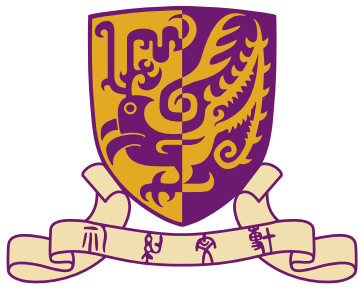
One-Shot Coding and Applications

Yanxiao Liu

Supervisors: Prof. Cheuk Ting Li and Prof. Raymond W. Yeung

Department of Information Engineering
The Chinese University of Hong Kong



Overview

Overview

My research focuses on one-shot information theory, which addresses scenarios in source coding and channel coding where the signal blocklength is 1.

In this presentation, I will describe three works:

- Part 1: One-shot coding over general noisy networks.
- Part 2: One-shot information hiding.
- Part 3: Compressing differential privacy mechanisms by one-shot channel simulation.

For these problems, we derive novel techniques that are rooted in the Poisson functional representation (Li and El Gamal, 2018), which also serves as a bridge between one-shot coding and differential privacy.

Liu, Y. & Li, C. T. (2024). One-Shot Coding over General Noisy Networks. ISIT 2024. Journal version has been submitted to IEEE Transactions on Information Theory.

Liu, Y. & Li, C. T. (2024). One-Shot Information Hiding. ITW 2024.

Liu, Y., Chen, W. N., Özgür, A. & Li, C. T. (2024). Universal exact compression of differentially private mechanisms. NeurIPS 2024.

Publications

- Nonasymptotic Oblivious Relaying and Variable-Length Noisy Lossy Source Coding. **Yanxiao Liu**, Sepehr Heidari Advary, Cheuk Ting Li. ISIT 2025.
- (Part 1 of the presentation) One-Shot Coding over General Noisy Networks. **Yanxiao Liu**, Cheuk Ting Li. ISIT 2024. Journal version has been submitted to IEEE Transactions on Information Theory.
- (Part 2 of the presentation) One-Shot Information Hiding. **Yanxiao Liu**, Cheuk Ting Li. ITW 2024. Journal version to be submitted.
- (Part 3 of the presentation) Universal Exact Compression of Differentially Private Mechanisms. **Yanxiao Liu**, Wei-Ning Chen, Ayfer Özgür, Cheuk Ting Li. NeurIPS 2024.
- Weighted Parity-Check Codes for Channels with State and Asymmetric Channels. Chih Wei Ling*, **Yanxiao Liu***, Cheuk Ting Li. ISIT 2022 and Transactions on Information Theory, August 2024.
- Wireless Network Scheduling with Discrete Propagation Delays: Theorems and Algorithms. Shenghao Yang, Jun Ma, **Yanxiao Liu**. Transactions on Information Theory, March 2024.
- Reliable Throughput of Generalized Collision Channel Without Synchronization. Yijun Fan, **Yanxiao Liu**, Yi Chen, Shenghao Yang, and Raymond W. Yeung. ISIT 2023.
- Continuity of Link Scheduling Rate Region for Wireless Networks with Propagation Delays. Yijun Fan, **Yanxiao Liu**, Shenghao Yang. ISIT 2022.
- Rate Region of Scheduling a Wireless Network with Discrete Propagation Delays. Jun Ma, **Yanxiao Liu**, Shenghao Yang. INFOCOM 2021.
- Joint Scheduling and Multiflow Maximization in Wireless Networks. **Yanxiao Liu**, Shenghao Yang, Cheuk Ting Li. 2025. To be submitted.

Outline

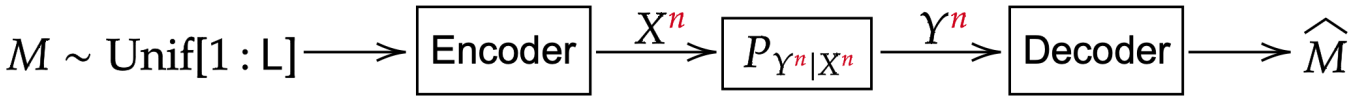
Background

One-Shot Information Theory

What do we mean by one-shot information theory and why we study it?

- In conventional Shannon theory, we study information limits assuming the signal blocklength n approaches infinity, but in practice it never does.
  - For large n, the decoder must wait for a long delay.
  - In modern applications (e.g., IoT), packets can even be very short!
- Finite-blocklength regime (Polyanskiy et al., 2010): n is finite.
- One-shot information theory: the extreme case where $n = 1$.
  - Sources and channels are *arbitrary* and used *once*, not necessarily memoryless.
  - Point-to-point: (Feinstein, 1954; Shannon, 1957; Hayashi, 2009).
  - Multiuser case: (Verdú, 2012; Yassaee et al., 2013; Watanabe et al., 2015).
- Clarification:
  1. Only 1 bit is sent?
     1. No! We transmit *one symbol* X; it can substituted by a sequence X^n or anything.
  2. Too troublesome to prove?
     1. No! With appropriate tools, the coding schemes and proofs can be even simpler.
- Goal: one-shot achievability results that can recover existing asymptotic results when applied to memoryless sources and channels.

One-Shot Information Theory



From Asymptotic Capacity to Nonasymptotic Behavior

- From Shannon, when $n \to \infty$, the channel capacity is $C = \max_{P_X} I(X;Y)$;
- What if the blocklength $n = 1$?

- What kind of guarantees are we looking for?
- Example: Dependence testing bound by (Polyanskiy et al., 2010):

$$\mathbf{P}\{M \neq \hat{M}\} \leq \mathbf{E}\left[\min\left\{\frac{\mathsf{L}-1}{2} \cdot 2^{-\iota_{X;Y}(X;Y)}, 1\right\}\right], \tag{1}$$

where $\iota_{X;Y}(X;Y) := \log(P_{X|Y}(x|y)/P_X(x))$ is the **information density** and we write $\iota(X;Y)$ when the context is clear; $I(X;Y) = \mathbf{E}[\iota(X;Y)]$.

Asymptotic Capacity:

1. Let $\mathsf{L} = 2^{nR}$, assuming the channel is memoryless, we know $P_{Y^n|X^n}(y^n|x^n) = \prod_{i=1}^n P_{Y|X}(y_i|x_i)$ in the absense of feedback.
2. Apply to (1), we get $P_e \leq \mathbf{E}\left[\min\{2^{nR - \sum_{i=1}^n \iota(X_i;Y_i)}, 1\}\right]$.
3. Let $n \to \infty$, by the law of large numbers, $\frac{1}{n}\sum_{i=1}^n \iota(X_i;Y_i) \approx I(X;Y)$.
4. Hence $P_e \to 0$ if $R < I(X;Y)$.

Poisson Functional Representation

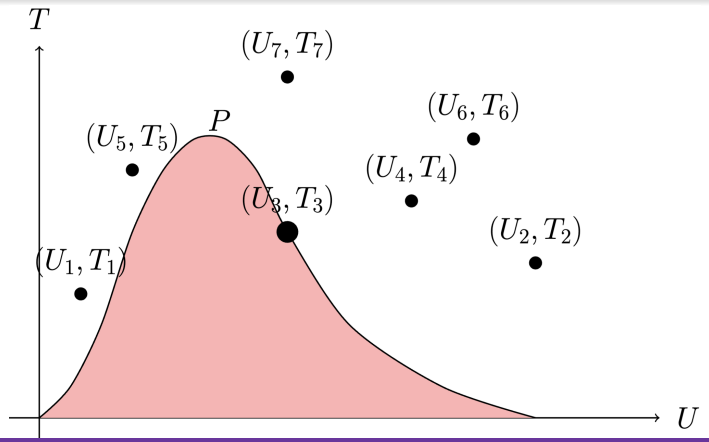
Poisson Functional Representation (Li and El Gamal (2018))

- Let $(T_i)_i$ be a Poisson process with rate 1, i.e., $T_1, T_2 - T_1, \ldots \sim \mathrm{Exp}(1)$.
- Let $(U_i)_i \overset{iid}{\sim} \mu$ be a sequence independent of $(T_i)_i$.
- Denote $\mathbf{U} := (U_i, T_i)_i$, which can be viewed as a "soft codebook".
- Fix a distribution P over $\mathcal{U}$ s.t. $P \ll \mu$ (P is absolutely continuous w.r.t μ).
- The **Poisson functional representation** selects[a]

$$\mathbf{U}_P := U_K, \quad \text{where} \quad K = \mathrm{argmin}_i \left(T_i \cdot \left(\frac{\mathrm{d}P}{\mathrm{d}\mu}(U_i) \right)^{-1} \right),$$

where $\frac{\mathrm{d}P}{\mathrm{d}\mu}(\cdot)$ is the Radon-Nikodym derivative.

