One-Shot Coding over General Noisy Networks

Yanxiao Liu and Cheuk Ting Li
Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong, China
Email: yanxiaoliu@link.cuhk.edu.hk, ctli@ie.cuhk.edu.hk

Abstract—**THIS PAPER IS ELIGIBLE FOR THE STUDENT PAPER AWARD. We present a unified one-shot coding framework designed for communication and compression of messages among multiple nodes across a general acyclic noisy network. Our setting can be seen as a one-shot version of the acyclic discrete memoryless network studied by Lee and Chung, and noisy network coding studied by Lim, Kim, El Gamal and Chung. We design a proof technique, called the exponential process refinement lemma, that is rooted in the Poisson matching lemma by Li and Anantharam, and can significantly simplify the analyses of one-shot coding over multi-hop networks. Our one-shot coding theorem not only recovers a wide range of existing asymptotic results, but also yields novel one-shot achievability results in different multi-hop network information theory problems. In a broader context, our framework provides a unified one-shot bound applicable to any combination of source coding, channel coding and coding for computing problems.**

I. Introduction

In information theory, we often employ the law of large numbers to study the asymptotic behavior of networks in the large blocklength limit. However, due to the fact that packet lengths are not infinite in real applications, *finite blocklength information theory* has been widely studied in the past decade. See [1]–[4]. In particular, one-shot coding is the setting where each source and channel is only used once. It is general in the sense that the sources and channels can be arbitrary, and does not have to be memoryless or ergodic. The goal is to derive one-shot results that can imply existing asymptotic [5] and finite-blocklength [1]–[4] results. See [6]–[13] for existing one-shot results for specific multiuser settings.

This paper studies a general class of networks, called *acyclic discrete networks*. Each node plays the role of an encoder and/or a decoder in source or channel coding settings. It is a one-shot version of the asymptotic acyclic discrete memoryless network [14] and noisy network coding [15],[1] and includes a wide range of settings as special cases, such as source and channel coding, primitive relay channel [5], [19]–[22], relay-with-unlimited-look-ahead [24], Gelfand-Pinsker [25], Wyner-Ziv [26], [27], coding for computing [28], multiple access channels [29]–[31] and broadcast channels [32].

To track a large number of auxiliary random variables, we propose the *exponential process refinement lemma* based on the Poisson matching lemma [9], which simplifies the analyses of the evolution of the posterior distribution of the sources, messages and/or auxiliary random variables at the decoder. We prove a one-shot achievability result for general acyclic discrete networks. It recovers one-shot results similar to existing ones in the settings in [6]–[9] and gives novel one-shot results for multi-hop problems, e.g., primitive relay channels [5], [19]–[21] and relay-with-unlimited-look-ahead [23], [24].

Some proofs are in the appendix due to space constraint.

Notations

We assume all random variables are from finite alphabets and logarithm and entropy are to the base 2 unless otherwise stated. Logarithm to the base e is denoted as $\ln(x)$. Write $[i..j] := \{i, i+1, \ldots, j\}$, $[j] := [1..j]$. For a set $\mathcal{S} \subseteq [k]$ and random sequence $U_1, \ldots, U_k$, write $U^k := (U_1, \ldots, U_k)$, $U_{\mathcal{S}} := (U_j)_{j \in \mathcal{S}}$. For a statement S, $\mathbf{1}\{S\}$ is its indicator, i.e., $\mathbf{1}\{S\}$ is 1 if S holds and 0 otherwise. $\iota_{X;Y|Z}(x;y|z) := \log(P_{X,Y|Z}(x,y|z)/(P_{X|Z}(x|z)P_{Y|Z}(y|z)))$ denotes the conditional information density. We use $\iota(X;Y)$ instead of $\iota_{X;Y}(X;Y)$ when the random variables are clear from the context. δ_a denotes the degenerate distribution $\mathbf{P}\{X = a\} = 1$.

II. Exponential Process Refinement Lemma

In this section, we design a tool for proving one-shot achievability results over noisy multi-hop networks. It is called the *exponential process refinement lemma* and can be viewed as a refinement of the Poisson matching lemma [9], which in turn is based on the Poisson functional representation [33].

Consider a finite set $\mathcal{U}$. Let $\mathbf{U} := (Z_u)_{u \in \mathcal{U}}$ be i.i.d. $\mathrm{Exp}(1)$ random variables.[2] Given a distribution P over $\mathcal{U}$,

$$\mathbf{U}_P := \mathrm{argmin}_u Z_u / P(u) \tag{1}$$

is the Poisson functional representation [33]. We have $\mathbf{U}_P \sim P$. We can generalize this by letting $\mathbf{U}_P(1), \ldots, \mathbf{U}_P(|\mathcal{U}|) \in \mathcal{U}$ be the elements of $\mathcal{U}$ sorted in ascending order of $Z_u/P(u)$:

$$Z_{\mathbf{U}_P(1)}/P(\mathbf{U}_P(1)) \le \cdots \le Z_{\mathbf{U}_P(|\mathcal{U}|)}/P(\mathbf{U}_P(|\mathcal{U}|)).$$

We break ties arbitrarily and treat $1/0 = \infty$. This is similar to the mapped Poisson process in the generalized Poisson matching lemma [9], though unlike [9], $\mathbf{U}_P(1), \ldots, \mathbf{U}_P(|\mathcal{U}|)$ is not an i.i.d. sequence following P. Write $\mathbf{U}_P^{-1} : \mathcal{U} \to [|\mathcal{U}|]$ for the inverse function of $i \mapsto \mathbf{U}_P(i)$. The following is a direct corollary of the generalized Poisson matching lemma [9].

Lemma 1 (Generalized Poisson matching lemma [9]): For distributions P, Q over $\mathcal{U}$, we have the following almost surely:

$$\mathbf{E}\big[\mathbf{U}_Q^{-1}(\mathbf{U}_P) \,\big|\, \mathbf{U}_P\big] \le P(\mathbf{U}_P)/Q(\mathbf{U}_P) + 1.$$

[1]Technically, general relay channels [16]–[18] and noisy network coding [15] cannot be treated as one-shot settings since the relay encodes in an iterative manner. The acyclic discrete network in this paper provides the best "one-shot approximation", and includes the primitive relay channel [5], [19]–[22] and the relay-with-unlimited-look-ahead [23], [24] as special cases.

[2]When the space $\mathcal{U}$ is continuous, a Poisson process is used in [9], [33].

We now define a convenient tool.

Definition 1 (Refining a distribution by an exponential process): For a joint distribution $Q_{V,U}$ over $\mathcal{V}\times\mathcal{U}$, the refinement of $Q_{V,U}$ by $\mathbf{U}$, denoted as $Q^{\mathbf{U}}_{V,U}$, is a joint distribution

$$Q^{\mathbf{U}}_{V,U}(v,u) := Q_V(v)\Big/\Big(\mathbf{U}^{-1}_{Q_{U|V}(\cdot|v)}(u)\sum_{i=1}^{|\mathcal{U}|} i^{-1}\Big)$$

for all (v,u) in the support of $Q_{V,U}$, where Q_V is the V-marginal of $Q_{V,U}$ and $Q_{U|V}$ is the conditional distribution of U given V. When $V=\emptyset$, the above definition becomes

$$Q^{\mathbf{U}}_{U}(u) = \frac{1}{\mathbf{U}^{-1}_{Q_U}(u)\sum_{i=1}^{|\mathcal{U}|} i^{-1}}.$$

While the Poisson functional representation $\mathbf{U}_{Q_U}$ (which only gives one value of U) is used for the unique decoding of U, the refinement $Q^{\mathbf{U}}_{U}(u)$ is for the *soft decoding* of U, which gives a distribution over U, with $\mathbf{U}_{Q_U}$ having the largest probability. This is useful in non-unique decoding. For example, if we want to decode U_1 uniquely, while utilizing U_2 via non-unique decoding, we can first obtain the distribution $(\mathbf{U}_2)_{Q_{U_2}}$, and then compute the marginal distribution of U_1 in $(\mathbf{U}_2)_{Q_{U_2}} P_{U_1|U_2}$ and use this marginal distribution to recover U_1 via the Poisson functional representation.

