathbf{1}\{f(v) > 0\} \max\left\{\frac{g(v)f(u)}{g(u)} - f(v),\, 0\right\}\right) \mu(dv) \\
&= t\int \max\left\{\frac{g(v)f(u)}{g(u)} - f(v),\, 0\right\} \mu(dv) \\
&= tf(u)\int \max\left\{\frac{g(v)}{g(u)} - \frac{f(v)}{f(u)},\, 0\right\} \mu(dv) \\
&=: t\alpha(u).
\end{aligned}
$$

For B, since $\{\tilde{U}_P(k), \tilde{T}_P(k)\}_{k<j}$ has the same distribution as $j-1$ i.i.d. points following $P \times \mathrm{Unif}[0,t]$ sorted in ascending order of the second coordinate, B follows the binomial distribution with number of trials $j-1$ and success probability

$$
\begin{aligned}
&(P \times \mathrm{Unif}[0,t])\left(\{(v,s) :\, sf(v)/g(v) < tf(u)/g(u)\}\right) \\
&= t^{-1}\int \min\left\{\frac{tg(v)f(u)}{f(v)g(u)},\, t\right\} f(v)\mu(dv) \\
&= f(u)\int \min\left\{\frac{g(v)}{g(u)},\, \frac{f(v)}{f(u)}\right\} \mu(dv) \\
&=: \beta(u).
\end{aligned}
$$

Conditioned on $\tilde{U}_P(j) = u$ (without conditioning on $\tilde{T}_P(j)$), we have $\tilde{T}_P(j) \sim \mathrm{Erlang}(j,1)$, and $(A,B)|\{\tilde{T}_P(j) = t\} \sim \mathrm{Poi}(t\alpha(u)) \times \mathrm{Bin}(j-1, \beta(u))$. Hence, conditioned on $\tilde{U}_P(j) = u$, the distribution of $\Upsilon_{P\|Q}(j) - 1 = A + B$ is

$$
\mathrm{NegBin}\left(j,\, 1 - \frac{1}{1+\alpha(u)}\right) + \mathrm{Bin}(j-1, \beta(u)), \tag{15}
$$

i.e., the sum of a negative binomial random variable and an independent binomial random variable. The mean is

$$
\begin{aligned}
&\mathbf{E}\left[\Upsilon_{P\|Q}(j) \,\middle|\, \tilde{U}_P(j) = u\right] - 1 \\
&= j\alpha(u) + (j-1)\beta(u) \\
&= jf(u)\int \max\left\{\frac{g(v)}{g(u)} - \frac{f(v)}{f(u)},\, 0\right\} \mu(dv) + (j-1)f(u)\int \min\left\{\frac{g(v)}{g(u)},\, \frac{f(v)}{f(u)}\right\} \mu(dv) \\
&= jf(u)\int \frac{g(v)}{g(u)}\mu(dv) - f(u)\int \min\left\{\frac{g(v)}{g(u)},\, \frac{f(v)}{f(u)}\right\} \mu(dv) \\
&\le j\frac{f(u)}{g(u)} \\
&= j\frac{dP}{dQ}(u).
\end{aligned}
$$

Also,

$$
\begin{aligned}
&\mathbf{P}\left\{\Upsilon_{P\|Q}(1) > 1 \,\middle|\, \tilde{U}_P(j) = u\right\} \\
&= 1 - \mathbf{P}\left\{A = 0 \text{ and } B = 0 \,\middle|\, \tilde{U}_P(j) = u\right\} \\
&= 1 - \frac{(1-\beta(u))^{j-1}}{(1+\alpha(u))^j} \\
&\le 1 - \left(\frac{1-\beta(u)}{1+\alpha(u)}\right)^j \\
&\le 1 - (1 - \min\{\alpha(u) + \beta(u), 1\})^j \\
&= 1 - \left(1 - \min\left\{f(u)\int \frac{g(v)}{g(u)}\mu(dv),\, 1\right\}\right)^j \\
&= 1 - \left(1 - \min\left\{\frac{f(u)}{g(u)},\, 1\right\}\right)^j
\end{aligned}
$$

$$= 1 - \left(1 - \min\left\{\frac{dP}{dQ}(u),\, 1\right\}\right)^j.$$

For $j = 1$,

$$\begin{aligned}
&\mathbf{P}\left\{\Upsilon_{P\|Q}(1) > k \,\middle|\, \tilde{U}_P(1) = u\right\} \\
&= \left(1 - (1+\alpha(u))^{-1}\right)^k \\
&= \left(1 - \left(1 + f(u)\int \max\left\{\frac{g(v)}{g(u)} - \frac{f(v)}{f(u)},\, 0\right\}\mu(dv)\right)^{-1}\right)^k \\
&\le \left(1 - \left(1 + f(u)\int \frac{g(v)}{g(u)}\mu(dv)\right)^{-1}\right)^k \\
&= \left(1 - (1+f(u)/g(u))^{-1}\right)^k \\
&\le \left(1 - \left(1 + \frac{dP}{dQ}(u)\right)^{-1}\right)^k \\
&= \exp\left(-k\ln\left(\left(\frac{dP}{dQ}(u)\right)^{-1} + 1\right)\right) \\
&\le \exp\left(-\ln\left(k\left(\frac{dP}{dQ}(u)\right)^{-1} + 1\right)\right) \\
&= 1 - \left(1 + k^{-1}\frac{dP}{dQ}(u)\right)^{-1}.
\end{aligned} \tag{16}$$

B. Proof of the Conditional Poisson Matching Lemma

The conditional Poisson matching lemma is intuitively obvious. The Poisson matching lemma can be equivalently stated as: for any probability measures $\nu, \xi \ll \mu$, the following holds for ν-almost all u:

$$\mathbf{P}_{\{\bar{U}_i,T_i\}_i \sim P_{\{\bar{U}_i,T_i\}_i \,|\, \tilde{U}_\nu = u}}\{\tilde{U}_\xi(\{\bar{U}_i,T_i\}_i) \neq u\} \le 1 - \left(1 + \frac{d\nu}{d\xi}(u)\right)^{-1},$$

where $P_{\{\bar{U}_i,T_i\}_i \,|\, \tilde{U}_\nu = u}$ is the conditional distribution of the Poisson process given $\tilde{U}_\nu = u$. Intuitively, we can consider the Poisson matching lemma to be a statement with 3 parameters ν, ξ, u (ignore the almost-all condition on u for the moment). Since the statement holds for (almost) any (ν, ξ, u), it also holds for any random choice of (ν, ξ, u). In particular, it holds for $(\nu, \xi, u) = (P_{U|X}(\cdot|X),\, Q_{U|Y}(\cdot|Y),\, U)$, where $(X, U, Y) \sim P_{X,U,Y}$, which gives the conditional Poisson matching lemma. Note that the probability in the conditional Poisson matching lemma is conditional on (X, U, Y), where $(X, U, Y) \leftrightarrow (\nu, \xi, u) \leftrightarrow \{\bar{U}_i, T_i\}_i$, and hence conditioning on (X, U, Y) has the same effect on $\{\bar{U}_i, T_i\}_i$ as conditioning on the parameters (ν, ξ, u).