[a] The "marked" Poisson process $\mathbf{U}$ supports a "query operation": with input P, it outputs one sample $\mathbf{U}_P \sim \mu$.



Poisson Functional Representation

When P is over a discrete space $\mathcal{U}$, we can use exponential random variables.

Poisson Functional Representation (PFR) for discrete $\mathcal{U}$

- For a finite set $\mathcal{U}$, let $\mathbf{U} := (Z_u)_{u\in\mathcal{U}}$ be i.i.d. $\mathrm{Exp}(1)$ random variable.
- Given a distribution P over $\mathcal{U}$, the **Poisson functional representation** is:

$$\mathbf{U}_P := \mathrm{argmin}_u \frac{Z_u}{P(u)}. \tag{2}$$

- PFR was introduced for one-shot channel simulation and some compression tasks in information theory (Li and El Gamal, 2018).
- Part 1 and 2: $\mathbf{U}_P$ can be viewed as an encoder or a decoder, with input P.
- Part 3: $\mathbf{U}_P$ "simulates" a channel P, i.e., $\mathbf{U}_P \sim P$ (properties of exponential random variables).
- PFR has been utilized in studying minimax learning (Li et al., 2020), neural estimation (Lei et al., 2023), reinforcement learning (Kobus and Gündüz, 2025) and many other settings.

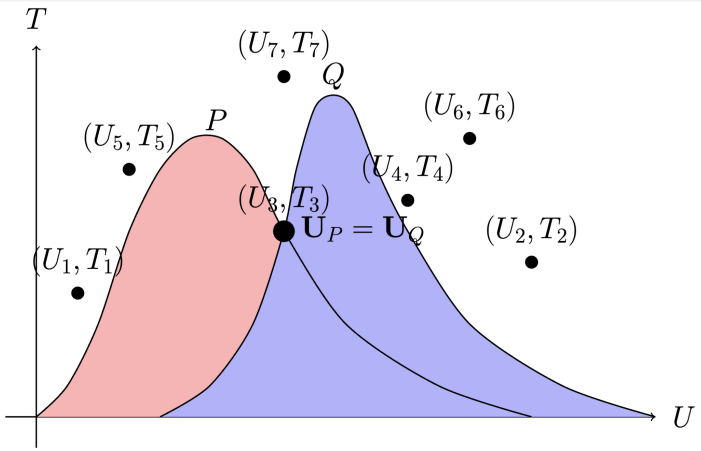
Poisson Matching Lemma

- Poisson Matching Lemma (Li and Anantharam, 2021) bounds the probability of mismatch between the PFRs applied on different distributions.

Poisson Matching Lemma (PML)

- $\mathbf{U}_P(1), \ldots, \mathbf{U}_P(|\mathcal{U}|)$ are elements of $\mathcal{U}$ ascendingly sorted by $Z_u/P(u)$.
- Let $\mathbf{U}_P^{-1} : \mathcal{U} \to [|\mathcal{U}|]$ be the inverse function of $i \mapsto \mathbf{U}_P(i)$.
- For distributions P, Q over $\mathcal{U}$, we have the following almost surely:

$$\mathbf{E}\left[\mathbf{U}_Q^{-1}(\mathbf{U}_P) \,\middle|\, \mathbf{U}_P\right] \leq \frac{P(\mathbf{U}_P)}{Q(\mathbf{U}_P)} + 1.$$



PML has been studied in hypothesis testing (Guo et al., 2024), unequal message protection (Khisti et al., 2024) and secret key generation (Hentilä et al., 2024).

Poisson Matching Lemma on Channel Coding

Example (channel coding): how is the PML used (Li and Anantharam, 2021)?

1. Let $P := P_X \times \delta_m$, $Q := P_{X|Y}(\cdot|Y) \times P_M$ and δ_m denote $\mathbf{P}(M = m) = 1$.
2. Encoder observes $M \sim \mathrm{Unif}[\mathsf{L}]$ and sends the X-component of $\mathbf{U}_P$ to $P_{Y|X}$;
3. Decoder observes Y and recovers the M-component of $\mathbf{U}_Q$.

$$
\begin{aligned}
P_e &\le \mathbf{P}\left((X, M) \neq \mathbf{U}_{P_{X|Y}(\cdot|Y)\times P_M}\right) \\
&\le \mathbf{E}\left[\min\left\{\mathbf{P}\left(\mathbf{U}_{P_X\times\delta_m} \neq \mathbf{U}_{P_{X|Y}(\cdot|Y)\times P_M} \middle| \mathbf{U}_{P_X\times\delta_m}\right), 1\right\}\right] \\
&= \mathbf{E}\left[\min\left\{\mathbf{P}\left(\mathbf{U}^{-1}_{P_{X|Y}(\cdot|Y)\times P_M}\left(\mathbf{U}_{P_X\times\delta_m}\right) > 1 \middle| \mathbf{U}_{P_X\times\delta_m}\right), 1\right\}\right] \\
&\overset{(a)}{\le} \mathbf{E}\left[\min\left\{\mathbf{E}\left[\mathbf{U}^{-1}_{P_{X|Y}(\cdot|Y)\times P_M}\left(\mathbf{U}_{P_X\times\delta_m}\right) - 1 \middle| \mathbf{U}_{P_X\times\delta_m}\right], 1\right\}\right] \\
&\overset{(b)}{\le} \mathbf{E}\left[\min\left\{P(\mathbf{U}_{P_X\times\delta_m})\big/Q(\mathbf{U}_{P_X\times\delta_m}), 1\right\}\right] \\
&= \mathbf{E}\left[\min\left\{\mathsf{L}\cdot 2^{-\iota(X;Y)}, 1\right\}\right].
\end{aligned}
$$

where (a) is by the Markov's inequality and (b) is by the PML.

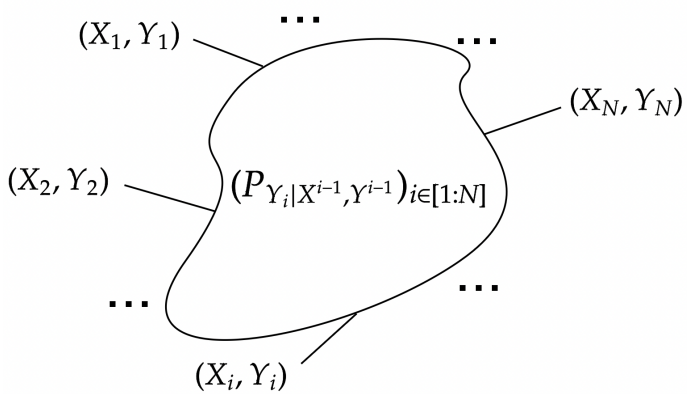
In the rest of this presentation, we will present novel techniques based on the PFR and PML, and extend their capabilities to different tasks.