Loosely speaking, if the distribution $Q_{V,U}$ represents our "prior distribution" of (V,U), then the refinement $Q^{\mathbf{U}}_{V,U}$ is our updated "posterior distribution" after taking the exponential process $\mathbf{U}$ into account. In multiterminal coding settings that a node decodes multiple random variables, the prior distribution of those random variables will be refined by multiple exponential processes. To keep track of the evolution of the "posterior probability" of the correct values of those random variables through the refinement process, we use the following lemma, called the *exponential process refinement lemma*. Although its proof (given in Appendix A) still relies on the Poisson matching lemma [9], it significantly simplifies our analyses.

Lemma 2 (Exponential Process Refinement Lemma): For a distribution P over $\mathcal{U}$ and a joint distribution $Q_{V,U}$ over a finite $\mathcal{V}\times\mathcal{U}$, for every $v\in\mathcal{V}$, we have, almost surely,

$$\mathbf{E}\left[\frac{1}{Q^{\mathbf{U}}_{V,U}(v,\mathbf{U}_P)}\middle|\mathbf{U}_P\right] \le \frac{\ln|\mathcal{U}|+1}{Q_V(v)}\left(\frac{P(\mathbf{U}_P)}{Q_{U|V}(\mathbf{U}_P|v)}+1\right).$$

III. Network Model

We describe a general N-node network model, which is the one-shot version of the *acyclic discrete memoryless network (ADMN)* [14]. There are N nodes labelled $1,\ldots,N$. Node i observes $Y_i\in\mathcal{Y}_i$ and produces $X_i\in\mathcal{X}_i$ ($\mathcal{X}_i,\mathcal{Y}_i$ are finite). Unlike conventional asymptotic settings (e.g. [14]), here X_i is only one symbol, instead of a sequence $(X_{i,1},\ldots,X_{i,n})$. The transmission is performed sequentially, and each Y_i is allowed to depend on all previous inputs and outputs (i.e., X^{i-1},Y^{i-1}) in a stochastic manner. Therefore, we can formally define an *N-node acyclic discrete network (ADN)* as a collection of channels $(P_{Y_i|X^{i-1},Y^{i-1}})_{i\in[N]}$, where $P_{Y_i|X^{i-1},Y^{i-1}}$ is a conditional distribution from $\prod_{j=1}^{i-1}\mathcal{X}_j\times\prod_{j=1}^{i-1}\mathcal{Y}_j$ to $\mathcal{Y}_i$. Note that the asymptotic ADMN [14] can be considered as the n-fold ADN $(P^n_{Y_i|X^{i-1},Y^{i-1}})_{i\in[N]}$, where $P^n_{Y_i|X^{i-1},Y^{i-1}}$ denotes the n-fold product conditional distribution (i.e., n copies of a memoryless channel), and we take the blocklength $n\to\infty$.

We remark that, similar to [14], the X_i's and Y_i's can represent sources, states, channel inputs, outputs and messages in source coding and channel coding settings. For example, for point-to-point channel coding, we take Y_1 to be the message, which the encoder (node 1) encodes into the channel input X_1, which in turn is sent through the channel $P_{Y_2|X_1}$. The decoder (node 2) observes Y_2 and outputs X_2, which is the decoded message. For lossless source coding, Y_1 is the source, $X_1=Y_2$ is the description by the encoder, and X_2 is the reconstruction.

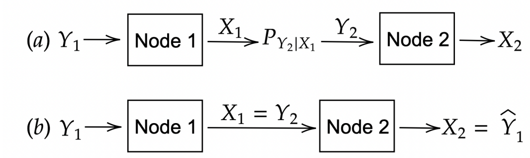

Fig. 1. (a) Channel coding. (b) Source coding.

We give the definition of a coding scheme below.

Definition 2: A *deterministic coding scheme* consists of a sequence of encoding functions $(f_i)_{i\in[N]}$, where $f_i:\mathcal{Y}_i\to\mathcal{X}_i$. For $i=1,\ldots,N$, the following operations are performed:

- **Noisy channel.** The output $\tilde{Y}_i$ is generated conditional on $\tilde{X}^{i-1},\tilde{Y}^{i-1}$ according to $P_{Y_i|X^{i-1},Y^{i-1}}$. For $i=1$, $\tilde{Y}_1\sim P_{Y_1}$ can be regarded as a source or a channel state.
- **Node operation.** Node i sees $\tilde{Y}_i$ and outputs $\tilde{X}_i=f_i(\tilde{Y}_i)$.

We sometimes allow an additional unlimited public randomness available to all nodes.

Definition 3: A *public-randomness coding scheme* for the network consists of a pair $(P_W,(f_i)_{i\in[N]})$, where P_W is the distribution of the public randomness $W\in\mathcal{W}$ available to all nodes and $f_i:\mathcal{Y}_i\times\mathcal{W}\to\mathcal{X}_i$ is the encoding function of node i mapping its observation Y_i and the public randomness W to its output X_i. The operations are as follows. First, generate $W\sim P_W$. For $i=1,\ldots,N$, generate $\tilde{Y}_i$ conditional on $\tilde{X}^{i-1},\tilde{Y}^{i-1}$ according to $P_{Y_i|X^{i-1},Y^{i-1}}$, and take $\tilde{X}_i=f_i(\tilde{Y}_i,W)$.

We do not impose any constraint on the public randomness W. In reality, to carry out a public-randomness coding scheme, the nodes share a common random seed to initialize their pseudorandom number generators before the scheme commences.

We use $\tilde{X}_i,\tilde{Y}_i$ to denote the actual random variables from the coding scheme. In contrast, X_i,Y_i usually denote the random variables following an ideal distribution. For example, in channel coding, the ideal distribution is $Y_1=X_2\sim\mathrm{Unif}[\mathsf{L}]$ (i.e., the message is decoded without error), independent of $(X_1,Y_2)\sim P_{X_1}P_{Y_2|X_1}$. If we ensure that the actual $\tilde{X}^2,\tilde{Y}^2$ is "close to" the ideal X^2,Y^2, this would imply that $\tilde{Y}_1=\tilde{X}_2$ with high probability as well, giving a small error probability.

The goal (the "achievability") is to make the actual joint distribution $P_{\tilde{X}^N,\tilde{Y}^N}$ "approximately as good as" the ideal joint distribution P_{X^N,Y^N}. If we have an "error set" $\mathcal{E}\subseteq\big(\prod_{i=1}^N\mathcal{X}_i\big)\times\big(\prod_{i=1}^N\mathcal{Y}_i\big)$ that we do not want $(\tilde{X}^N,\tilde{Y}^N)$ to fall into (e.g., for channel coding, $\mathcal{E}$ is the set where $\tilde{Y}_1 \neq \tilde{X}_2$, i.e., an error coours; for lossy source coding, $\mathcal{E}$ is the set where $d(\tilde{Y}_1, \tilde{X}_2) > \mathsf{D}$, i.e., the distortion exceeds the limit), we want

$$\mathbf{P}\big((\tilde{X}^N, \tilde{Y}^N) \in \mathcal{E}\big) \lesssim \mathbf{P}\big((X^N, Y^N) \in \mathcal{E}\big). \tag{2}$$

If $P_{\tilde{X}^N,\tilde{Y}^N}$ is close to P_{X^N,Y^N} in total variation distance, i.e.,

$$\delta_{\mathrm{TV}}\big(P_{X^N,Y^N},\, P_{\tilde{X}^N,\tilde{Y}^N}\big) \approx 0, \tag{3}$$

then (2) is guaranteed. For public-randomness coding, we show that (3) can be achieved, which can be seen as a channel simulation [34], [35] or a coordination [36] result. For deterministic coding, since the node operations are deterministic, there might not be sufficient randomness to make $P_{\tilde{X}^N,\tilde{Y}^N}$ close to P_{X^N,Y^N}, and hence we use the error bound in (2).