We now prove the conditional Poisson matching lemma rigorously. Let $(\Omega, \mathcal{F}, P_{\{\bar{U}_i,T_i\}_i})$ be the probability space for $\{\bar{U}_i, T_i\}_i$, the points of a Poisson process with intensity measure $\mu \times \lambda_{\mathbb{R}_{\ge 0}}$ on $\mathcal{U} \times \mathbb{R}_{\ge 0}$ (let $\mathcal{E}$ be the Borel σ-algebra of $\mathcal{U}$). The Poisson matching lemma can be equivalently stated as: for any probability measures $\nu, \xi \ll \mu$, and $\kappa : \mathcal{U} \times \mathcal{F} \to [0, 1]$ a regular conditional probability distribution (RCPD) of $\{\bar{U}_i, T_i\}_i$ conditioned on $\tilde{U}_\nu(\{\bar{U}_i, T_i\}_i)$ (i.e., κ is a probability kernel, and $P_{\{\bar{U}_i,T_i\}_i}(A \cap \tilde{U}_\nu^{-1}(B)) = \int_B \kappa(u, A)\nu(du)$ for any $A \in \mathcal{F},\ B \in \mathcal{E}$, where $\tilde{U}_\nu^{-1}(B)$ denotes the preimage of B under $\tilde{U}_\nu : \Omega \to \mathcal{U}$, note that $\tilde{U}_\nu(\{\bar{U}_i, T_i\}_i) \sim \nu$), then we have

$$\int \mathbf{1}\{\tilde{U}_\xi(\{\bar{u}_i, t_i\}_i) \neq u\}\kappa(u,\, d\{\bar{u}_i, t_i\}_i) \le 1 - \left(1 + \frac{d\nu}{d\xi}(u)\right)^{-1} \tag{17}$$

for ν-almost all u.

Consider the conditional Poisson matching lemma. We have the following for $P_{X,U,Y}$-almost all (x, u, y):

$$\begin{aligned}
&\mathbf{P}\left\{\tilde{U}_{Q_{U|Y}(\cdot|Y)} \neq U \,\middle|\, X = x, U = u, Y = y\right\} \\
&= \int \mathbf{1}\{\tilde{U}_{Q_{U|Y}(\cdot|y)}(\{\bar{u}_i, t_i\}_i) \neq u\} P_{\{\bar{U}_i,T_i\}_i|X,U,Y}(d\{\bar{u}_i, t_i\}_i|x, u, y) \\
&\overset{(a)}{=} \int \mathbf{1}\{\tilde{U}_{Q_{U|Y}(\cdot|y)}(\{\bar{u}_i, t_i\}_i) \neq u\} P_{\{\bar{U}_i,T_i\}_i|X,U}(d\{\bar{u}_i, t_i\}_i|x, u) \\
&\overset{(b)}{\le} 1 - \left(1 + \frac{dP_{U|X}(\cdot|x)}{dQ_{U|Y}(\cdot|y)}(u)\right)^{-1},
\end{aligned}$$

where (a) holds for $P_{X,U,Y}$-almost all (x,u,y) due to $Y \leftrightarrow (X,U) \leftrightarrow \{\bar{U}_i,T_i\}_i$, and (b) is by (17) with $(\nu,\xi,\kappa) \leftarrow (P_{U|X}(\cdot|x),\, Q_{U|Y}(\cdot|y),\, P_{\{\bar{U}_i,T_i\}_i|X,U}(\cdot|x,\cdot))$, which holds for $P_{U|X}(\cdot|x)$-almost all u, and hence holds for $P_{X,U,Y}$-almost all (x,u,y). We now check that $P_{\{\bar{U}_i,T_i\}_i|X,U}(\cdot|x,\cdot)$ satisfies the RCPD condition for P_X-almost all x. Since $X \perp\!\!\!\perp \{\bar{U}_i,T_i\}_i$, we have $P_{\{\bar{U}_i,T_i\}_i}(\cdot) = P_{\{\bar{U}_i,T_i\}_i|X}(\cdot|x)$ for P_X-almost all x. Since $U = \tilde{U}_{P_{U|X}(\cdot|X)}(\{\bar{U}_i,T_i\}_i)$, we have $P_{\{\bar{U}_i,T_i\}_i|X,U}(\tilde{U}^{-1}_{P_{U|X}(\cdot|x)}(\{u\}) \,|\, x,u) = 1$ for $P_{X,U}$-almost all (x,u). Hence the following conditions are satisfied for P_X-almost all x:

$$P_{\{\bar{U}_i,T_i\}_i}(\cdot) = P_{\{\bar{U}_i,T_i\}_i|X}(\cdot|x), \tag{18}$$

$$P_{\{\bar{U}_i,T_i\}_i|X,U}(\tilde{U}^{-1}_{P_{U|X}(\cdot|x)}(\{u\})|x,u) = 1 \text{ for } P_{U|X}(\cdot|x)\text{-almost all } u. \tag{19}$$

For any x satisfying (18) and (19), we have the following: for all $A \in \mathcal{F}$, $B \in \mathcal{E}$,

$$\begin{aligned}
&P_{\{\bar{U}_i,T_i\}_i}(A \cap \tilde{U}^{-1}_{P_{U|X}(\cdot|x)}(B)) \\
&\stackrel{(a)}{=} P_{\{\bar{U}_i,T_i\}_i|X}(A \cap \tilde{U}^{-1}_{P_{U|X}(\cdot|x)}(B)|x) \\
&= \int P_{\{\bar{U}_i,T_i\}_i|X,U}(A \cap \tilde{U}^{-1}_{P_{U|X}(\cdot|x)}(B) \,|\, x,u) P_{U|X}(du|x) \\
&\stackrel{(b)}{=} \int P_{\{\bar{U}_i,T_i\}_i|X,U}(A \cap \tilde{U}^{-1}_{P_{U|X}(\cdot|x)}(B) \cap \tilde{U}^{-1}_{P_{U|X}(\cdot|x)}(\{u\}) \,|\, x,u) P_{U|X}(du|x) \\
&= \int \mathbf{1}\{u \in B\} P_{\{\bar{U}_i,T_i\}_i|X,U}(A \cap \tilde{U}^{-1}_{P_{U|X}(\cdot|x)}(\{u\}) \,|\, x,u) P_{U|X}(du|x) \\
&\stackrel{(c)}{=} \int_B P_{\{\bar{U}_i,T_i\}_i|X,U}(A|x,u) P_{U|X}(du|x),
\end{aligned}$$

where (a) is by (18), and (b), (c) are by (19).