Outline

A Unified One-Shot Coding Framework

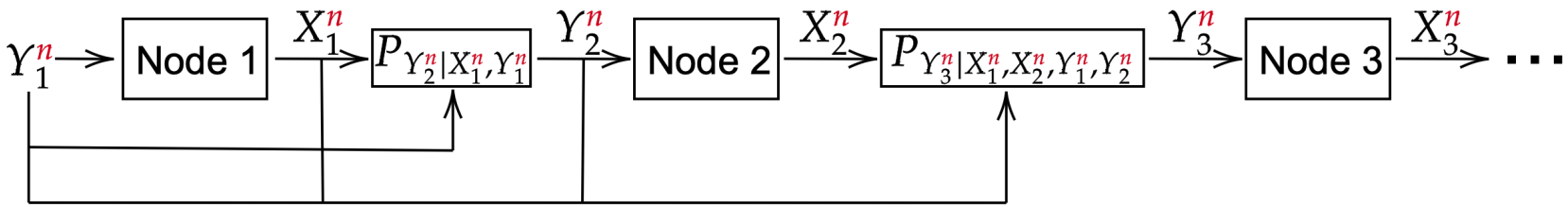
Overview

- Li and Anantharam (2021) have studied one-shot results of various settings.
- All the settings are simple (≤ 2 senders/receivers) and single-hop (no relays).
- We consider one-shot coding over general noisy acyclic networks.

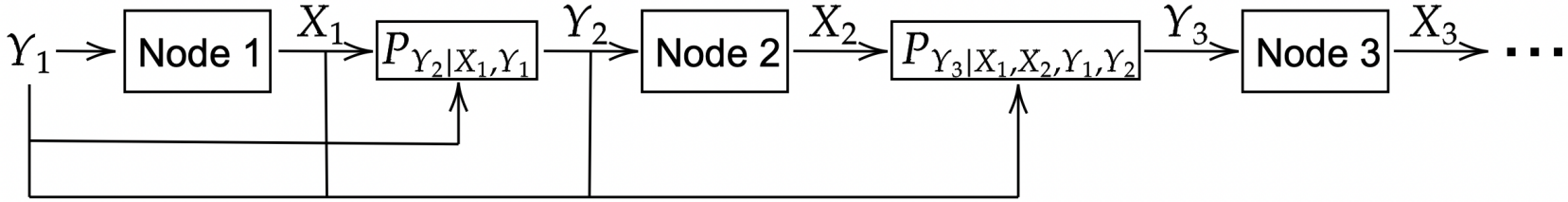


- When a large number of nodes are involved, it becomes difficult to use the original Poisson matching lemma for analysis.
- We look for a one-shot counterpart of:
  - unified random coding bound by (Lee and Chung, 2018);
  - noisy network coding (Lim et al., 2011).

A Unified One-Shot Coding Framework



- Lee and Chung (2018) introduced a unified random coding bound that
  1. unified and generalized many known relaying strategies;
  2. can yield *asymptotic* bounds without complicated error analysis.
- We introduce a unified one-shot coding framework over noisy networks
  - that is applicable to any combination of source coding, channel coding and coding for computing problems in one-shot scenarios.



- We recover many known one-shot achievability results.
- We derive novel one-shot achievability results for:
  - one-shot (primitive) relay channel: partial-decode-forward bound and compress-forward bound;
  - one-shot cascade multiterminal source coding.

New Technique: Exponential Process Refinement

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 $\mathbf{U}$:

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

  for all (v,u) in the support of $Q_{V,U}$.
- If $Q_{V,U}$ represents our "**prior distribution**" of (V,U), then the refinement $Q^{\mathbf{U}}_{V,U}$ is the updated "**posterior distribution**" after taking $\mathbf{U}$ into account.

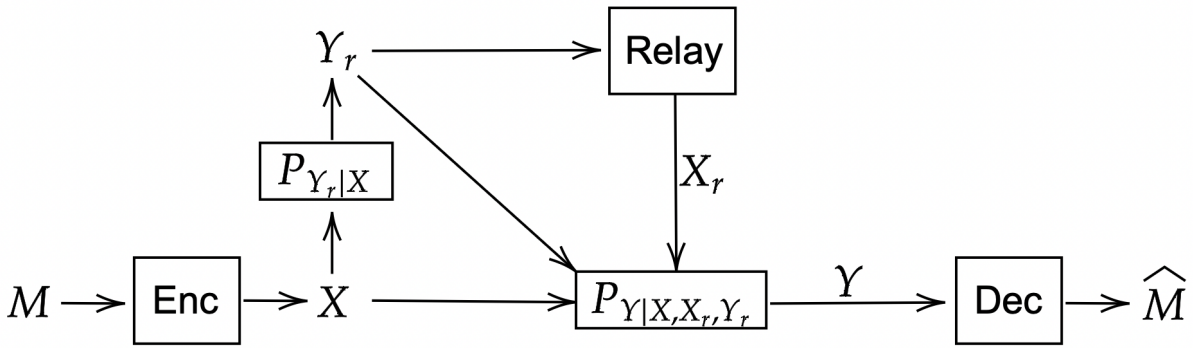
Exponential Process Refinement Lemma (EPRL)

- 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,

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

- It keeps track of the evolution of the "posterior probability" of the correct values of a large number of random variables through the refinement process.

One-Shot Relay Channel



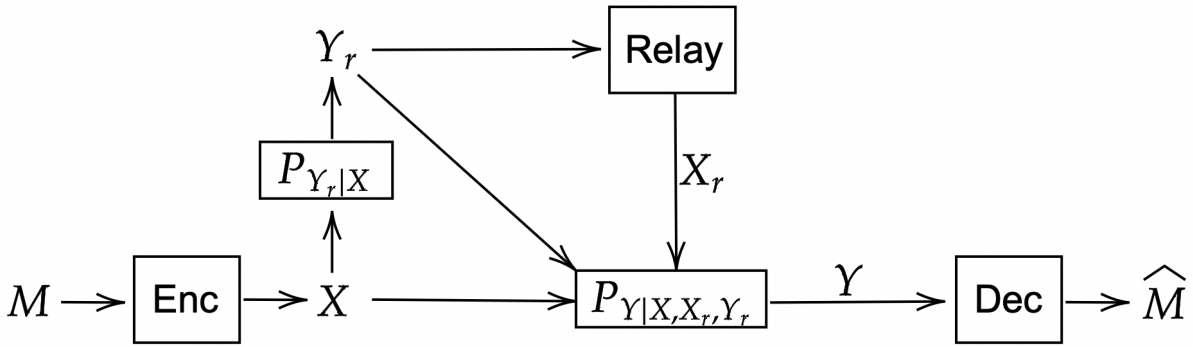
One-Shot Relay Channel

1. Encoder observes $M \sim \mathrm{Unif}[\mathrm{L}]$ and outputs X, which is sent to $P_{Y_\mathrm{r}|X}$.
2. Relay observes Y_r and outputs X_r.
3. $(X, X_\mathrm{r}, Y_\mathrm{r})$ is passed through the channel $P_{Y|X,X_\mathrm{r},Y_\mathrm{r}}$.
4. Decoder observes Y and recovers $\hat{M}$.

- One-shot case of relay-with-unlimited-look-ahead (El Gamal et al., 2007).
- "Best one-shot approximation" of the conventional relay channels (Van Der Meulen, 1971; Cover and Gamal, 1979).[1]

[1] One-shot settings cannot model "networks with causality", e.g., conventional relay channels.