IV. Main Theorem for Acyclic Discrete Networks

We show a one-shot achievability result for ADN via public-randomness coding scheme. The proof is in Appendix B.

Theorem 3: Fix any ADN $(P_{Y_i|X^{i-1},Y^{i-1}})_{i\in[N]}$. For any collection of indices $(a_{i,j})_{i\in[N],j\in[d_i]}$ where $(a_{i,j})_{j\in[d_i]}$ is a sequence of distinct indices in $[i-1]$ for each i, any sequence $(d'_i)_{i\in[N]}$ with $0 \le d'_i \le d_i$ and any collection of conditional distributions $(P_{U_i|Y_i,\overline{U}'_i}, P_{X_i|Y_i,U_i,\overline{U}'_i})_{i\in[N]}$ (where $\overline{U}_{i,\mathcal{S}} := (U_{a_{i,j}})_{j\in\mathcal{S}}$ for $\mathcal{S} \subseteq [d_i]$ and $\overline{U}'_i := \overline{U}_{i,[d'_i]}$), which induces the joint distribution of X^N, Y^N, U^N (the "ideal distribution"), there exists a public-randomness coding scheme $(P_W, (f_i)_{i\in[N]})$ such that the joint distribution of $\tilde{X}^N, \tilde{Y}^N$ induced by the scheme (the "actual distribution") satisfies

$$\delta_{\mathrm{TV}}\big(P_{X^N,Y^N},\, P_{\tilde{X}^N,\tilde{Y}^N}\big) \le \mathbf{E}\bigg[\min\bigg\{\sum_{i=1}^{N}\sum_{j=1}^{d'_i} B_{i,j},\, 1\bigg\}\bigg],$$

where[3] $\gamma_{i,j} := \prod_{k=j+1}^{d_i}(\ln|\mathcal{U}_{a_{i,k}}|+1)$, and $B_{i,j} :=$

$$\gamma_{i,j}\prod_{k=j}^{d_i}\big(2^{-\iota(\overline{U}_{i,k};\overline{U}_{i,[d_i]\backslash[j..k]},Y_i)+\iota(\overline{U}_{i,k};\overline{U}'_{a_{i,k}},Y_{a_{i,k}})} + \mathbf{1}\{k>j\}\big). \tag{4}$$

The sequences $(a_{i,j})_j$ control which auxiliaries U_j node i decodes and in which order. Node i uniquely decodes $\overline{U}'_i = (U_{a_{i,j}})_{j\in[d'_i]}$ while utilizing $(U_{a_{i,j}})_{j\in[d'_i+1..d_i]}$ by non-unique decoding via the exponential process refinement (Def 1). For brevity, we say "the *decoding order* of node i is $\overline{U}_{i,1}, \ldots, \overline{U}_{i,d'_i}, \overline{U}_{i,d'_i+1}?, \ldots, \overline{U}_{i,d_i}?$" where "?" means the random variable is only used in non-unique decoding. Node i decodes $\overline{U}'_i$, creates its own U_i by using the Poisson functional representation on $P_{U_i|Y_i,\overline{U}'_i}$, and generates X_i from $P_{X_i|Y_i,U_i,\overline{U}'_i}$.

We also have the following result for deterministic coding schemes. The proof is given in Appendix B.

Theorem 4: Fix any ADN $(P_{Y_i|X^{i-1},Y^{i-1}})_{i\in[N]}$. For any $(a_{i,j})_{i\in[N],j\in[d_i]}$, $(d'_i)_{i\in[N]}$, $(P_{U_i|Y_i,\overline{U}'_i}, P_{X_i|Y_i,U_i,\overline{U}'_i})_{i\in[N]}$ as defined in Theorem 3, which induce the joint distribution of X^N, Y^N, U^N, and any set $\mathcal{E} \subseteq \big(\prod_{i=1}^N \mathcal{X}_i\big)\times\big(\prod_{i=1}^N \mathcal{Y}_i\big)$, there is a deterministic coding scheme $(f_i)_{i\in[N]}$ such that with $B_{i,j}$ defined in Theorem 3, $\tilde{X}^N, \tilde{Y}^N$ induced by the scheme satisfy

$$\begin{aligned}&\mathbf{P}\big((\tilde{X}^N, \tilde{Y}^N) \in \mathcal{E}\big)\\ &\le \mathbf{E}\bigg[\min\bigg\{\mathbf{1}\big\{(X^N, Y^N) \in \mathcal{E}\big\} + \sum_{i=1}^{N}\sum_{j=1}^{d'_i} B_{i,j},\, 1\bigg\}\bigg].\end{aligned} \tag{5}$$

Theorem 3 implies the following result for the asymptotic ADMN [14] by directly applying the law of large numbers.

Corollary 5: Fix any ADN $(P_{Y_i|X^{i-1},Y^{i-1}})_{i\in[N]}$. Fix any $(a_{i,j})_{i\in[N],j\in[d_i]}$, $(d'_i)_{i\in[N]}$, $(P_{U_i|Y_i,\overline{U}'_i}, P_{X_i|Y_i,U_i,\overline{U}'_i})_{i\in[N]}$ as defined in Theorem 3, which induces the joint distribution of X^N, Y^N, U^N. If for every $i \in [N]$, $j \in [d'_i]$,

$$\begin{aligned}&I(\overline{U}_{i,j};\overline{U}_{i,[d_i]\backslash\{j\}}, Y_i) - I(\overline{U}_{i,j};\overline{U}'_{a_{i,j}}, Y_{a_{i,j}}) > \sum_{k=j+1}^{d_i}\Big(\\ &\max\big\{I(\overline{U}_{i,k};\overline{U}'_{a_{i,k}}, Y_{a_{i,k}}) - I(\overline{U}_{i,k};\overline{U}_{i,[d_i]\backslash[j..k]}, Y_i),\, 0\big\}\Big),\end{aligned}$$

then there is a sequence of public-randomness coding (indexed by n) for the n-fold ADN $(P^n_{Y_i|X^{i-1},Y^{i-1}})_{i\in[N]}$ such that the induced $\tilde{X}^{N,n}, \tilde{Y}^{N,n}$ (write $\tilde{X}^{N,n} = (\tilde{X}_{i,j})_{i\in[N],j\in[n]}$) satisfy

$$\lim_{n\to\infty}\delta_{\mathrm{TV}}\big(P^n_{X^N,Y^N},\, P_{\tilde{X}^{N,n},\tilde{Y}^{N,n}}\big) = 0. \tag{6}$$

While this result is not as strong as the general asymptotic result in [14], a one-shot analogue of [14] will likely be significantly more complicated than Theorem 3. We choose to present Theorem 3 since it is simple and already powerful enough to give a wide range of tight one-shot results.

V. One-shot Relay Channel

To explain our scheme, we discuss a *one-shot relay channel* in Figure 2. An encoder observes $M \sim \mathrm{Unif}[\mathsf{L}]$ and outputs X, which is passed through the channel $P_{Y_\mathrm{r}|X}$. The relay observes Y_r and outputs X_r. Then $(X, X_\mathrm{r}, Y_\mathrm{r})$ is passed through the channel $P_{Y|X,X_\mathrm{r},Y_\mathrm{r}}$. The decoder sees Y and recovers $\hat{M}$. For generality, we allow Y to depend on all of $X, X_\mathrm{r}, Y_\mathrm{r}$, and X_r may interfere with (X, Y_r), which can happen if the relay outputs X_r instantaneously or the channel has a long memory, or it is a storage device. It is a one-shot version of the *relay-without-delay* and *relay-with-unlimited-look-ahead* [23], [24], and is an ADN by taking $Y_1 = M$, $X_1 = X$, $Y_2 = Y_\mathrm{r}$, $X_2 = X_\mathrm{r}$, $Y_3 = Y$, and $X_3 = M$ (in the ideal distribution).