C. Proof of Theorem 1

Let $\{\bar{X}_i,T_i\}_{i\in\mathbb{N}}$ be the points of a Poisson process with intensity measure $P_X \times \lambda_{\mathbb{R}_{\ge 0}}$ independent of M. The encoding function is $m \mapsto \tilde{X}_{P_X}(m)$ (i.e., $X = \tilde{X}_{P_X}(M)$), and the decoding function is $y \mapsto \Upsilon_{P_{X|Y}(\cdot|y)\|P_X}(1)$. We have $(M,X,Y) \sim P_M \times P_X P_{Y|X}$,

$$\begin{aligned}
&\mathbf{P}\{M \neq \Upsilon_{P_{X|Y}(\cdot|Y)\|P_X}(1)\} \\
&\stackrel{(a)}{=} \mathbf{P}\{\Upsilon_{P_X\|P_{X|Y}(\cdot|Y)}(M) > 1\} \\
&= \mathbf{E}\left[\mathbf{P}\left\{ \Upsilon_{P_X\|P_{X|Y}(\cdot|Y)}(M) > 1 \,\middle|\, M,X,Y\right\}\right] \\
&\stackrel{(b)}{\le} \mathbf{E}\left[1 - \left(1 - \min\left\{\frac{dP_X}{dP_{X|Y}(\cdot|Y)}(X),\, 1\right\}\right)^M\right] \\
&= \mathbf{E}\left[1 - \left(1 - \min\left\{2^{-\iota_{X;Y}(X;Y)},\, 1\right\}\right)^M\right] \\
&\stackrel{(c)}{\le} \mathbf{E}\left[1 - \left(1 - \min\left\{2^{-\iota_{X;Y}(X;Y)},\, 1\right\}\right)^{(\mathsf{L}+1)/2}\right],
\end{aligned}$$

where (a) is by the definition of Υ, (b) is by the conditional generalized Poisson matching lemma on $(\emptyset, M, X, Y, P_{X|Y})$, and (c) is by $M \perp\!\!\!\perp (X,Y)$ and Jensen's inequality. Therefore there exists a fixed $\{\bar{x}_i,t_i\}_{i\in\mathbb{N}}$ attaining the desired bound.

■

A noteworthy property of this construction is that both the encoder and the decoder do not require knowledge of L. The code can transmit any integer $m \in \mathbb{N}$ with error probability $\mathbf{E}\left[1 - (1 - \min\{2^{-\iota_{X;Y}(X;Y)},\, 1\})^m\right]$, assuming unlimited common randomness $\{\bar{X}_i,T_i\}_{i\in\mathbb{N}}$ between the encoder and the decoder.

D. Dispersion of Joint Source-Channel Coding

We show a second order result for joint source-channel coding using Theorem 4 that coincides with the optimal dispersion in [10]. Consider an i.i.d. source sequence W^k of length k, separable distortion measure $\mathsf{d}(w^k,\hat{z}^k) = \frac{1}{k}\sum_{i=1}^k \mathsf{d}(w_i,\hat{z}_i)$, and n uses of the memoryless channel $P_{Y|X}$. Let $P_{Z|W}$ attain the infimum of the rate-distortion function

$$R(\mathsf{D}) := \inf_{P_{Z|W}:\, \mathbf{E}[\mathsf{d}(W,Z)] \le \mathsf{D}} I(W;Z).$$

The D-tilted information [40] is defined as

$$\jmath_W(w,\mathsf{D}) := -\log \mathbf{E}\left[2^{\nu^*(\mathsf{D}-\mathsf{d}(w,Z))}\right],$$

where $Z \sim P_Z$ (the unconditional Z-marginal of $P_W P_{Z|W}$), and $\nu^* = -R'(\mathsf{D})$ (the derivative exists if the infimum in $R(\mathsf{D})$ is achieved by a unique $P_{Z|W}$ [40]). We invoke a lemma in [40]:

Lemma 5 ([40], Lemma 2)**.** *If the following conditions hold:*

- $\inf\{\tilde{\mathsf{D}} \geq 0 :\ R(\tilde{\mathsf{D}}) < \infty\} < \mathsf{D} < \inf_{z\in\mathcal{Z}} \mathbf{E}[\mathsf{d}(W,z)]$,
- *the infimum in* $R(\mathsf{D})$ *is achieved by a unique* $P_{Z|W}$*,*
- *there exists a finite set* $\tilde{\mathcal{Z}} \subseteq \mathcal{Z}$ *such that* $\mathbf{E}[\min_{z\in\tilde{\mathcal{Z}}} \mathsf{d}(W,z)] < \infty$*, and*
- $\mathbf{E}_{P_W \times P_Z}[(\mathsf{d}(W,Z))^9] < \infty$ *(computed assuming* W, Z *independent),*

then there exist constants $\alpha, \beta, \gamma, k_0 > 0$ *such that for* $k \geq k_0$*,*

$$\mathbf{P}\left\{-\log P_{Z^k}(\mathcal{B}_{\mathsf{D}}(W^k)) \leq \sum_{i=1}^{k} \jmath_W(W_i,\mathsf{D}) + \alpha \log k + \beta\right\} \geq 1 - \frac{\gamma}{\sqrt{k}},$$

where $W^k \stackrel{iid}{\sim} P_W$*, and* $P_{Z^k} = P_Z^{\otimes k}$*.*

We now show a second order result.

Proposition 5. *Fix* P_X*,* $0 < \epsilon < 1$*,* $n, k \in \mathbb{N}$*. We have* $P_e = \mathbf{P}\{\mathsf{d}(W^k, \hat{Z}^k) > \mathsf{D}\} \leq \epsilon$ *if the conditions in Lemma 5 are satisfied,* $k \geq k_0$*, and*

$$nC - kR(\mathsf{D}) \geq \sqrt{nV + k\mathcal{V}(\mathsf{D})}\,\mathcal{Q}^{-1}\left(\epsilon - \frac{\eta}{\sqrt{\min\{n,k\}}}\right) + \alpha \log k + \frac{1}{2}\log n + \beta,$$

where $C := I(X;Y)$*,* $V := \mathrm{Var}[\iota_{X;Y}(X;Y)]$*,* $\mathcal{V}(\mathsf{D}) := \mathrm{Var}[\jmath_W(W,\mathsf{D})]$*, and* $\eta > 0$ *is a constant that depends on* $P_{X,Y}$ *and the distribution of* $\jmath_W(W,\mathsf{D})$*.*

Proof: We have

$$\begin{aligned}
P_e &= \mathbf{P}\{\mathsf{d}(W^k, \hat{Z}^k) > \mathsf{D}\} \\
&\stackrel{(a)}{\leq} \mathbf{P}\left\{-\log P_{Z^k}(\mathcal{B}_{\mathsf{D}}(W^k)) > \sum_{i=1}^{k} \jmath_W(W_i,\mathsf{D}) + \alpha\log k + \beta\right\} \\
&\quad + \mathbf{E}\left[\left(1 + 2^{-\sum_{i=1}^k \jmath_W(W_i,\mathsf{D}) - \alpha\log k - \beta} 2^{\iota_{X^n;Y^n}(X^n;Y^n)}\right)^{-1}\right] \\
&\stackrel{(b)}{\leq} \frac{\gamma}{\sqrt{k}} + \frac{1}{\sqrt{n}} + \mathbf{P}\left\{2^{\sum_{i=1}^k \jmath_W(W_i,\mathsf{D}) - \iota_{X^n;Y^n}(X^n;Y^n) + \alpha\log k + \beta} > \frac{1}{\sqrt{n}}\right\} \\
&= \frac{\gamma}{\sqrt{k}} + \frac{1}{\sqrt{n}} + \mathbf{P}\left\{\sum_{i=1}^{n}(\iota_{X;Y}(X_i;Y_i) - C) - \sum_{i=1}^{k}(\jmath_W(W_i,\mathsf{D}) - R(\mathsf{D})) < -nC + kR(\mathsf{D}) + \alpha\log k + \frac{1}{2}\log n + \beta\right\} \\
&\leq \frac{\gamma}{\sqrt{k}} + \frac{1}{\sqrt{n}} + \mathbf{P}\left\{\sum_{i=1}^{n}(\iota_{X;Y}(X_i;Y_i) - C) - \sum_{i=1}^{k}(\jmath_W(W_i,\mathsf{D}) - R(\mathsf{D})) < -\sqrt{nV + k\mathcal{V}(\mathsf{D})}\,\mathcal{Q}^{-1}\left(\epsilon - \frac{\eta}{\sqrt{\min\{n,k\}}}\right)\right\} \\
&\stackrel{(c)}{\leq} \frac{\gamma}{\sqrt{k}} + \frac{1}{\sqrt{n}} + \epsilon - \frac{\eta}{\sqrt{\min\{n,k\}}} + \frac{\eta - \gamma - 1}{\sqrt{\min\{n,k\}}} \\
&\leq \epsilon
\end{aligned}$$