One-Shot Relay Channel



Theorem (One-Shot Achievable Bound)

For any P_X, $P_{U|Y_\mathrm{r}}$, function $x_\mathrm{r}(y_\mathrm{r}, u)$, there is a coding scheme for the one-shot relay channel such that the error probability satisfies

$$P_e \le \mathbf{E}\left[\min\left\{\gamma \mathsf{L} 2^{-\imath(X;U,Y)}\left(2^{-\imath(U;Y)+\imath(U;Y_\mathrm{r})}+1\right), 1\right\}\right],$$

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

Proof

- "Random codebooks" $\mathbf{U}_1$, $\mathbf{U}_2$: independent exponential processes.
- Encoder: $U_1 = (\mathbf{U}_1)_{P_{U_1} \times \delta_M}$.
- Relay: $U_2 = (\mathbf{U}_2)_{P_{U_2|Y_\mathrm{r}}(\cdot|Y_\mathrm{r})}$, then outputs $X_\mathrm{r} = x_\mathrm{r}(Y_\mathrm{r}, U_2)$.

One-Shot Relay Channel

- Decoder observes Y, and refine $P_{U_2|Y}(\cdot|Y)$ to $Q_{U_2} := P_{U_2|Y}^{\mathbf{U}_2}$. By EPRL:

$$\mathbf{E}\left[\frac{1}{Q_{U_2}(U_2)}\middle| U_2, Y, Y_{\mathrm{r}}\right] \le (\ln|\mathcal{U}_2|+1)\left(\frac{P_{U_2|Y_{\mathrm{r}}}(U_2)}{P_{U_2|Y}(U_2)}+1\right). \tag{3}$$

  - Compute $Q_{U_2}P_{U_1|U_2,Y}$ over $\mathcal{U}_1\times\mathcal{U}_2$, and let its U_1-marginal be $\tilde{Q}_{U_1}$.
  - Let $\tilde{U}_1 = (\mathbf{U}_1)_{\tilde{Q}_{U_1}\times P_M}$, and output its M-component.
- Error analysis:

$$\begin{aligned}
&\mathbf{P}(\tilde{U}_1 \neq U_1 \,|\, X, Y_{\mathrm{r}}, U_2, X_{\mathrm{r}}, Y, M)\\
&\overset{\text{PML}}{\le} \mathbf{E}\left[\min\left\{\frac{P_{U_1}(U_1)\delta_M(M)}{P_{U_1|U_2,Y}(U_1|U_2,Y)Q_{U_2}(U_2)P_M(M)},1\right\}\middle| X, Y_{\mathrm{r}}, U_2, X_{\mathrm{r}}, Y, M\right]\\
&\overset{(a)}{\le} \min\left\{\mathsf{L}\frac{P_{U_1}(U_1)}{P_{U_1|U_2,Y}(U_1|U_2,Y)}(\ln|\mathcal{U}_2|+1)\left(\frac{P_{U_2|Y_{\mathrm{r}}}(U_2)}{P_{U_2|Y}(U_2)}+1\right),1\right\}\\
&= \min\left\{(\ln|\mathcal{U}_2|+1)\mathsf{L}2^{-\iota(X;U_2,Y)}\left(2^{-\iota(U_2;Y)+\iota(U_2;Y_{\mathrm{r}})}+1\right),1\right\},
\end{aligned}$$

  where (a) by Jensen's inequality, (3), and $\delta_M(M)=1, P_M(M)=1/\mathsf{L}$.

Implied asymptotic rate: $R \le I(X;U,Y) - \max\left\{I(U;Y_{\mathrm{r}}) - I(U;Y),\, 0\right\}$.

One-Shot Relay Channel: Partial-Decode-Forward Bound

Splitting the message $M \sim \mathrm{Unif}[\mathsf{L}]$ into $M_1 \sim \mathrm{Unif}[\mathsf{J}]$ and $M_2 \sim \mathrm{Unif}[\mathsf{L}/\mathsf{J}]$, we can let the relay decodes part of the message and provide the following bound.

Corollary (Partial-Decode-Forward Bound)

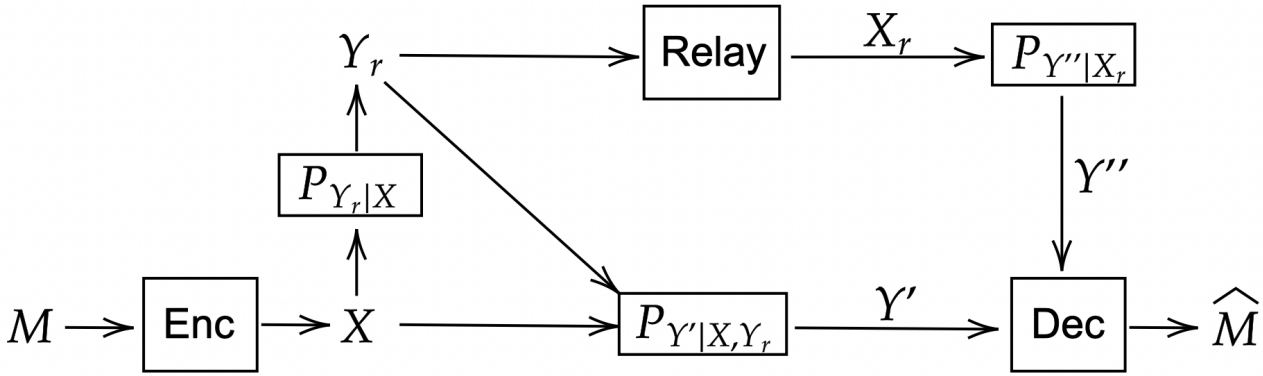
Fix any $P_{X,V}$, $P_{U|Y_\mathrm{r},V}$, function $x_\mathrm{r}(y_\mathrm{r},u,v)$, and J which is a factor of L. There exists a coding scheme for the one-shot relay channel with

$$P_e \le \mathbf{E}\Big[\min\Big\{\mathsf{J}2^{-\imath(V;Y_\mathrm{r})} + (\ln(\mathsf{J}|\mathcal{U}|)+1)(\ln(\mathsf{J}|\mathcal{V}|)+1)\mathsf{L}\mathsf{J}^{-1}2^{-\imath(X;U,Y|V)} \cdot \left(2^{-\imath(U;V,Y)+\imath(U;V,Y_\mathrm{r})}+1\right)\left(\mathsf{J}2^{-\imath(V;Y)}+1\right), 1\Big\}\Big],$$

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

It recovers the partial-decode-forward bounds for relay-with-unlimited-look-ahead (El Gamal et al., 2007) and primitive relay channels (Kim, 2007).

One-Shot Primitive Relay Channel



One-shot primitive relay channels (Kim, 2007): $Y = (Y', Y'')$ and $P_{Y|X,X_\mathrm{r},Y_\mathrm{r}} = P_{Y'|X,Y_\mathrm{r}} P_{Y''|X_\mathrm{r}}$ can be decomposed into two orthogonal components.

Theorem

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

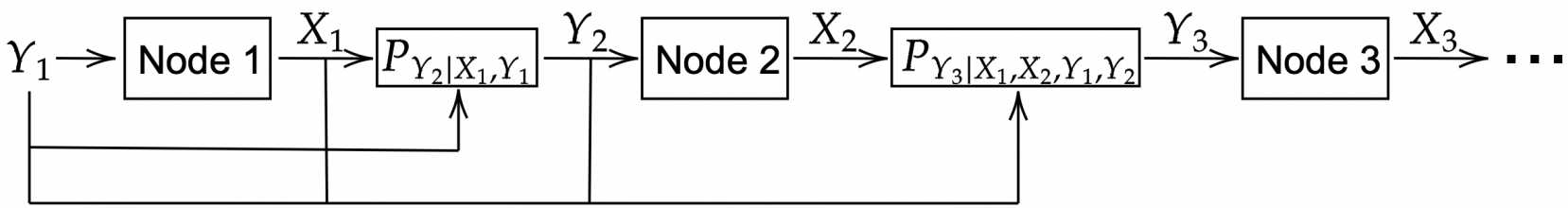
$$P_e \le \mathbf{E}\left[\min\left\{\left(\ln(|\mathcal{U}'||\mathcal{X}_\mathrm{r}|)+1\right)\mathsf{L}2^{-\imath(X;U',Y')}\left(2^{-\imath(X_\mathrm{r};Y'')+\imath(U';Y_\mathrm{r}|Y')}+1\right), 1\right\}\right],$$

$(X, Y_\mathrm{r}, U', Y') \sim P_X P_{Y_\mathrm{r}|X} P_{U'|Y_\mathrm{r}} P_{Y'|X,Y_\mathrm{r}}$ independent of $(X_\mathrm{r}, Y'') \sim P_{X_\mathrm{r}} P_{Y''|X_\mathrm{r}}$.