In case if $Y = (Y', Y'')$ consists of two components and the channel $P_{Y|X,X_\mathrm{r},Y_\mathrm{r}} = P_{Y'|X,Y_\mathrm{r}}P_{Y''|X_\mathrm{r}}$ can be decomposed into two orthogonal components (so X_r does not interfere with (X, Y_r)), this becomes the one-shot version of the *primitive relay channel* [5], [19]–[21] since the n-fold version of this ADN (with $n \to \infty$) is precisely the asymptotic primitive relay channel. However, the n-fold version of the ADN in Figure 2 in general is not the conventional relay channel [16]–[18] (it is the relay-with-unlimited-look-ahead instead). The conventional relay channel, due to its causal assumption that the relay can only look at past $Y_{\mathrm{r},t}$'s, has no one-shot counterpart.

We use the corollary below to explain the use of Lemma 2.

[3]The logarithmic terms $\gamma_{i,j}$ do not affect the first and second order results.

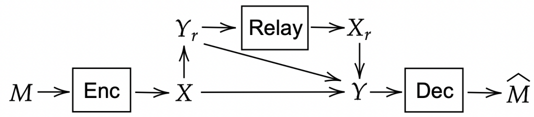

Fig. 2. One-shot relay channel.

Corollary 6: For any P_X, $P_{U|Y_\mathrm{r}}$, function $x_\mathrm{r}(y_\mathrm{r}, u)$, there is a deterministic coding scheme for the one-shot relay channel such that with $\gamma := \ln|\mathcal{U}| + 1$, the error probability satisfies

$$P_e \le \mathbf{E}\Big[\min\big\{\gamma \mathsf{L} 2^{-\iota(X;U,Y)}\big(2^{-\iota(U;Y)+\iota(U;Y_\mathrm{r})}+1\big), 1\big\}\Big], \quad (7)$$

where $(X, Y_\mathrm{r}, U, X_\mathrm{r}, Y) \sim P_X P_{Y_\mathrm{r}|X} P_{U|Y_\mathrm{r}} \delta_{x_\mathrm{r}(Y_\mathrm{r},U)} P_{Y|X,Y_\mathrm{r},X_\mathrm{r}}$.

Proof: Let $U_1 := (X, M)$, $U_2 := U$. Let $\mathbf{U}_1$, $\mathbf{U}_2$ be two independent exponential processes, which serve as the "random codebooks". The encoder (node 1) uses the Poisson functional representation (1) to compute $U_1 = (\mathbf{U}_1)_{P_{U_1}\times\delta_M}$ and outputs X-component of U_1. The relay (node 2) computes $U_2 = (\mathbf{U}_2)_{P_{U_2|Y_\mathrm{r}}(\cdot|Y_\mathrm{r})}$ and outputs $X_\mathrm{r} = x_\mathrm{r}(Y_\mathrm{r}, U_2)$. Note that $X, Y_\mathrm{r}, U_2, X_\mathrm{r}, Y$ follow the ideal distribution in the corollary due to the property of Poisson functional representation, and hence we write X_r instead of $\tilde{X}_\mathrm{r}$. The decoder (node 3) observes Y, and performs the following steps.

1) Refine $P_{U_2|Y}(\cdot|Y)$ (written as $P_{U_2|Y}$ for brevity) to $Q_{U_2} := P_{U_2|Y}^{\mathbf{U}_2}$ using Definition 1. By the exponential process refinement lemma (Lemma 2, with $V = \emptyset$),

$$\mathbf{E}\bigg[\frac{1}{Q_{U_2}(U_2)}\bigg|\, U_2, Y, Y_\mathrm{r}\bigg] \le (\ln|\mathcal{U}_2| + 1)\bigg(\frac{P_{U_2|Y_\mathrm{r}}(U_2)}{P_{U_2|Y}(U_2)} + 1\bigg).$$

2) Compute the joint distribution $Q_{U_2} P_{U_1|U_2,Y}$ over $\mathcal{U}_1 \times \mathcal{U}_2$, the semidirect product between Q_{U_2} and $P_{U_1|U_2,Y}(\cdot|\cdot, Y)$. Let its U_1-marginal be $\tilde{Q}_{U_1}$.

3) Let $\tilde{U}_1 = (\mathbf{U}_1)_{\tilde{Q}_{U_1}\times P_M}$, and output its M-component.

Let $A := (X, Y_\mathrm{r}, U_2, X_\mathrm{r}, Y, M)$ and $\gamma := \ln|\mathcal{U}_2| + 1$,

$$\begin{aligned}
&\mathbf{P}(\tilde{U}_1 \ne U_1 \,|\, A)\\
&\overset{(a)}{\le} \mathbf{E}\left[\min\left\{\frac{P_{U_1}(U_1)\delta_M(M)}{P_{U_1|U_2,Y}(U_1|U_2,Y)Q_{U_2}(U_2)P_M(M)}, 1\right\} \,\middle|\, A\right]\\
&\overset{(b)}{=} \mathbf{E}\left[\min\left\{\mathsf{L}\frac{P_{U_1}(U_1)}{P_{U_1|U_2,Y}(U_1|U_2,Y)Q_{U_2}(U_2)}, 1\right\} \,\middle|\, A\right]\\
&\overset{(c)}{\le} \min\left\{\mathsf{L}\frac{P_{U_1}(U_1)}{P_{U_1|U_2,Y}(U_1|U_2,Y)}\gamma\bigg(\frac{P_{U_2|Y_\mathrm{r}}(U_2)}{P_{U_2|Y}(U_2)} + 1\bigg), 1\right\}\\
&= \min\big\{\gamma\mathsf{L}2^{-\iota(X;U_2,Y)}\big(2^{-\iota(U_2;Y)+\iota(U_2;Y_\mathrm{r})} + 1\big), 1\big\},
\end{aligned}$$

where (a) is by the generalized Poisson matching lemma [9] (Lemma 1), (b) is by $\delta_M(M) = 1$ and $P_M(M) = 1/\mathsf{L}$, and (c) is by step 1) and Jensen's inequality. Taking expectation over A gives the desired error bound. Although the codebooks $\mathbf{U}_1$, $\mathbf{U}_2$ are random (so this is a public-randomness scheme), we can convert it to a deterministic scheme by fixing one particular choice $(\mathbf{u}_1, \mathbf{u}_2)$ that satisfies the error bound. ■

Corollary 6 yields the following asymptotic achievable rate:

$$R \le I(X; U, Y) - \max\big\{I(U; Y_\mathrm{r}) - I(U; Y),\, 0\big\}$$

for some $P_{U|Y_\mathrm{r}}$ and function $x_\mathrm{r}(y_\mathrm{r}, u_2)$.

For the one-shot primitive relay channel that $P_{Y|X,X_\mathrm{r},Y_\mathrm{r}} = P_{Y'|X,Y_\mathrm{r}} P_{Y''|X_\mathrm{r}}$ can be decomposed into two orthogonal components, consider (X, Y_r, Y') independent of (X_r, Y'') in the ideal distribution and take $U = (U', X_\mathrm{r})$ where U' follows $P_{U'|Y_\mathrm{r}}$, Corollary 6 specializes to the following result.

Corollary 7: For any P_X, P_{X_r}, $P_{U'|Y_\mathrm{r}}$, there is a deterministic coding scheme for the one-shot primitive relay channel with $M \sim \mathrm{Unif}[\mathsf{L}]$ such that the error probability of satisfies

$$P_e \le \mathbf{E}\Big[\min\Big\{\gamma\mathsf{L}2^{-\iota(X;U',Y')}\big(2^{-\iota(X_\mathrm{r};Y'')+\iota(U';Y_\mathrm{r}|Y')}+1\big), 1\Big\}\Big],$$

where $(X, Y_\mathrm{r}, U', Y') \sim P_X P_{Y_\mathrm{r}|X} P_{U'|Y_\mathrm{r}} P_{Y'|X,Y_\mathrm{r}}$ is independent of $(X_\mathrm{r}, Y'') \sim P_{X_\mathrm{r}} P_{Y''|X_\mathrm{r}}$, and $\gamma := \ln(|\mathcal{U}'||\mathcal{X}_\mathrm{r}|) + 1$.