where (a) is by Theorem 4, (b) is by Lemma 5, and (c) is by the Berry-Esseen theorem [20], [21], [22] if we let $\eta - \gamma - 1$ be a constant given by the Berry-Esseen theorem. ■

This coincides with the optimal dispersion in [10]. Although this is not a self-contained proof (it requires the lemma in [40] for the dispersion of lossy source coding), it shows how we can obtain the achievability of the dispersion in joint source-channel coding from a result on the dispersion of lossy source coding with little additional effort, using the Poisson matching lemma. This proof is considerably simpler than that in [10].

E. Properties of $\phi(t)$

Let $\phi : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$, $\phi(t) = ct^{-1}(\log(t+2))^{-2}$, where $c > 0$ such that $\sum_{j=1}^{\infty} \phi(j) = 1$. Note that $(\phi(t))^{-1}$ is convex. It can be checked numerically that $1 \leq c \leq 2$. We prove a useful inequality about $\phi(t)$.

Proposition 6. *For any* $s>0$, $t\ge 1$, *we have*

$$\min\{s(\phi(t))^{-1},\,1\}\le \min\left\{st\left(\log(s^{-1}+1)+1\right)^{2},\,1\right\}.$$

Moreover, if $st\le 2^{-\alpha}$, $t-1\le 2^{\beta}$, *and* $\tilde{\alpha}\ge\max\{\alpha,0\}$, *then*

$$\min\left\{st\left(\log(s^{-1}+1)+1\right)^{2},\,1\right\}\le 2^{-\alpha}\left(2(\tilde{\alpha}+\beta)^2+2\tilde{\alpha}^2+14\right).$$

Proof: Write $\phi^{-1}(t)$ for the inverse function of ϕ. Since

$$\phi\left(\frac{c}{t\left(\log\left(c/t+2\right)\right)^2}\right)=\frac{t\left(\log\left(c/t+2\right)\right)^2}{\left(\log\left(\frac{c}{t(\log(c/t+2))^2}+2\right)\right)^2}\ge t,$$

we have

$$\phi^{-1}(t)\ge\frac{c}{t\left(\log\left(c/t+2\right)\right)^2}.$$

By the convexity of $(\phi(t))^{-1}$,

$$\begin{aligned}\min\left\{\frac{s}{\phi(t)},\,1\right\}&\le\min\left\{\frac{t}{\phi^{-1}(s)},\,1\right\}\\&\le\min\left\{tc^{-1}s\left(\log\left(c/s+2\right)\right)^2,\,1\right\}\\&\le\min\left\{st\left(\log\left(2/s+2\right)\right)^2,\,1\right\}\\&=\min\left\{st\left(\log(s^{-1}+1)+1\right)^2,\,1\right\}.\end{aligned}$$

If $st\le 2^{-\alpha}$, $t-1\le 2^{\beta}$, and $\tilde{\alpha}\ge\max\{\alpha,0\}$,

$$\begin{aligned}&\min\left\{st\left(\log(s^{-1}+1)+1\right)^2,\,1\right\}\\&=\min\left\{st\left(\log(t/(st)+1)+1\right)^2,\,1\right\}\\&\le 2^{-\alpha}\left(\log((2^{\beta}+1)2^{\alpha}+1)+1\right)^2\\&\le 2^{-\alpha}\left(\log(2^{\tilde{\alpha}+\beta}+2^{\tilde{\alpha}}+1)+1\right)^2\\&\le 2^{-\alpha}\left(\max\{\tilde{\alpha}+\beta,\,\tilde{\alpha}\}+\log 3+1\right)^2\\&\le 2^{-\alpha}\left(2(\tilde{\alpha}+\beta)^2+2\tilde{\alpha}^2+14\right),\end{aligned}$$

where the last inequality follows from considering whether β is positive or negative, and the inequality $(x+y)^2\le 2x^2+2y^2$. ∎

F. Proof of Theorem 5 for Broadcast Channel with Common Message

The parameters $\mathsf{K}_1,\mathsf{K}_2$ correspond to rate splitting. We can split $M_1\in[1:\mathsf{L}_1]$ into $M_{10}\in[1:\mathsf{K}_1]$ and $M_{11}\in[1:\lceil\mathsf{L}_1\mathsf{K}_1^{-1}\rceil]$, and treat M_{10} as part of M_0 to be decoded by both decoders. Although M_{10} and M_{11} may not be uniformly distributed, we can apply a random cyclic shift to M_1 such that $M_1\sim\mathrm{Unif}[1:\mathsf{K}_1\lceil\mathsf{L}_1\mathsf{K}_1^{-1}\rceil]$ (and hence M_{10},M_{11} are also uniform), and condition on a fixed shift at the end. Also M_2 can be split similarly. Therefore we assume $\mathsf{K}_1=\mathsf{K}_2=1$ without loss of generality.

Let $\mathfrak{P}_0=\{(\bar{U}_{0,i},\bar{M}_{00,i}),T_{0,i}\}_{i\in\mathbb{N}}$, $\mathfrak{P}_1=\{(\bar{U}_{1,i},\bar{M}_{01,i},\bar{M}_{1,i}),T_{1,i}\}_{i\in\mathbb{N}}$, $\mathfrak{P}_2=\{(\bar{U}_{2,i},\bar{M}_{02,i},\bar{M}_{2,i}),T_{2,i}\}_{i\in\mathbb{N}}$ be three independent Poisson processes with intensity measures $P_{U_0}\times P_{M_0}\times\lambda_{\mathbb{R}_{\ge 0}}$, $P_{U_1}\times P_{M_0}\times P_{M_1}\times\lambda_{\mathbb{R}_{\ge 0}}$ and $P_{U_2}\times P_{M_0}\times P_{M_2}\times\lambda_{\mathbb{R}_{\ge 0}}$ respectively, independent of M_0,M_1,M_2.