It recovers the asymptotic compress-and-forward bound (Kim, 2007):

$$R \le I(X; U', Y') - \max\{I(U'; Y_\mathrm{r}|Y') - \max_{P_{X_\mathrm{r}}} I(X_\mathrm{r}; Y''),\, 0\}$$

General Acyclic Discrete Network



General Acyclic Discrete Network (ADN)

- Nodes are labelled by $1, \ldots, N$.
- Node i observes $Y_i \in \mathcal{Y}_i$ and produces $X_i \in \mathcal{X}_i$.
- **ADN**: a collection of channels $\left(P_{Y_i|X^{i-1},Y^{i-1}}\right)_{i\in[N]}$.

X_i and Y_i can represent sources, states or messages in source and channel coding

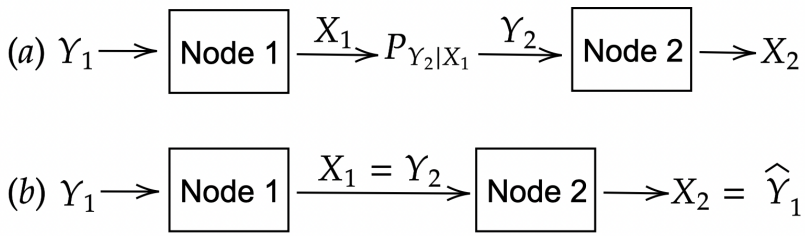


Figure 1: A unified view: (a) channel coding; (b) source coding.

- Channel coding: Message Y_1 is encoded by node 1 to X_1; node 2 sees Y_2 and outputs X_2.
- Lossless source coding: Y_1 is source, $X_1 = Y_2$ is description, X_2 is reconstruction.

General Acyclic Discrete Network

- $\tilde{X}_i, \tilde{Y}_i$: **actual** random variables from the coding scheme.
- X_i, Y_i: random variables following an **ideal** distribution.
  - Example 1 (channel coding): the ideal distribution is $Y_1 = X_2 \sim \mathrm{Unif}[\mathsf{L}]$ (decoding without error). If we ensure $\tilde{X}^2, \tilde{Y}^2$ is "close to" the ideal X^2, Y^2, it implies $\tilde{Y}_1 = \tilde{X}_2$ with high probability, giving a small error probability.
- Take an "error set" $\mathcal{E}$ that we do not want $(\tilde{X}^N, \tilde{Y}^N)$ to fall into.
  - Example 2 (channel coding): $\mathcal{E}$ is the set where $\tilde{Y}_1 \neq \tilde{X}_2$, i.e., an error occurs.
  - Example 3 (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.
- **Goal**: make $P_{\tilde{X}^N, \tilde{Y}^N}$ "approximately as good as" P_{X^N, Y^N}, i.e.,

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

  which can be guaranteed by ensuring the closeness in TV distance:

$$\|P_{X^N, Y^N} - P_{\tilde{X}^N, \tilde{Y}^N}\|_{\mathrm{TV}} \approx 0. \tag{4}$$

- With public randomness, (4) can be achieved; it can be viewed as a channel simulation (Cuff, 2013) or a coordination (Cuff et al., 2010) result.

Coding Scheme

Deterministic coding scheme $(f_i)_{i\in[N]}$

A sequence of functions $(f_i)_{i\in[N]}$, where $f_i : \mathcal{Y}_i \to \mathcal{X}_i$. For $i = 1, \ldots, N$:

- Encoding: $\tilde{X}_i = f_i(\tilde{Y}_i)$.
- $\tilde{Y}_i$ follows $P_{Y_i|X^{i-1},Y^{i-1}}$ conditional on $\tilde{X}^{i-1}, \tilde{Y}^{i-1}$.

Goal: $\mathbf{P}\big((\tilde{X}^N, \tilde{Y}^N) \in \mathcal{E}\big) \lesssim \mathbf{P}\big((X^N, Y^N) \in \mathcal{E}\big)$

To construct a deterministic coding scheme, we utllize a randomized coding scheme.

Public-randomness coding scheme $(P_W, (f_i)_{i\in[N]})$

1. Generate public randomness $W \in \mathcal{W}$ available to all nodes;
2. Encoding: $f_i : \mathcal{Y}_i \times \mathcal{W} \to \mathcal{X}_i$, $\tilde{X}_i = f_i(\tilde{Y}_i, W)$.

Goal: $\|P_{X^N,Y^N} - P_{\tilde{X}^N,\tilde{Y}^N}\|_{\mathrm{TV}} \approx 0$

If there is a good public-randomness coding scheme, then there is a good deterministic coding scheme by fixing the value of W.

Main Theorem

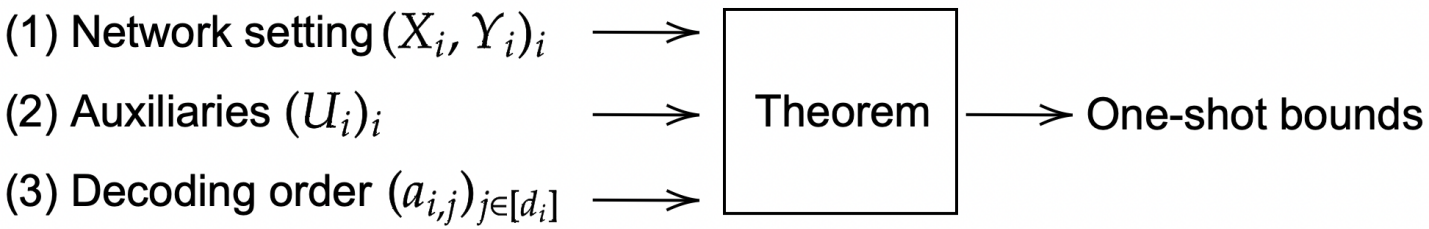
Theorem

Fix ADN $(P_{Y_i|X^{i-1},Y^{i-1}})_{i\in[N]}$. *For any collection of* $(a_{i,j})_{i\in[N],j\in[d_i]}$ *where for each* i, $(a_{i,j})_{j\in[d_i]}$ *is a sequence of distinct indices in* $[i-1]$, *any sequence* $(d_i')_{i\in[N]}$ *with* $0 \le d_i' \le d_i$ *and any collection of* $(P_{U_i|Y_i,\overline{U}_i'}, P_{X_i|Y_i,U_i,\overline{U}_i'})_{i\in[N]}$ *(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 joint distribution of* X^N, Y^N, U^N, *there exists a public-randomness coding scheme such that*