This gives the asymptotic achievable rate $R \le I(X; U', Y') - \max\{I(U'; Y_\mathrm{r}|Y') - C_\mathrm{r},\, 0\}$ where $C_\mathrm{r} = \max_{P_{X_\mathrm{r}}} I(X_\mathrm{r}; Y'')$ is the capacity of the channel $P_{Y''|X_\mathrm{r}}$. It implies the compress-forward bound [19], which is the maximum of $I(X; U', Y')$ subject to the constraint $C_\mathrm{r} \ge I(U'; Y_\mathrm{r}|Y')$ (where the random variables are distributed as in Corollary 7). Hence, Corollary 7 can be treated as a one-shot compress-and-forward bound.

A. Partial-Decode-and-Forward Bound

We extend Corollary 6 to allow partial decoding of the message [18], [19], [24]. To this end, we split the message and encoder into two. The message $M \sim \mathrm{Unif}[\mathsf{L}]$ is split into $M_1 \sim \mathrm{Unif}[\mathsf{J}]$ and $M_2 \sim \mathrm{Unif}[\mathsf{L}/\mathsf{J}]$ (assume J is a factor of L). The encoder controls two nodes (node 1 and 2), where node 1 observes $Y_1 = M_1$, outputs $X_1 = V$, and has an auxiliary $U_1 = (M_1, V)$; node 2 observes $Y_2 = (M_1, M_2, V)$, outputs $X_2 = X$, and has an auxiliary $U_2 = (M_1, M_2, X)$. The relay (node 3) observes $Y_3 = Y_r$, decodes U_1, outputs $X_3 = X_r$, and has an auxiliary $U_3 = (M_1, U)$. The decoder (node 4) observes $Y_4 = Y$ and uses the decoding order "$U_2, U_3?, U_1?$".

Corollary 8: Fix any $P_{X,V}$, $P_{U|Y_\mathrm{r},V}$, function $x_\mathrm{r}(y_\mathrm{r}, u, v)$, and J which is a factor of L. There exists a deterministic coding scheme for the one-shot relay channel with

$$\begin{aligned}
P_e \le \mathbf{E}\Big[\min\Big\{&\mathsf{J}2^{-\iota(V;Y_\mathrm{r})} + \gamma\mathsf{L}\mathsf{J}^{-1}2^{-\iota(X;U,Y|V)}\\
&\cdot\big(2^{-\iota(U;V,Y)+\iota(U;V,Y_\mathrm{r})} + 1\big)\big(\mathsf{J}2^{-\iota(V;Y)} + 1\big), 1\Big\}\Big],
\end{aligned}$$

where $(X, V, Y_\mathrm{r}, U, X_\mathrm{r}, Y) \sim P_{X,V} P_{Y_\mathrm{r}|X,V} P_{U|Y_\mathrm{r},V} \delta_{x_\mathrm{r}(Y_\mathrm{r},U,V)} P_{Y|X,Y_\mathrm{r},X_\mathrm{r}}$ and $\gamma := (\ln(\mathsf{J}|\mathcal{U}|) + 1)(\ln(\mathsf{J}|\mathcal{V}|) + 1)$.

The implied asymptotic achievable rate is in Appendix C. Taking $x_\mathrm{r}(y_\mathrm{r}, (v', x'_\mathrm{r})) = x'_\mathrm{r}$, $U = \emptyset$, $V = (V', X'_\mathrm{r})$, it gives an achievable rate $\min\{I(X, X_\mathrm{r}; Y),\, I(V'; Y_\mathrm{r}) + I(X; Y|X_\mathrm{r}, V')\}$, recovering the partial noncausal decode-forward bound for relay-with-unlimited-look-ahead [24, Prop. 3].

Specializing to the primitive relay channel, and again substituting $U = \emptyset$, $V = (V', X'_\mathrm{r})$, $x_\mathrm{r}(y_\mathrm{r}, (v', x'_\mathrm{r})) = x'_\mathrm{r}$, we have

$$\begin{aligned}
P_e \le \mathbf{E}\Big[\min\Big\{&\mathsf{J}2^{-\iota(V';Y_\mathrm{r})} + 2\gamma\mathsf{L}\mathsf{J}^{-1}2^{-\iota(X;Y'|V')}\\
&\cdot\big(\mathsf{J}2^{-\iota(V';Y')-\iota(X_\mathrm{r};Y'')} + 1\big), 1\Big\}\Big],
\end{aligned}$$

where $\gamma := (\ln\mathsf{J} + 1)(\ln(\mathsf{J}|\mathcal{V}'||\mathcal{X}_\mathrm{r}|) + 1)$ and $(X, V', Y_\mathrm{r}, Y') \sim P_{X,V'} P_{Y_\mathrm{r}|X} P_{Y'|X,Y_\mathrm{r}}$ is independent of $(X_\mathrm{r}, Y'') \sim P_{X_\mathrm{r}} P_{Y''|X_\mathrm{r}}$ and . It gives the asymptotic rate $\min\{I(V'; Y_\mathrm{r}) +$

$I(X;Y|V'),\ I(X;Y)+C_{\mathrm{r}}\}$ and recovers the partial decode-forward lower bound for primitive relay channels [18], [19]. One-shot versions of other asymptotic bounds for primitive relay channels (e.g., [20], [21]) are left for future studies.

VI. Examples of Acyclic Discrete Networks

In this section, we apply the Theorem 3 and Theorem 4 on network information theory settings: Gelfand-Pinsker [25], Wyner-Ziv [26], [27], coding for computing [28], multiple access channels [29]–[31] and broadcast channels [32].

A. Gelfand-Pinsker Problem [25]

Upon observing $M \sim \mathrm{Unif}[\mathsf{L}]$ and $S \sim P_S$, the encoder generates X and sends X through a channel $P_{Y|X,S}$. The decoder receives Y and recovers $\hat{M}$. This can be considered as an ADN: in the ideal situation, let $Y_1 := (M,S)$ represent all the information coming into node 1, $Y_2 := Y$, $P_{Y_2|Y_1,X_1}$ be $P_{Y|S,X}$, and $X_2 := M$. The auxiliary of node 1 is $U_1 = (U,M)$ for some U following $P_{U|S}$ given S. The decoding order of node 2 is "U_1" (i.e., it only wants U_1). Since node 2 has decoded U_1, X_2 is allowed to depend on $U_1 = (U,M)$, and hence the choice $X_2 := M$ is valid in the ideal situation. Nevertheless, in the actual situation where we have $\tilde{X}, \tilde{Y}$ instead of X, Y, the actual output $\tilde{X}_2$ will not be exactly M, though the error probability $P_e := \mathbf{P}(\tilde{X}_2 \neq M)$ can still be bounded. We have Corollary 9 by applying Theorem 4.

Corollary 9: Fix $P_{U|S}$ and function $x : \mathcal{U} \times \mathcal{S} \to \mathcal{X}$. There exists a deterministic coding scheme for $P_{Y|X,S}$ with

$$P_e \le \mathbf{E}\big[\min\big\{\mathsf{L}2^{-\iota(U;Y)+\iota(U;S)}, 1\big\}\big],$$

where $S,U,X,Y \sim P_S P_{U|S} \delta_{x(U,S)} P_{Y|X,S}$.
This bound is similar to the one given in [9] (which is stronger than the one-shot bounds in [6]–[8] in the second order). Both of them attain the second-order result in [37].

B. Wyner-Ziv Problem [26], [27] and Coding for Computing

Upon observing $X \sim P_X$, the encoder outputs $M \in [\mathsf{L}]$. The decoder receives M and the side information $T \sim P_{T|X}$, and recovers $\tilde{Z} \in \mathcal{Z}$ with probability of excess distortion $P_e := \mathbf{P}\{d(X,\tilde{Z}) > \mathsf{D}\}$, where $d : \mathcal{X} \times \mathcal{Z} \to \mathbb{R}_{\ge 0}$ is a distortion measure. This can be considered as an ADN: in the ideal situation, $Y_1 := X$, $X_1 := M$, $Y_2 := (M,T)$, $X_2 := Z$. The auxiliary of node 1 is $U_1 = (U,M)$ for some U following $P_{U|X}$ given X. By Theorem 4, we bound P_e as follows.