The encoder would generate X such that

$$\begin{aligned}&(M_0,M_1,M_2,U_0,J,\{\check{U}_{1j}\}_{j\in[1:\mathsf{J}]},U_1,U_2,X)\\&\sim P_{M_0}\times P_{M_1}\times P_{M_2}\times P_{U_0}P_JP_{U_1|U_0}^{\otimes\mathsf{J}}\delta_{\check{U}_{1J}}P_{U_2|U_0,U_1}\delta_{x(U_0,U_1,U_2)},\end{aligned}\tag{20}$$

where $P_J=\mathrm{Unif}[1:\mathsf{J}]$, and $\{\check{U}_{1j}\}_{j\in[1:\mathsf{J}]}\in\mathcal{U}_1^{\mathsf{J}}$ is an intermediate list (which can be regarded as a sub-codebook). The term $P_{U_1|U_0}^{\otimes\mathsf{J}}\delta_{\check{U}_{1J}}$ in (20) means that $\{\check{U}_{1j}\}_j$ are conditionally i.i.d. $P_{U_1|U_0}$ given U_0, and $U_1=\check{U}_{1J}$. To accomplish this, the encoder computes $U_0=(\tilde{U}_0)_{P_{U_0}\times\delta_{M_0}}$, $\check{U}_{1j}=(\tilde{U}_1)_{P_{U_1|U_0}(\cdot|U_0)\times\delta_{M_0}\times\delta_{M_1}}(j)$ for $j=1,\ldots,\mathsf{J}$,
$U_2=(\tilde{U}_2)_{\mathsf{J}^{-1}\sum_{j=1}^{\mathsf{J}}P_{U_2|U_0,U_1}(\cdot|U_0,\check{U}_{1j})\times\delta_{M_0}\times\delta_{M_2}}$ (which Poisson process we are referring to can be deduced from whether we

are discussing U_0, U_1 or U_2), $(J,U_1)|(U_0,\{\check{U}_{1j}\}_j,U_2) \sim P_{J,U_1|U_0,\{\check{U}_{1j}\}_j,U_2}$ (where $P_{J,U_1|U_0,\{\check{U}_{1j}\}_j,U_2}$ is derived from (20)), and outputs $X = x(U_0,U_1,U_2)$. It can be verified that (20) is satisfied.

For the decoding function at the decoder $a \in [1:2]$, let $(\check{U}_{0aj},\check{M}_{0aj}) = (\tilde{U}_0,\tilde{M}_{00})_{P_{U_0|Y_a}(\cdot|Y_a)\times P_{M_0}}(j)$ for $j \in \mathbb{N}$, $(\hat{U}_a,\hat{M}_{0a},\hat{M}_a) = (\tilde{U}_a,\tilde{M}_{0a},\tilde{M}_a)_{\sum_{j=1}^{\infty}\phi(j)(P_{U_a|U_0,Y_a}(\cdot|\check{U}_{0aj},Y_a)\times\delta_{\check{M}_{0aj}})\times P_{M_a}}$ where $\phi(j) \propto j^{-1}(\log(j+2))^{-2}$ with $\sum_{j=1}^{\infty}\phi(j) = 1$.

Let $K_a = \Upsilon_{P_{U_0}\times\delta_{M_0}\|P_{U_0|Y_a}(\cdot|Y_a)\times P_{M_0}}(1)$ (using the Poisson process $\mathfrak{P}_0$). By the conditional generalized Poisson matching lemma on $(M_0,\,1,\,(U_0,M_0),\,Y_a,\,P_{U_0|Y_a}\times P_{M_0})$, almost surely,

$$\begin{aligned}\mathbf{E}\left[K_a \mid U_0,Y_a,M_0\right] &\le \frac{dP_{U_0}\times\delta_{M_0}}{dP_{U_0|Y_a}(\cdot|Y_a)\times P_{M_0}}(U_0,M_0)+1\\ &= \mathsf{L}_0 2^{-\iota_{U_0;Y_a}(U_0;Y_a)}+1. \end{aligned} \tag{21}$$

By (20), $U_0 = (\tilde{U}_0)_{P_{U_0}\times\delta_{M_0}}$, $\check{U}_{1j} = (\tilde{U}_1)_{P_{U_1|U_0}(\cdot|U_0)\times\delta_{M_0}\times\delta_{M_1}}(j)$, and $(\check{U}_{01j},\check{M}_{01j}) = (\tilde{U}_0,\tilde{M}_{00})_{P_{U_0|Y_1}(\cdot|Y_1)\times P_{M_0}}(j)$, we have $(M_0,M_1,U_0,J) \perp\!\!\!\perp \mathfrak{P}_1$ and $(\{(\check{U}_{01j},\check{M}_{01j})\}_j,Y_1) \leftrightarrow (M_0,M_1,U_0,J,U_1) \leftrightarrow \mathfrak{P}_1$ (see Figure (2) middle). Hence by the conditional generalized Poisson matching lemma on $((M_0,M_1,U_0),\,J,\,(U_1,M_0,M_1),\,(\{(\check{U}_{01j},\check{M}_{01j})\}_j,Y_1),\,\sum_{j=1}^{\infty}\phi(j)$ $(P_{U_1|U_0,Y_1}(\cdot|\check{U}_{01j},Y_1)\times\delta_{\check{M}_{01j}})\times P_{M_1})$, almost surely,

$$\begin{aligned}&\mathbf{P}\left\{(\tilde{U}_1,\tilde{M}_{01},\tilde{M}_1)_{\sum_{j=1}^{\infty}\phi(j)(P_{U_1|U_0,Y_1}(\cdot|\check{U}_{01j},Y_1)\times\delta_{\check{M}_{01j}})\times P_{M_1}} \neq (U_1,M_0,M_1)\,\middle|\,U_0,U_1,U_2,J,Y_1,Y_2,M_0,M_1\right\}\\
&\overset{(a)}{=} \mathbf{P}\left\{(\tilde{U}_1,\tilde{M}_{01},\tilde{M}_1)_{\sum_{j=1}^{\infty}\phi(j)(P_{U_1|U_0,Y_1}(\cdot|\check{U}_{01j},Y_1)\times\delta_{\check{M}_{01j}})\times P_{M_1}} \neq (U_1,M_0,M_1)\,\middle|\,U_0,U_1,J,Y_1,M_0,M_1\right\}\\
&\overset{(b)}{\le} \mathbf{E}\left[\min\left\{\mathsf{J}\frac{dP_{U_1|U_0}(\cdot|U_0)\times\delta_{M_0}\times\delta_{M_1}}{d(\sum_{j=1}^{\infty}\phi(j)(P_{U_1|U_0,Y_1}(\cdot|\check{U}_{01j},Y_1)\times\delta_{\check{M}_{01j}}))\times P_{M_1}}(U_1,M_0,M_1),\,1\right\}\,\middle|\,U_0,U_1,J,Y_1,M_0,M_1\right]\\
&\le \mathbf{E}\left[\min\left\{\frac{\mathsf{L}_1\mathsf{J}}{\phi(K_1)}\frac{dP_{U_1|U_0}(\cdot|U_0)\times\delta_{M_0}}{dP_{U_1|U_0,Y_1}(\cdot|U_0,Y_1)\times\delta_{M_0}}(U_1,M_0),\,1\right\}\,\middle|\,U_0,U_1,J,Y_1,M_0,M_1\right]\\
&= \mathbf{E}\left[\min\left\{\frac{\mathsf{L}_1\mathsf{J}}{\phi(K_1)}2^{-\iota_{U_1;Y_1|U_0}(U_1;Y_1|U_0)},\,1\right\}\,\middle|\,U_0,U_1,J,Y_1,M_0,M_1\right]\\
&\overset{(c)}{\le} \mathbf{E}\left[K_1\mathsf{L}_1\mathsf{J}2^{-\iota_{U_1;Y_1|U_0}(U_1;Y_1|U_0)}\left(\log(\mathsf{L}_1^{-1}\mathsf{J}^{-1}2^{\iota_{U_1;Y_1|U_0}(U_1;Y_1|U_0)}+1)+1\right)^2\,\middle|\,U_0,U_1,J,Y_1,M_0,M_1\right]\\
&\overset{(d)}{\le} (\mathsf{L}_0 2^{-\iota_{U_0;Y_1}(U_0;Y_1)}+1)\mathsf{L}_1\mathsf{J}2^{-\iota_{U_1;Y_1|U_0}(U_1;Y_1|U_0)}\left(\log(\mathsf{L}_1^{-1}\mathsf{J}^{-1}2^{\iota_{U_1;Y_1|U_0}(U_1;Y_1|U_0)}+1)+1\right)^2,\end{aligned}$$