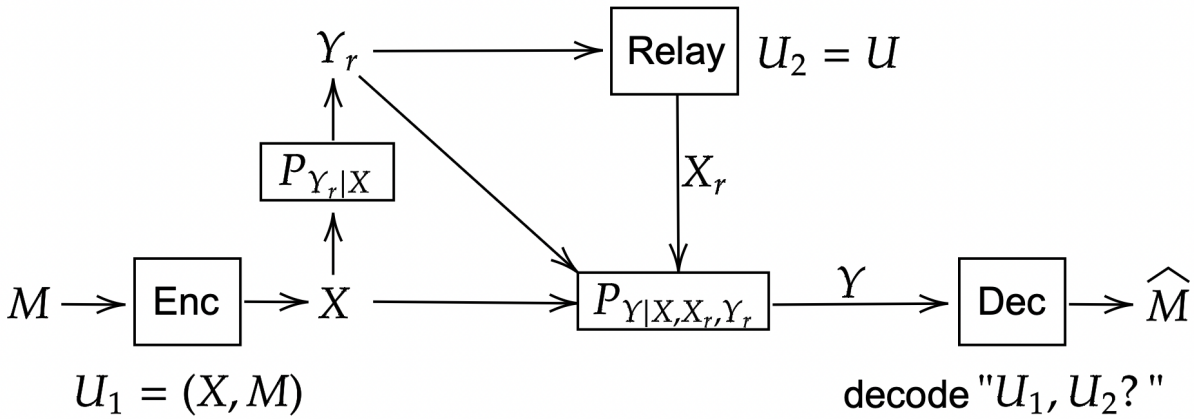
$$\|P_{X^N,Y^N} - P_{\tilde{X}^N,\tilde{Y}^N}\|_{\mathrm{TV}} \le \mathbf{E}\left[\min\left\{\sum_{i=1}^{N}\sum_{j=1}^{d_i'} B_{i,j},\, 1\right\}\right],$$

where $\gamma_{i,j} := \prod_{k=j+1}^{d_i}\left(\ln|\mathcal{U}_{a_{i,k}}|+1\right)$ *and*

$$B_{i,j} := \gamma_{i,j}\prod_{k=j}^{d_i}\left(2^{-\iota(\overline{U}_{i,k};\overline{U}_{i,[d_i]\setminus[j:k]},Y_i)+\iota(\overline{U}_{i,k};\overline{U}'_{a_{i,k}},Y_{a_{i,k}})}+\mathbf{1}\{k>j\}\right).$$



One-Shot Relay Channel



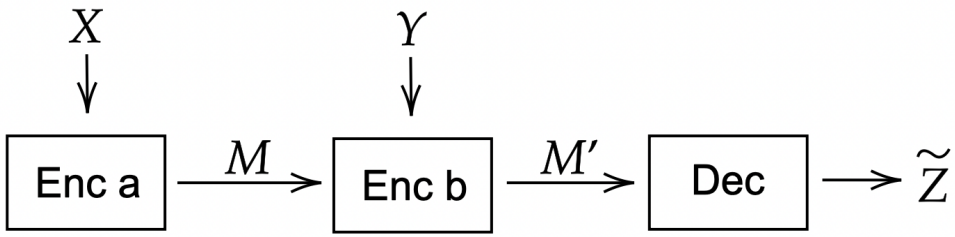
Main Theorem on One-Shot Relay Channel

1. **Network**:
   - Node 1 (encoder) has input $Y_1 = M$ and output $X_1 = X$.
   - Node 2 (relay) has input $Y_2 = Y_r$ and output $X_2 = X_r$.
   - Node 3 (decoder) has input $Y_3 = Y$ and output $X_3 = M$.
2. **Auxiliaries**: $U_1 = (X, M)$ and $U_2 = U$.
3. **Decoding order**: decode with order "U_1, U_2?", where $d_3' = 1$ and $d_3 = 2$.[a]
4. Applying our main theorem, we recover

$$P_e \leq (\ln |\mathcal{U}_2| + 1)\mathsf{L} 2^{-\iota(X;U_2,Y)} \left(2^{-\iota(U_2;Y)+\iota(U_2;Y_\mathrm{r})} + 1\right).$$

[a] "?" means the random variable is only utilized in non-unique decoding.

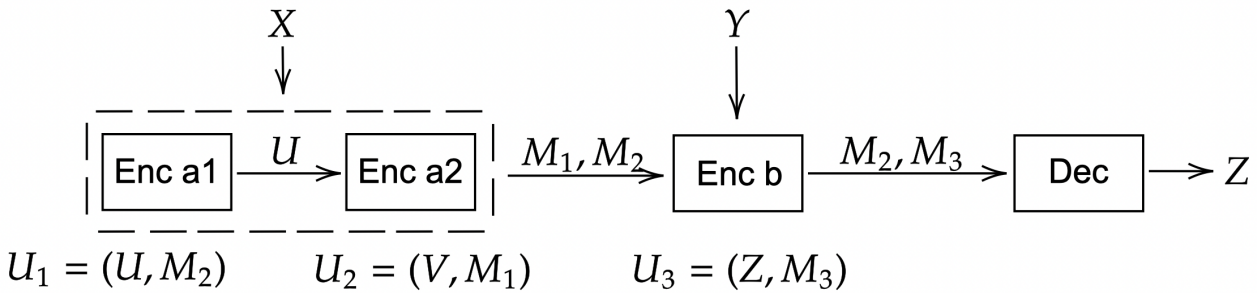
One-Shot Cascade Multiterminal Source Coding



One-Shot Cascade Multiterminal Source Coding (Cuff et al., 2009)

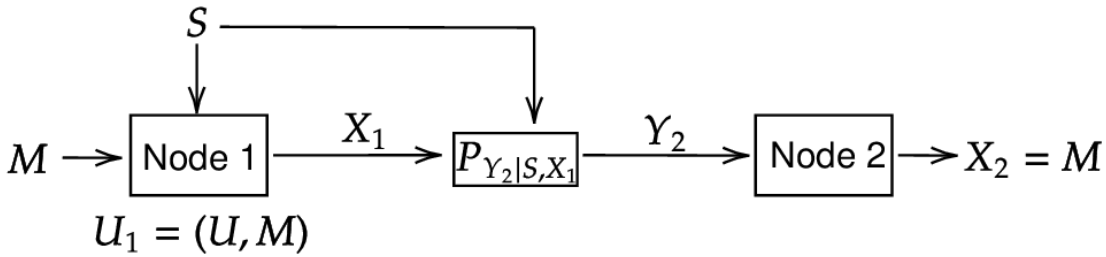
- Two sources $X, Y \sim P_{X,Y}$ are described by separate encoders.
- Encoder a observes X, sends $M \in [\mathsf{L}_1]$ to encoder b, which then creates $M' \in [\mathsf{L}_2]$ summarizing both sources and sends M' to the decoder.
- Decoder recovers $\tilde{Z}$ with the probability of excess distortion $P_e := \mathbf{P}\{d(X, Y, \tilde{Z}) > \mathsf{D}\}$, where $d : \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \to \mathbb{R}_{\geq 0}$ is a distortion measure.
- Applicable to scenarios where one needs to pass messages to neighbors in order to compute functions:
  - distributed data collection;
  - aggregating measurements in sensor networks;
  - federated computing.