Corollary 10: Fix $P_{U|X}$ and function $z : \mathcal{U} \times \mathcal{Y} \to \mathcal{Z}$. There is a deterministic coding scheme for Wyner-Ziv problem with

$$P_e \le \mathbf{E}\Big[\min\Big\{\mathbf{1}\{d(X,Z) > \mathsf{D}\} + \mathsf{L}^{-1}2^{-\iota(U;T)+\iota(U;X)}, 1\Big\}\Big], \tag{8}$$

where $X,Y,U,Z \sim P_X P_{Y|X} P_{U|X} \delta_{z(U,Y)}$.

This bound is similar to the one given in [9] (which in turn is stronger than the one-shot bounds in [6], [8]).

This reduces to lossy source coding with $T = \emptyset$. Let $U = Z$, we have $P_e \le \mathbf{P}(d(X,Z) > \mathsf{D}) + \mathbf{E}\left[\min\left\{\mathsf{L}^{-1}2^{\iota(Z;X)}, 1\right\}\right]$.

We also consider coding for computing [28], where node 2 recovers a function $f(X,T)$ of X and T with respect to distortion level D with a distortion measure $d(\cdot,\cdot)$. The probability of excess distortion is $P_e := \mathbf{P}\{d(f(X,T),\tilde{Z}) > \mathsf{D}\}$. We obtain a result similar to Corollary 10, where (8) is changed to $P_e \le \mathbf{E}[\min\{\mathbf{1}\{d(f(X,T),Z) > \mathsf{D}\} + \mathsf{L}^{-1}2^{-\iota(U;T)+\iota(U;X)}, 1\}]$.

C. Multiple Access Channel [29]–[31]

There are two encoders, one decoder, and two independent messages $M_j \sim \mathrm{Unif}[\mathsf{L}_j]$ for $j = 1,2$. Encoder j observes M_j and creates X_j for $j = 1,2$. The decoder observes the output Y of the channel $P_{Y|X_1,X_2}$ and produces the reconstructions $(\hat{M}_1, \hat{M}_2)$ of the messages. The error probability is defined as $P_e := \mathbf{P}\{(M_1,M_2) \neq (\hat{M}_1,\hat{M}_2)\}$. To consider this as an ADN, in the ideal situation, we let $Y_1 := M_1$, $Y_2 := M_2$, $Y_3 := Y$ and $X_3 := (M_1,M_2)$. We let $U_1 := (X_1,M_1)$ and $U_2 := (X_2,M_2)$. The decoding order of node 3 is "U_2, U_1". By Theorem 4, we have the following result.

Corollary 11: Fix P_{X_1}, P_{X_2}. There exists a deterministic coding scheme for the multiple access channel $P_{Y|X_1,X_2}$ with

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

where $\gamma := \ln(\mathsf{L}_1|\mathcal{X}_1|)+1$, $(X_1,X_2,Y) \sim P_{X_1}P_{X_2}P_{Y|X_1,X_2}$.

This bound is similar to the one-shot bounds in [6], [9]. In the asymptotic setting, this will give the region $R_1 < I(X_1;Y|X_2)$, $R_2 < I(X_2;Y|X_1)$, $R_1 + R_2 < I(X_1,X_2;Y)$.

D. Broadcast Channel with Private Messages [32]

Upon observing independent messages $M_j \sim \mathrm{Unif}[\mathsf{L}_j]$ for $j = 1,2$, the encoder produces X and sends it through a channel $P_{Y_1,Y_2|X}$. Decoder j observes Y_j and reconstructs $\hat{M}_j$ for $j = 1,2$. By Theorem 4, we have the following result.

Corollary 12: Fix P_{U_1,U_2} and $x : \mathcal{U}_1 \times \mathcal{U}_2 \to \mathcal{X}$. There is a deterministic coding scheme for the broadcast channel with

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

where $(U_1,U_2,X,Y_1,Y_2) \sim P_{U_1,U_2}\delta_{x(U_1,U_2)}P_{Y_1,Y_2|X}$.

In the asymptotic case, this gives a corner point in Marton's region [32]: $R_1 < I(U_1;Y_1)$, $R_2 < I(U_2;Y_2) - I(U_1;U_2)$. Another corner point can be obtained by swapping the decoders.

VII. Concluding Remarks

In this paper, we have provided a unified one-shot coding framework over a general noisy network, applicable to any combination of source coding, channel coding and coding for computing problems. A unified coding scheme may be useful for designing automated theorem proving tools for one-shot results. For example, [38] gives an algorithm for deriving asymptotic inner and outer bounds for general ADMN [14]. Extension of [38] to one-shot results is left for future study.

VIII. Acknowledgements

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

References

[1] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[2] 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, 2013.

[3] D. Wang, A. Ingber, and Y. Kochman, "The dispersion of joint source-channel coding," in *2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sep. 2011, pp. 180–187.

[4] V. Y. Tan and O. Kosut, "On the dispersions of three network information theory problems," *IEEE Transactions on Information Theory*, vol. 60, no. 2, pp. 881–903, 2013.

[5] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge university press, 2011.

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

[7] 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*. IEEE, 2013, pp. 1287–1291.

[8] S. Watanabe, S. Kuzuoka, and V. Y. F. Tan, "Nonasymptotic and second-order achievability bounds for coding with side-information," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1574–1605, April 2015.

[9] C. T. Li and V. Anantharam, "A unified framework for one-shot achievability via the poisson matching lemma," *IEEE Transactions on Information Theory*, vol. 67, no. 5, pp. 2624–2651, 2021.

[10] 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*. IEEE, 2013, pp. 1849–1853.

[11] J. Liu, P. Cuff, and S. Verdú, "Resolvability in E_γ with applications to lossy compression and wiretap channels," in *2015 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2015, pp. 755–759.

[12] ——, "One-shot mutual covering lemma and marton's inner bound with a common message," in *2015 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2015, pp. 1457–1461.

[13] 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.

[14] S.-H. Lee and S.-Y. Chung, "A unified random coding bound," *IEEE Transactions on Information Theory*, vol. 64, no. 10, pp. 6779–6802, 2018.

[15] S. H. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung, "Noisy network coding," *IEEE Transactions on Information Theory*, vol. 57, no. 5, pp. 3132–3152, 2011.

[16] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge University Press, 2011.

[17] E. C. Van Der Meulen, "Three-terminal communication channels," *Advances in applied Probability*, vol. 3, no. 1, pp. 120–154, 1971.

[18] T. Cover and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on information theory*, vol. 25, no. 5, pp. 572–584, 1979.

[19] Y.-H. Kim, "Coding techniques for primitive relay channels," in *Proc. Forty-Fifth Annual Allerton Conf. Commun., Contr. Comput*, 2007, p. 2007.

[20] M. Mondelli, S. H. Hassani, and R. Urbanke, "A new coding paradigm for the primitive relay channel," *Algorithms*, vol. 12, no. 10, p. 218, 2019.

[21] A. El Gamal, A. Gohari, and C. Nair, "Achievable rates for the relay channel with orthogonal receiver components," in *2021 IEEE Information Theory Workshop (ITW)*. IEEE, 2021, pp. 1–6.

[22] ——, "A strengthened cutset upper bound on the capacity of the relay channel and applications," *IEEE Transactions on Information Theory*, vol. 68, no. 8, pp. 5013–5043, 2022.

[23] A. El Gamal and N. Hassanpour, "Relay-without-delay," in *Proceedings. International Symposium on Information Theory, 2005. ISIT 2005.* IEEE, 2005, pp. 1078–1080.

[24] A. El Gamal, N. Hassanpour, and J. Mammen, "Relay networks with delays," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3413–3431, 2007.

[25] 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.

[26] A. 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, 1976.

[27] 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.

[28] H. Yamamoto, "Wyner-ziv theory for a general function of the correlated sources (corresp.)," *IEEE Transactions on Information Theory*, vol. 28, no. 5, pp. 803–807, 1982.

[29] R. Ahlswede, "Multi-way communication channels," in *2nd Int. Symp. Inform. Theory, Tsahkadsor, Armenian SSR*, 1971, pp. 23–52.