where (a) is due to $(U_2,Y_2) \leftrightarrow (M_0,M_1,U_0,J,U_1,Y_1) \leftrightarrow \mathfrak{P}_1$ (see Figure 2 middle), (b) is due to the aforementioned application of the conditional generalized Poisson matching lemma, (c) is by Proposition 6, and (d) is due to (21) and $K_1 \leftrightarrow (U_0,Y_1,M_0) \leftrightarrow (J,U_1,M_1)$ (see Figure 2 middle).

Also, since $(M_0,M_2,U_0,\{\check{U}_{1j}\}_j) \perp\!\!\!\perp \mathfrak{P}_2$ and $(\{(\check{U}_{02j},\check{M}_{02j})\}_j,Y_2) \leftrightarrow (M_0,M_2,U_0,\{\check{U}_{1j}\}_j,U_2) \leftrightarrow \mathfrak{P}_2$ (see Figure 2 right), by the conditional Poisson matching lemma on $((M_0,M_2,U_0,\{\check{U}_{1j}\}_j),\,(U_2,M_0,M_2),\,(\{(\check{U}_{02j},\check{M}_{02j})\}_j,Y_2),\,\sum_{j=1}^{\infty}\phi(j)(P_{U_2|U_0,Y_2}$ $(\cdot|\check{U}_{02j},Y_2)\times\delta_{\check{M}_{02j}})\times P_{M_2})$, almost surely,

$$\begin{aligned}&\mathbf{P}\left\{(\tilde{U}_2,\tilde{M}_{02},\tilde{M}_2)_{\sum_{j=1}^{\infty}\phi(j)(P_{U_2|U_0,Y_2}(\cdot|\check{U}_{02j},Y_2)\times\delta_{\check{M}_{02j}})\times P_{M_2}} \neq (U_2,M_0,M_2)\,\middle|\,U_0,U_1,U_2,Y_1,Y_2,M_0,M_2\right\}\\
&\overset{(a)}{=} \mathbf{P}\left\{(\tilde{U}_2,\tilde{M}_{02},\tilde{M}_2)_{\sum_{j=1}^{\infty}\phi(j)(P_{U_2|U_0,Y_2}(\cdot|\check{U}_{02j},Y_2)\times\delta_{\check{M}_{02j}})\times P_{M_2}} \neq (U_2,M_0,M_2)\,\middle|\,U_0,U_2,Y_2,M_0,M_2\right\}\\
&\overset{(b)}{\le} \mathbf{E}\left[\mathbf{E}\left[\min\left\{\frac{d(\mathsf{J}^{-1}\sum_{j=1}^{\mathsf{J}}P_{U_2|U_0,U_1}(\cdot|U_0,\check{U}_{1j}))\times\delta_{M_0}\times\delta_{M_2}}{d(\sum_{j=1}^{\infty}\phi(j)(P_{U_2|U_0,Y_2}(\cdot|\check{U}_{02j},Y_2)\times\delta_{\check{M}_{02j}}))\times P_{M_2}}(U_2,M_0,M_2),\,1\right\}\right.\right.\\
&\qquad\left.\left.\middle|\,\{\check{U}_{1j}\}_j,U_0,U_2,Y_2,M_0,M_2\right]\,\middle|\,U_0,U_2,Y_2,M_0,M_2\right]\\
&\le \mathbf{E}\left[\min\left\{\frac{\mathsf{L}_2}{\phi(K_2)}\frac{d(\mathsf{J}^{-1}\sum_{j=1}^{\mathsf{J}}P_{U_2|U_0,U_1}(\cdot|U_0,\check{U}_{1j}))\times\delta_{M_0}}{dP_{U_2|U_0,Y_2}(\cdot|U_0,Y_2)\times\delta_{M_0}}(U_2,M_0),\,1\right\}\,\middle|\,U_0,U_2,Y_2,M_0,M_2\right]\\
&= \mathbf{E}\left[\min\left\{\frac{\mathsf{L}_2\mathsf{J}^{-1}}{\phi(K_2)}\sum_{j=1}^{\mathsf{J}}2^{\iota_{U_1;U_2|U_0}(\check{U}_{1j};U_2|U_0)-\iota_{U_2;Y_2|U_0}(U_2;Y_2|U_0)},\,1\right\}\,\middle|\,U_0,U_2,Y_2,M_0,M_2\right]\\
&\overset{(c)}{\le} \mathbf{E}\left[\min\left\{\frac{\mathsf{L}_2\mathsf{J}^{-1}}{\phi(K_2)}2^{-\iota_{U_2,Y_2|U_0}(U_2;Y_2|U_0)}(2^{\iota_{U_1;U_2|U_0}(U_1;U_2|U_0)}+\mathsf{J}-1),\,1\right\}\,\middle|\,U_0,U_2,Y_2,M_0,M_2\right]\end{aligned}$$

$$\overset{(d)}{\leq} \mathbf{E}\Big[K_2 \mathsf{L}_2 \mathsf{J}^{-1} 2^{-\iota_{U_2,Y_2|U_0}(U_2;Y_2|U_0)} (2^{\iota_{U_1;U_2|U_0}(U_1;U_2|U_0)} + \mathsf{J} - 1)$$
$$\Big(\log(\mathsf{L}_2^{-1} \mathsf{J} 2^{\iota_{U_2,Y_2|U_0}(U_2;Y_2|U_0)} (2^{\iota_{U_1;U_2|U_0}(U_1;U_2|U_0)} + \mathsf{J} - 1)^{-1} + 1) + 1\Big)^2 \,\Big|\, U_0, U_2, Y_2, M_0, M_2\Big]$$
$$\overset{(e)}{\leq} (\mathsf{L}_0 2^{-\iota_{U_0;Y_2}(U_0;Y_2)} + 1) \mathsf{L}_2 \mathsf{J}^{-1} 2^{-\iota_{U_2,Y_2|U_0}(U_2;Y_2|U_0)} (2^{\iota_{U_1;U_2|U_0}(U_1;U_2|U_0)} + \mathsf{J} - 1)$$
$$\Big(\log(\mathsf{L}_2^{-1} \mathsf{J} 2^{\iota_{U_2,Y_2|U_0}(U_2;Y_2|U_0)} (2^{\iota_{U_1;U_2|U_0}(U_1;U_2|U_0)} + \mathsf{J} - 1)^{-1} + 1) + 1\Big)^2,$$