One-Shot Cascade Multiterminal Source Coding



- Let $M_i \in [\mathsf{L}_i]$. We split the encoder a. We have a **network**:
  1. Encoder a1 (node 1) has input $Y_1 = X$ and output $X_1 = U$.
  2. Encoder a2 (node 2) has input $Y_2 = (U, X)$ and output $X_2 = M_1$.
  3. Encoder b (node 3) has input $Y_3 = (Y, M_1, M_2)$ and output $X_3 = M_3$.
  4. Decoder (node 4) has input $Y_4 = (M_2, M_3)$ and output $X_4 = Z$.
- **Auxiliaries**: $U_1 = (U, M_2)$, $U_2 = (V, M_1)$ and $U_3 = (Z, M_3)$.
  1. Encoder b (node 3) **decodes with the order** "U_2, U_1" where $d_3' = d_3 = 2$:
$$B_{3,1} = (\ln(|\mathcal{U}|\tilde{\mathsf{L}}_2) + 1)\left(\tilde{\mathsf{L}}_1^{-1}\tilde{\mathsf{L}}_2^{-1}2^{\iota(U,V;X|Y)} + \tilde{\mathsf{L}}_1^{-1}2^{-\iota(V;U,Y)+\iota(V;U,X)}\right),$$
$$B_{3,2} = \tilde{\mathsf{L}}_2^{-1}2^{-\iota(U;V,Y)+\iota(U;X)},$$
  2. Decoder (node 4) **decodes with the order** "U_3, U_1?" where $d_4' = 1, d_4 = 2$:
$$B_{4,1} = \left(\ln(|\mathcal{U}|\tilde{\mathsf{L}}_2) + 1\right)\tilde{\mathsf{L}}_3^{-1}2^{\iota(Z;V,Y|U)}\left(\tilde{\mathsf{L}}_2^{-1}2^{\iota(U;X)} + 1\right).$$
- One-shot Bound: $P_e \le \mathbf{E}\left[\min\left\{\mathbf{1}\{d(X,Y,Z) > \mathsf{D}\} + B_{3,1} + B_{3,2} + B_{4,1}, 1\right\}\right]$
- It recovers the best-known bound (local-computing-and-forwarding):
$$R_1 > I(X;U,V|Y), \quad R_2 > I(X;U) + I(Z;V,Y|U)$$
and $D > \mathbf{E}[d(X,Y,Z)]$, $X, Y, Z, U, V \sim P_X P_{Y|X} P_{U,V|X} P_{Z|Y,U,V}$.

Gelfand-Pinsker Problem



- **Network**: $Y_1 = (M, S)$, $Y_2 = Y$, $P_{Y_2|Y_1,X_1}$ is $P_{Y|S,X}$ and $X_2 = M$.
- **Auxiliary**: $U_1 = (U, M)$ for some U following $P_{U|S}$ given S.
- **Decoding order**: on node 2 decode "U_1".

Corollary (Gelfand and Pinsker (1980))

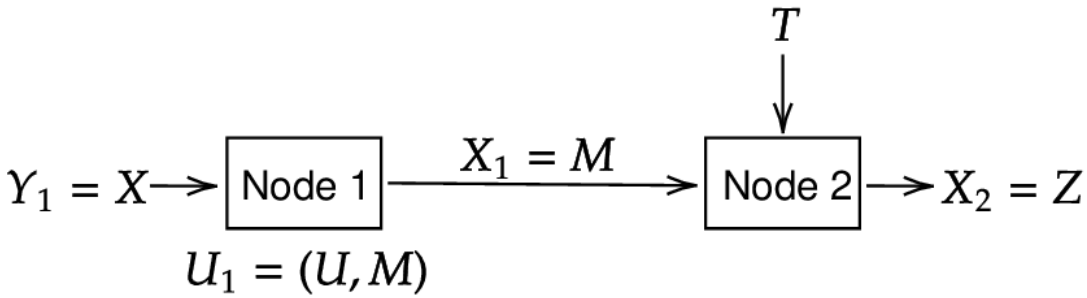
Fix $P_{U|S}$ and function $x : \mathcal{U} \times \mathcal{S} \to \mathcal{X}$. There exists a deterministic coding scheme for the channel $P_{Y|X,S}$ with $S \sim P_S, M \sim \mathrm{Unif}[\mathsf{L}]$ such that

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

where $S, U, X, Y \sim P_S P_{U|S} \delta_{x(U,S)} P_{Y|X,S}$.

Similar to the one-shot result in (Li and Anantharam, 2021), above implies the best known second order result of the Gelfand-Pinsker problem (Scarlett, 2015).

Wyner-Ziv Problem



- **Network**: $Y_1 = X$, $X_1 = M$, $Y_2 = (M, T)$ and $X_2 = Z$.
- **Auxiliary**: $U_1 = (U, M)$ for some U following $P_{U|X}$ given X.
- **Decoding order**: on node 2 decode "U_1".

Corollary (Wyner and Ziv (1976))

Fix $P_{U|X}$ and function $z : \mathcal{U} \times \mathcal{Y} \to \mathcal{Z}$. There exists a deterministic coding scheme with source $X \sim P_X$, side information $P_{T|X}$ and $M \in [\mathsf{L}]$ such that

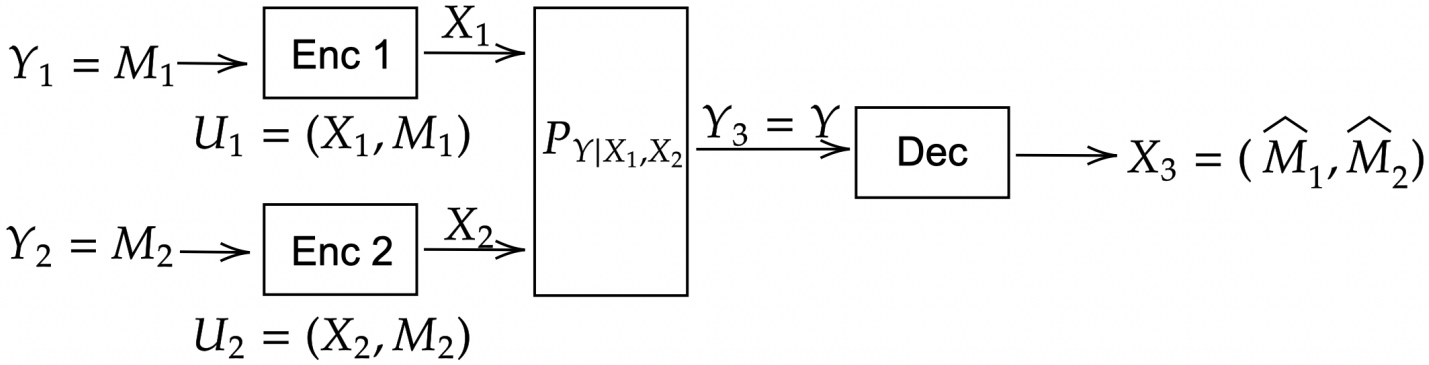
$$P_e \leq \mathbf{E}\left[\min\left\{\mathbf{1}\{d(X,Z) > \mathsf{D}\} + \mathsf{L}^{-1}2^{-\imath(U;T)+\imath(U;X)}, 1\right\}\right],$$

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

In coding for computing (Yamamoto, 1982) where node 2 lossily recovers $f(X, T)$:

$$P_e \leq \mathbf{E}[\min\{\mathbf{1}\{d(f(X,T),Z) > \mathsf{D}\} + \mathsf{L}^{-1}2^{-\imath(U;T)+\imath(U;X)}, 1\}].$$

Multiple Access Channel



- **Network**: $Y_1 = M_1$, $Y_2 = M_2$, $Y_3 = Y$ and $X_3 = (M_1, M_2)$.
- **Auxiliaries**: $U_1 = (X_1, M_1)$ and $U_2 = (X_2, M_2)$.
- **Decoding order**: on node 3 decode "U_2, U_1".

Corollary (Multiple Access Channel (Liao (1972), Ahlswede (1974)))

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 $M_j \sim \mathrm{Unif}[\mathsf{L}_j]$ for $j = 1, 2$ such that

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

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

Asymptotic region: $R_1 < I(X_1; Y|X_2)$, $R_2 < I(X_2; Y|X_1)$, $R_1 + R_2 < I(X_1, X_2; Y)$.

Summary

Summary

- We provide a **unified one-shot coding framework** for communication and compression over general acyclic noisy networks.
- We design a proof technique "**exponential process refinement lemma**" that can keep track of a large number of auxiliary random variables and greatly simplify the analysis.
- We provide **novel** one-shot achievability results for various settings.
- Our results can recover existing one-shot and asymptotic bounds on many settings.

Outline

Information Hiding

- In previous section, we considered a "multi-hop generalization" of the PML.
- Back to point-to-point channels, what if the channel itself is uncertain?