[30] H. Liao, "Multiple access channels," Ph.D. dissertation, University of Hawaii, Honolulu, HI, 1972.

[31] 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.

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

[33] 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, 2018.

[34] C. H. Bennett, P. W. Shor, J. Smolin, and A. V. Thapliyal, "Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem," *IEEE Trans. Inf. Theory*, vol. 48, no. 10, pp. 2637–2655, 2002.

[35] P. Cuff, "Distributed channel synthesis," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7071–7096, 2013.

[36] P. Cuff, H. Permuter, and T. M. Cover, "Coordination capacity," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4181–4206, Sept 2010.

[37] 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.

[38] C. T. Li, "An automated theorem proving framework for information-theoretic results," *IEEE Transactions on Information Theory*, vol. 69, no. 11, pp. 6857–6877, 2023.

Appendix

A. Proof of Lemma 2

We have

$$\begin{aligned}
&\mathbf{E}\left[\frac{1}{Q^{\mathbf{U}}_{V,U}(v,\mathbf{U}_P)}\,\middle|\,\mathbf{U}_P\right]\\
&\stackrel{(a)}{=} \mathbf{E}\left[\frac{\mathbf{U}^{-1}_{Q_{U|V(\cdot|v)}}(\mathbf{U}_P)\sum_{i=1}^{|\mathcal{U}|} i^{-1}}{Q_V(v)}\,\middle|\,\mathbf{U}_P\right]\\
&\stackrel{(b)}{\le} \frac{\sum_{i=1}^{|\mathcal{U}|} i^{-1}}{Q_V(v)}\left(\frac{P(\mathbf{U}_P)}{Q_{U|V}(\mathbf{U}_P|v)}+1\right)\\
&\stackrel{(c)}{\le} (\ln|\mathcal{U}|+1)\,\frac{1}{Q_V(v)}\left(\frac{P(\mathbf{U}_P)}{Q_{U|V}(\mathbf{U}_P|v)}+1\right).
\end{aligned}$$

where (a) is by Definition 1, (b) is by Lemma 1 and (c) is by $\sum_{i=1}^{n} i^{-1} \le \int_1^n x^{-1}dx + 1 = \ln n + 1$.

B. Proof of Theorem 3 and Theorem 4

We first generate N independent exponential processes $\mathbf{U}_i$ for $i \in [N]$ according to Section II, which serve as the random codebooks. Each node i will perform two steps: the *decoding step* and the *encoding step*.

We describe the decoding step at node i. The node observes Y_i and wants to decode $\overline{U}'_i = (U_{a_{i,j}})_{j\in[d'_i]}$, while utilizing $(U_{a_{i,j}})_{j\in[d'_i+1..d_i]}$ by non-unique decoding. For the sake of notational simplicity, we omit the subscript i and write $d = d_i$, $d' = d'_i$, $a_k = a_{i,k}$, $\overline{U}_{\mathcal{S}} = \overline{U}_{i,\mathcal{S}} = (U_{a_{i,j}})_{j\in\mathcal{S}}$, $\overline{\mathbf{U}}_k := \mathbf{U}_{a_{i,k}}$. For each $j = 1,\ldots,d'$, the node will perform soft decoding via the exponential process refinement (see Section II) on $\overline{U}_d$, and then on $\overline{U}_{d-1}$, and so on up to $\overline{U}_{j+1}$, and then use all the distributions obtained to decode $\overline{U}_j$ uniquely using the Poisson functional representation. For example, when $d = 3$, $d' = 2$, the decoding process will be: $\overline{U}_3$ (soft), $\overline{U}_2$ (soft), $\overline{U}_1$ (unique), $\overline{U}_3$ (soft), $\overline{U}_2$ (unique). The choice of the sequence $a_{i,k}$ controls the decoding ordering of the random variables. The goal is to obtain the decoded variables $\hat{\overline{U}}_1,\ldots,\hat{\overline{U}}_{d'}$ that equal $\overline{U}_1,\ldots,\overline{U}_{d'}$ with high probability.

More precisely, for $j = 1,\ldots,d'$, the node computes the decoded variable $\hat{\overline{U}}_j \in \overline{\mathcal{U}}_j$ by first computing the joint distributions $Q^{(j)}_{\overline{U}_{[k..d]}}$ over $\overline{\mathcal{U}}_k\times\ldots\times\overline{\mathcal{U}}_d$ for $k = d, d-1,\ldots,j+1$ recursively using the exponential process refinement as

$$Q^{(j)}_{\overline{U}_{[k..d]}} := \left(Q^{(j)}_{\overline{U}_{[k+1..d]}} P_{\overline{U}_k|\overline{U}_{[k+1..d]},\overline{U}_{[j-1]},Y_i}(\cdot\,|\,\cdot,\hat{\overline{U}}_{[j-1]},Y_i)\right)^{\overline{\mathbf{U}}_k},$$

i.e., first compute the semidirect product between $Q^{(j)}_{\overline{U}_{[k+1..d]}}$ and the conditional distribution $P_{\overline{U}_k|\overline{U}_{[k+1..d]},\overline{U}_{[j-1]},Y_i}(\cdot\,|\,\cdot,\hat{\overline{U}}_{[j-1]},Y_i)$ (computed using the ideal joint distribution of X^N,Y^N,U^N) to obtain a distribution over $\overline{\mathcal{U}}_k\times\ldots\times\overline{\mathcal{U}}_d$, and then refine it by $\overline{\mathbf{U}}_k$ using Definition 1. For the base case, we assume $Q^{(j)}_{\overline{U}_{[d+1..d]}}$ is the degenerate distribution. After we have computed $Q^{(j)}_{\overline{U}_{[j+1..d]}}$, we can obtain $\hat{\overline{U}}_j$ using the Poisson functional representation (1) as $\hat{\overline{U}}_j = (\overline{\mathbf{U}}_j)_{\tilde{Q}^{(j)}_{\overline{U}_j}}$, where $\tilde{Q}^{(j)}_{\overline{U}_j}$ is the $\overline{U}_j$-marginal of

$$Q^{(j)}_{\overline{U}_{[j+1..d]}} P_{\overline{U}_j|\overline{U}_{[j+1..d]},\overline{U}_{[j-1]},Y_i}(\cdot\,|\,\cdot,\hat{\overline{U}}_{[j-1]},Y_i). \tag{9}$$

The node repeats this process for $j = 1,\ldots,d'$ to obtain $\hat{\overline{U}}'_i = (\hat{\overline{U}}_1,\ldots,\hat{\overline{U}}_{d'})$.

We then describe the encoding step at node i. It uses the Poisson functional representation (see Section II) to obtain

$$U_i = (\mathbf{U}_i)_{P_{U_i|Y_i,\overline{U}'_i}(\cdot|Y_i,\hat{\overline{U}}'_i)}. \tag{10}$$

Finally, it generates X_i from the conditional distribution $P_{X_i|Y_i,U_i,\overline{U}'_i}(\cdot|Y_i,U_i,\hat{\overline{U}}'_i)$.