where (a) is due to $(U_1, Y_1) \leftrightarrow (U_0, U_2, Y_2, M_0, M_2) \leftrightarrow \mathfrak{P}_2$ (see Figure 2 right), (b) is due to the aforementioned application of the conditional Poisson matching lemma, (c) is by the same arguments as in the proof of Theorem 6, (d) is by Proposition 6, and (e) is due to (21) and $K_2 \leftrightarrow (U_0, Y_2, M_0) \leftrightarrow (U_2, M_2)$ (see Figure 2 right). Hence

$$\mathbf{P}\{(M_0, M_0, M_1, M_2) \neq (\hat{M}_{00}, \hat{M}_{01}, \hat{M}_1, \hat{M}_2)\}$$
$$\leq \mathbf{E}\Big[\min\Big\{(\mathsf{L}_0 2^{-\iota_{U_0;Y_1}(U_0;Y_1)} + 1) \mathsf{L}_1 \mathsf{J} 2^{-\iota_{U_1;Y_1|U_0}(U_1;Y_1|U_0)} \Big(\log(\mathsf{L}_1^{-1} \mathsf{J}^{-1} 2^{\iota_{U_1;Y_1|U_0}(U_1;Y_1|U_0)} + 1) + 1\Big)^2$$
$$+ (\mathsf{L}_0 2^{-\iota_{U_0;Y_2}(U_0;Y_2)} + 1) \mathsf{L}_2 \mathsf{J}^{-1} 2^{-\iota_{U_2,Y_2|U_0}(U_2;Y_2|U_0)} (2^{\iota_{U_1;U_2|U_0}(U_1;U_2|U_0)} + \mathsf{J} - 1)$$
$$\Big(\log(\mathsf{L}_2^{-1} \mathsf{J} 2^{\iota_{U_2,Y_2|U_0}(U_2;Y_2|U_0)} (2^{\iota_{U_1;U_2|U_0}(U_1;U_2|U_0)} + \mathsf{J} - 1)^{-1} + 1) + 1\Big)^2, 1\Big\}\Big]$$
$$\leq \mathbf{E}\Big[\min\Big\{\mathsf{L}_0 \mathsf{L}_1 \mathsf{J} A 2^{-\iota_{U_0,U_1;Y_1}(U_0,U_1;Y_1)} + \mathsf{L}_1 \mathsf{J} A 2^{-\iota_{U_1;Y_1|U_0}(U_1;Y_1|U_0)}$$
$$+ \mathsf{L}_0 \mathsf{L}_2 \mathsf{J}^{-1} B 2^{\iota_{U_1;U_2|U_0}(U_1;U_2|U_0) - \iota_{U_0,U_2;Y_2}(U_0,U_2;Y_2)} + \mathsf{L}_0 \mathsf{L}_2 (1 - \mathsf{J}^{-1}) B 2^{-\iota_{U_0,U_2;Y_2}(U_0,U_2;Y_2)}$$
$$+ \mathsf{L}_2 \mathsf{J}^{-1} B 2^{\iota_{U_1;U_2|U_0}(U_1;U_2|U_0) - \iota_{U_2,Y_2|U_0}(U_2;Y_2|U_0)} + \mathsf{L}_2 (1 - \mathsf{J}^{-1}) B 2^{-\iota_{U_2,Y_2|U_0}(U_2;Y_2|U_0)}, 1\Big\}\Big],$$

where $A = (\log(\mathsf{L}_1^{-1} \mathsf{J}^{-1} 2^{\iota_{U_1;Y_1|U_0}(U_1;Y_1|U_0)} + 1) + 1)^2$, $B = \Big(\log((\mathsf{L}_2 \mathsf{J}^{-1} 2^{\iota_{U_1;U_2|U_0}(U_1;U_2|U_0) - \iota_{U_2,Y_2|U_0}(U_2;Y_2|U_0)} + \mathsf{L}_2(1 - \mathsf{J}^{-1}) 2^{-\iota_{U_2,Y_2|U_0}(U_2;Y_2|U_0)})^{-1} + 1) + 1\Big)^2$.

For (6), if the event in (6) does not occur, by Proposition 6,

$$(\mathsf{L}_0 2^{-\iota_{U_0;Y_1}(U_0;Y_1)} + 1) \mathsf{L}_1 \mathsf{J} 2^{-\iota_{U_1;Y_1|U_0}(U_1;Y_1|U_0)} \Big(\log(\mathsf{L}_1^{-1} \mathsf{J}^{-1} 2^{\iota_{U_1;Y_1|U_0}(U_1;Y_1|U_0)} + 1) + 1\Big)^2$$
$$+ (\mathsf{L}_0 2^{-\iota_{U_0;Y_2}(U_0;Y_2)} + 1) \mathsf{L}_2 \mathsf{J}^{-1} 2^{-\iota_{U_2,Y_2|U_0}(U_2;Y_2|U_0)} (2^{\iota_{U_1;U_2|U_0}(U_1;U_2|U_0)} + \mathsf{J} - 1)$$
$$\Big(\log(\mathsf{L}_2^{-1} \mathsf{J} 2^{\iota_{U_2,Y_2|U_0}(U_2;Y_2|U_0)} (2^{\iota_{U_1;U_2|U_0}(U_1;U_2|U_0)} + \mathsf{J} - 1)^{-1} + 1) + 1\Big)^2$$
$$\leq 2^{1-\gamma} \left(2(\iota_{U_1;Y_1|U_0}(U_1;Y_1|U_0))^2 + 2\gamma^2 + 14\right) + 2^{2-\gamma} \left(2(\iota_{U_2,Y_2|U_0}(U_2;Y_2|U_0))^2 + 2\gamma^2 + 14\right)$$
$$\leq 2^{-\gamma} \left(8(\iota_{U_1;Y_1|U_0}(U_1;Y_1|U_0))^2 + 8(\iota_{U_2,Y_2|U_0}(U_2;Y_2|U_0))^2 + 12\gamma^2 + 84\right).$$

References

[1] 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.

[2] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Contr. and Inf. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[3] S. Verdú, "Non-asymptotic achievability bounds in multiuser information theory," in *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*, Oct 2012, pp. 1–8.

[4] J. Liu, P. Cuff, and S. Verdú, "One-shot mutual covering lemma and Marton's inner bound with a common message," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 1457–1461.

[5] M. H. Yassaee, M. R. Aref, and A. Gohari, "A technique for deriving one-shot achievability results in network information theory," in *2013 IEEE International Symposium on Information Theory*, July 2013, pp. 1287–1291.

[6] E. C. Song, P. Cuff, and H. V. Poor, "The likelihood encoder for lossy compression," *IEEE Transactions on Information Theory*, vol. 62, no. 4, pp. 1836–1849, 2016.

[7] M. H. Yassaee, M. R. Aref, and A. Gohari, "Non-asymptotic output statistics of random binning and its applications," in *2013 IEEE International Symposium on Information Theory*, July 2013, pp. 1849–1853.

[8] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, January 1976.

[9] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder-ii. general sources," *Information and control*, vol. 38, no. 1, pp. 60–80, 1978.

[10] V. Kostina and S. Verdú, "Lossy joint source-channel coding in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 59, no. 5, pp. 2545–2575, May 2013.

[11] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications*, G. Longo, Ed. New York: Springer-Verlag, 1978, pp. 171–231.