For the error analysis, we create a fictitious "ideal network" (with N "ideal nodes") that is almost identical to the actual network. The only difference is that the ideal node i uses the true $\overline{U}'_i$ (supplied by a genie) instead of the decoded $\hat{\overline{U}}'_i$ for the encoding step. The random variables induced by the ideal network will have the same distribution as the ideal distribution of X^N,Y^N,U^N in Theorem 3. Hence, we assume X^N,Y^N,U^N are induced by the ideal network. We couple the channels in the ideal network and the channels in the actual network, such that $Y_i = \tilde{Y}_i$ if $(X^{i-1},Y^{i-1}) = (\tilde{X}^{i-1},\tilde{Y}^{i-1})$ (i.e., the "channel noises" in the two networks are the same). If none of the actual nodes makes an error (i.e., $\hat{\overline{U}}'_i = \overline{U}'_i$ for all i), the actual network would coincide with the ideal network, and $(\tilde{X}^N,\tilde{Y}^N) = (X^N,Y^N)$. We consider the error probability conditional on $A := (X^N,Y^N,U^N)$:

$$F := \mathbf{P}\big(\exists i:\ \hat{\overline{U}}'_i = \overline{U}'_i\,\big|\,A\big).$$

Note that F is a random variable and is a function of $A = (X^N,Y^N,U^N)$. We have

$$\begin{aligned}
F &= \mathbf{P}\big(\exists i\in[N], j\in[d'_i]:\ \hat{\overline{U}}_{i,j} \ne \overline{U}_{i,j}\,\big|\,A\big)\\
&= \sum_{i=1}^{N}\sum_{j=1}^{d'_i}\mathbf{P}\Big(\hat{\overline{U}}'_{[i-1]} = \overline{U}'_{[i-1]},\ \hat{\overline{U}}_{i,[j-1]} = \overline{U}_{i,[j-1]},\\
&\qquad\qquad \hat{\overline{U}}_{i,j} \ne \overline{U}_{i,j}\,\big|\,A\Big).
\end{aligned}$$

For the term inside the summation (which is the probability that the first error we make is at $\overline{U}_{i,j}$), by (9), (10) and the Poisson matching lemma [9] (we again omit the subscripts i as in the description of the decoding step, e.g., we write $\overline{U}_j = \overline{U}_{i,j} = U_{a_j} = U_{a_{i,j}}$; we also simply write $P(\overline{U}_j|Y_{a_j},\overline{U}'_{a_j}) =$

$P_{\overline{U}_j|Y_{a_j},\overline{U}'_{a_j}}(\overline{U}_j|Y_{a_j},\overline{U}'_{a_j})$), we have

$$
\begin{aligned}
&\mathbf{P}\big(\hat{\overline{U}}'_{[i-1]} = \overline{U}'_{[i-1]},\, \hat{\overline{U}}_{i,[j-1]} = \overline{U}_{i,[j-1]},\, \hat{\overline{U}}_{i,j} \neq \overline{U}_{i,j} \,\big|\, A\big) \\
&\le \mathbf{E}\bigg[\frac{P(\overline{U}_j|Y_{a_j},\overline{U}'_{a_j})}{Q^{(j)}(\overline{U}_{[j+1..d]})P(\overline{U}_j\,|\,\overline{U}_{[j+1..d]},\,\overline{U}_{[j-1]},Y_i)}\,\bigg|\,A\bigg] \\
&\overset{(a)}{\le} \mathbf{E}\bigg[\frac{P(\overline{U}_j|Y_{a_j},\overline{U}'_{a_j})}{P(\overline{U}_j\,|\,\overline{U}_{[j+1..d]},\,\overline{U}_{[j-1]},Y_i)} \\
&\qquad \cdot \mathbf{E}\bigg[\frac{1}{Q^{(j)}(\overline{U}_{[j+1..d]})}\,\bigg|\,\overline{U}_{[d]},Y_i,Y_{a_j},\overline{U}'_{a_j},\overline{\mathbf{U}}_{[j+1..d]}\bigg]\,\bigg|\,A\bigg] \\
&\overset{(b)}{\le} \mathbf{E}\bigg[\frac{P(\overline{U}_j|Y_{a_j},\overline{U}'_{a_j})}{P(\overline{U}_j\,|\,\overline{U}_{[j+1..d]},\,\overline{U}_{[j-1]},Y_i)}(\ln|\overline{\mathcal{U}}_{j+1}|+1) \\
&\quad \cdot \frac{1}{Q^{(j)}(\overline{U}_{[j+2..d]})}\bigg(\frac{P(\overline{U}_{j+1}|Y_{a_{j+1}},\overline{U}'_{a_{j+1}})}{P(\overline{U}_{j+1}|\overline{U}_{[j+2..d]},\overline{U}_{[j-1]},Y_i)}+1\bigg)\,\bigg|\,A\bigg] \\
&\overset{(c)}{\le} \mathbf{E}\bigg[\frac{P(\overline{U}_j|Y_{a_j},\overline{U}'_{a_j})}{P(\overline{U}_j\,|\,\overline{U}_{[j+1..d]},\,\overline{U}_{[j-1]},Y_i)} \\
&\quad \cdot \prod_{k=j+1}^{d'}(\ln|\overline{\mathcal{U}}_k|+1)\bigg(\frac{P(\overline{U}_k|Y_{a_k},\overline{U}'_{a_k})}{P(\overline{U}_k|\overline{U}_{[k+1..d]},\overline{U}_{[j-1]},Y_i)}+1\bigg)\,\bigg|\,A\bigg] \\
&= B_{i,j},
\end{aligned}
$$

where (a) is by Jensen's inequality, (b) is due to Lemma 2, (c) is by applying the same steps as (a) and (b) $d'-j-1$ times, and $\beta_{i,j}$ is given in (4). The proof of Theorem 3 is completed by noting that $\delta_{\mathrm{TV}}(P_{X^N,Y^N},P_{\tilde{X}^N,\tilde{Y}^N}) \le \mathbf{P}((X^N,Y^N) \neq (\tilde{X}^N,\tilde{Y}^N)) \le \mathbf{E}[F] = \mathbf{E}[\min\{F,1\}]$.

We now prove Theorem 4. Recall that the scheme we have constructed requires the public randomness W, which we have to fix in order to construct a deterministic coding scheme for Theorem 4. We have

$$
\begin{aligned}
&\mathbf{E}\big[\mathbf{P}\big((\tilde{X}^N,\tilde{Y}^N)\in\mathcal{E}\,\big|\,W\big)\big] \\
&= \mathbf{P}((\tilde{X}^N,\tilde{Y}^N)\in\mathcal{E}) \\
&\le \mathbf{P}\big((X^N,Y^N)\in\mathcal{E} \text{ or } (X^N,Y^N)\neq(\tilde{X}^N,\tilde{Y}^N)\big) \\
&= \mathbf{E}\big[\mathbf{P}\big((X^N,Y^N)\in\mathcal{E} \text{ or } (X^N,Y^N)\neq(\tilde{X}^N,\tilde{Y}^N)\,\big|\,A\big)\big] \\
&\le \mathbf{E}\big[\min\big\{\mathbf{1}\{(X^N,Y^N)\in\mathcal{E}\} \\
&\qquad + \mathbf{P}\big((X^N,Y^N)\neq(\tilde{X}^N,\tilde{Y}^N)\,\big|\,A\big),\,1\big\}\big] \\
&\le \mathbf{E}\big[\min\big\{\mathbf{1}\{(X^N,Y^N)\in\mathcal{E}\}+F,\,1\big\}\big].
\end{aligned}
$$

Therefore, there exists a value w such that $\mathbf{P}((\tilde{X}^N,\tilde{Y}^N)\in\mathcal{E}\,|\,W=w)$ satisfies the upper bound. Fixing the value of W to w gives a deterministic coding scheme.

C. Asymptotic Rate of Corollary 8

Applying the law of large numbers to Corollary 8, and Fourier-Motzkin elimination (using the PSITIP software [38]), we obtain the following asymptotic achievable rate:

$$
\min\left\{\begin{array}{l}
I(V;Y)+I(U,Y;X|V), \\
I(V;Y_{\mathrm{r}})+I(U,Y;X|V), \\
I(V,U;Y)+I(U,Y;X|V)-I(U;Y_{\mathrm{r}}|V), \\
I(V;Y_{\mathrm{r}})+I(U;Y|V)+I(U,Y;X|V)-I(U;Y_{\mathrm{r}}|V)
\end{array}\right\},
$$

where $(X,V,Y_{\mathrm{r}},U,X_{\mathrm{r}},Y) \sim P_{X,V}P_{Y_{\mathrm{r}}|X,V}P_{U|Y_{\mathrm{r}},V}\delta_{x_{\mathrm{r}}(Y_{\mathrm{r}},U,V)}P_{Y|X,Y_{\mathrm{r}},X_{\mathrm{r}}}$, subject to the constraint $I(U;Y_{\mathrm{r}}|V) \le I(U;Y|V)+I(U,Y;X|V)$.