[12] S.-Y. Tung, "Multiterminal source coding," Ph.D. dissertation, Cornell University, Ithaca, NY, 1978.

[13] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Transactions on Information Theory*, vol. 25, no. 3, pp. 306–311, May 1979.

[14] J. F. C. Kingman, *Poisson Processes*. Oxford University Press, 1993.

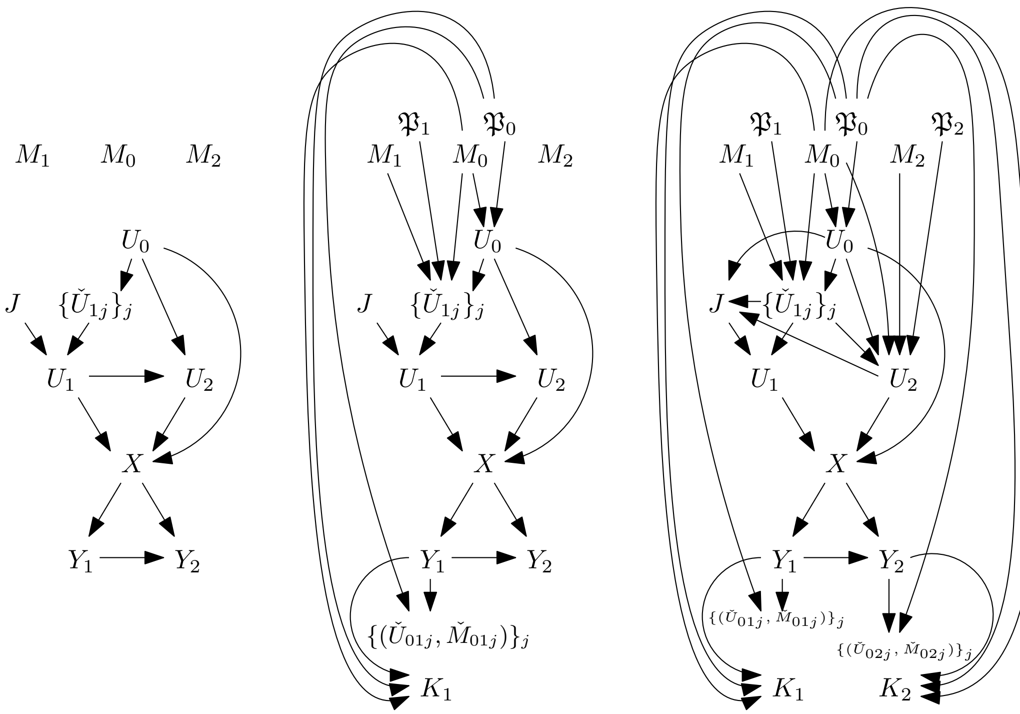

Figure 2. Left: The Bayesian network described in (20). Middle: The Bayesian network deduced from (20), $U_0 = (\tilde{U}_0)_{P_{U_0} \times \delta_{M_0}}$, $\check{U}_{1j} = (\tilde{U}_1)_{P_{U_1|U_0}(\cdot|U_0) \times \delta_{M_0} \times \delta_{M_1}}(j)$, and $(\check{U}_{01j}, \check{M}_{01j}) = (\tilde{U}_0, \tilde{M}_{00})_{P_{U_0|Y_1}(\cdot|Y_1) \times P_{M_0}}(j)$. Right: The Bayesian network describing the scheme. Note that all three are valid Bayesian networks, and the desired conditional independence relations can be deduced using d-separation.

[15] G. Last and M. Penrose, *Lectures on the Poisson process*. Cambridge University Press, 2017, vol. 7.
[16] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge University Press, 2011.
[17] 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.
[18] J. Scarlett, "On the dispersions of the Gel'fand–Pinsker channel and dirty paper coding," *IEEE Transactions on Information Theory*, vol. 61, no. 9, pp. 4569–4586, 2015.
[19] S. Watanabe, S. Kuzuoka, and V. Y. F. Tan, "Nonasymptotic and second-order achievability bounds for coding with side-information," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1574–1605, April 2015.
[20] A. C. Berry, "The accuracy of the Gaussian approximation to the sum of independent variates," *Transactions of the American Mathematical Society*, vol. 49, no. 1, pp. 122–136, 1941.
[21] C.-G. Esseen, "On the Liapunov limit error in the theory of probability," *Ark. Mat. Astr. Fys.*, vol. 28, pp. 1–19, 1942.
[22] W. Feller, *An introduction to probability theory and its applications*, 2nd ed. Wiley, New York, 1971, vol. 2.
[23] Y. Liang and G. Kramer, "Rate regions for relay broadcast channels," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3517–3535, Oct 2007.
[24] S. I. Gel'fand and M. S. Pinsker, "Capacity of a broadcast channel with one deterministic component," *Problemy Peredachi Informatsii*, vol. 16, no. 1, pp. 24–34, 1980.
[25] Y. Liang, G. Kramer, and H. V. Poor, "On the equivalence of two achievable regions for the broadcast channel," *IEEE Transactions on Information Theory*, vol. 57, no. 1, pp. 95–100, Jan 2011.
[26] V. Bentkus, "On the dependence of the Berry–Esseen bound on dimension," *Journal of Statistical Planning and Inference*, vol. 113, no. 2, pp. 385–402, 2003.
[27] R. Ahlswede, "Multi-way communication channels," in *2nd Int. Symp. Inform. Theory, Tsahkadsor, Armenian SSR*, 1971, pp. 23–52.
[28] H. Liao, "Multiple access channels," Ph.D. dissertation, University of Hawaii, Honolulu, HI, 1972.
[29] R. Ahlswede, "The capacity region of a channel with two senders and two receivers," *The annals of probability*, vol. 2, no. 5, pp. 805–814, 1974.
[30] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 752–772, May 1993.
[31] P. Cuff, "Distributed channel synthesis," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7071–7096, Nov 2013.
[32] M. Hayashi, "General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1562–1575, April 2006.
[33] A. D. Wyner, "The wire-tap channel," *Bell system technical journal*, vol. 54, no. 8, pp. 1355–1387, 1975.
[34] ——, "The common information of two dependent random variables," *IEEE Transactions on Information Theory*, vol. 21, no. 2, pp. 163–179, 1975.
[35] M. H. Yassaee, "One-shot achievability via fidelity," in *Proc. IEEE Int. Symp. Inf. Theory*. IEEE, 2015, pp. 301–305.
[36] J. Liu, P. Cuff, and S. Verdú, "E_γ-resolvability," *IEEE Transactions on Information Theory*, vol. 63, no. 5, pp. 2629–2658, May 2017.
[37] P. Harsha, R. Jain, D. McAllester, and J. Radhakrishnan, "The communication complexity of correlation," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 438–449, Jan 2010.
[38] C. T. Li, X. Wu, A. Ozgur, and A. El Gamal, "Minimax learning for remote prediction," in *2018 IEEE International Symposium on Information Theory (ISIT)*, June 2018, pp. 541–545.
[39] J. Liu and S. Verdú, "Rejection sampling and noncausal sampling under moment constraints," in *2018 IEEE International Symposium on Information Theory (ISIT)*, June 2018, pp. 1565–1569.
[40] 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.