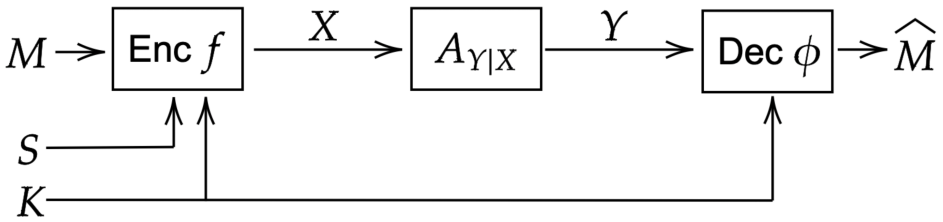


Figure 2: Information Hiding: $A_{Y|X}$ is chosen by an attacker from a set $\mathcal{A}$,

Information Hiding (Moulin and O'sullivan (2003))

- Game-theoretic setting: an encoder-decoder team is against an attacker:
  - **Encoder**: upon observing message $M \sim [\mathsf{L}]$, host signal $S \in \mathcal{S}$ and common randomness $K \in \mathcal{K}$ $(S, K \sim P_{S,K})$, it produces $X = f(S, K, M)$.
    - X is expected to "look like" S: $d_1(S, X)$ is small with $d_1 : \mathcal{S} \times \mathcal{X} \to [0, \infty)$.
  - **Attacker**: it chooses a channel $A_{Y|X} \in \mathcal{A}$ to destroy M.
    - Attacker knows the distributions (not the values) of S, M, K, and the code in use.
  - **Decoder**: upon observing Y, K, it recovers $\hat{M} = \phi(K, Y)$.
    - Decoder is uninformed of the attacker's strategy.
  - We bound the following worst case failure probability:

$$P_e := \sup_{A_{Y|X} \in \mathcal{A}} \mathbf{P}\Big(d_1(S, X) > \mathsf{D}_1 \ \text{ OR } \ M \neq \hat{M}\Big).$$

One-Shot Information Hiding

- By (Moulin and O'Sullivan, 2003), asymptotic hiding capacity was derived.
- Wide range of applications: watermarking, fingerprinting, steganography...

Our Contributions

1. We derive one-shot results that apply to any host distribution, and any class of attack channels (not memoryless or subject to distortion constraints).
2. Our techniques include the Poisson matching lemma together with a covering argument (Blackwell et al., 1959).
3. We recover the asymptotic capacity, hence give an alternative proof.
4. Unlike (Moulin and O'Sullivan, 2003) which assumed the decoder knows the attack channel, we let the decoder be uninformed of the attacker.[a]

[a] It was also dropped by (Somekh-Baruch and Merhav, 2004), in which K is a shared key of unlimited size independent of M, S that can be chosen as part of the code, but our side information K is given and cannot be changed.

One-Shot Achievability Results

- To account for all possible $A \in \mathcal{A}$, we need a penalty on the "size" of $\mathcal{A}$.
- For $\mathcal{A}$ with infinite cardinality, we find a finite $\tilde{\mathcal{A}} \subseteq \mathcal{A}$ such that every $A \in \mathcal{A}$ is close enough to some $\tilde{A} \in \tilde{\mathcal{A}}$ (Blackwell et al., 1959).
- Given a set of channels $\mathcal{A}$ from $\mathcal{X}$ to $\mathcal{Y}$, its ϵ-covering number is

$$N_\epsilon(\mathcal{A}) := \min\left\{|\tilde{\mathcal{A}}| : \tilde{\mathcal{A}} \subseteq \mathcal{A}, \sup_{A\in\mathcal{A}} \min_{\tilde{A}\in\tilde{\mathcal{A}}} \sup_{x\in\mathcal{X}} \|A_{Y|X}(\cdot|x) - \tilde{A}_{Y|X}(\cdot|x)\|_{\mathrm{TV}} \le \epsilon\right\}.$$

Theorem

Fix any $P_{U,X|S,K}$ and channel $\hat{A}_{Y|X}$. For any $\epsilon \ge 0$, there exists an information hiding scheme satisfying $P_e \le$

$$N_\epsilon(\mathcal{A}) \sup_{A_{Y|X}\in\mathcal{A}} \mathbf{E}_{Y|X\sim A_{Y|X}}\left[1 - \mathbf{1}\{d_1(S,X) \le \mathsf{D}_1\}\left(1 + \mathsf{L}2^{-\hat{\imath}(U;Y|K)+\imath(U;S|K)}\right)^{-1}\right] + \epsilon,$$

where $(S,K,U,X,Y) \sim P_{S,K}P_{U,X|S,K}A_{Y|X}$ in the expectation, and $\hat{\imath}(U;Y|K)$ is the information density computed by $P_{S,K}P_{U,X|S,K}\hat{A}_{Y|X}$ (instead of $A_{Y|X}$), assuming that $\imath(U;S|K), \hat{\imath}(U;Y|K)$ are almost surely finite for every $A_{Y|X} \in \mathcal{A}$.

When $K = \emptyset$, $d_1(s,x) = 0$, and $\mathcal{A} = \{A_{Y|X}\}$, taking $\hat{A}_{Y|X} = A_{Y|X}$, above reduces to the one-shot Gelfand-Pinsker coding result (Li and Anantharam, 2021).

One-Shot Achievability Results: Proof

- Design the code assuming $\hat{A}_{Y|X}$ is attacker; show it works for all $A_{Y|X} \in \mathcal{A}$.
- Codebook: $\mathbf{U} := \overline{((\bar{U}_i, \bar{M}_i), T_i)_i \text{ s.t. } (T_i)_i \sim \mathrm{PP}(1)}$, $\bar{U}_i \overset{\text{iid}}{\sim} P_U$, $\bar{M}_i \overset{\text{iid}}{\sim} \mathrm{Unif}[\mathsf{L}]$.
- Encoder calculates $\mathbf{U}_{P_{U|S,K}(\cdot|S,K)\times\delta_M}$ and sends $X|(S,K,U) \sim P_{X|S,K,U}$.
- Decoder calculates $\hat{M}$ by $\mathbf{U}_{\hat{P}_{U|Y,K}(\cdot|Y,K)\times P_M}$ where $\hat{P}_{U|Y,K}$ is by using $\hat{A}_{Y|X}$.

$$P_e(A_{Y|X}) := 1 - \mathbf{P}_{Y|X\sim A_{Y|X}}\Big(d_1(S,X) \le \mathsf{D}_1 \ \text{ AND } \ M = \hat{M}\Big)$$

$$\le \mathbf{E}\Big[1 - \mathbf{1}\{d_1(S,X) \le \mathsf{D}_1\} \cdot \mathbf{P}\Big((U,M) = \mathbf{U}_{\hat{P}_{U|Y,K}(\cdot|Y,K)\times P_M}|M,S,U,Y,K\Big)\Big]$$

$$\overset{\text{PML}}{\le} \mathbf{E}\Big[1 - \mathbf{1}\{d_1(S,X) \le \mathsf{D}_1\} \cdot \Big(1 + \frac{\mathrm{d}P_{U|S,K}(\cdot|S,K)\times\delta_M}{\mathrm{d}\hat{P}_{U|Y,K}(\cdot|Y,K)\times P_M}(U,M)\Big)^{-1}\Big]$$

$$\le \sup_{A_{Y|X}\in\mathcal{A}} \mathbf{E}_{Y|X\sim A_{Y|X}}\Big[1 - \mathbf{1}\{d_1(S,X) \le \mathsf{D}_1\}\Big(1 + \mathsf{L}2^{-\hat{\imath}(U;Y|K)+\imath(U;S|K)}\Big)^{-1}\Big] =: \overline{P_e}.$$

- We're left to fix the codebook. Let $\tilde{\mathcal{A}} \subseteq \mathcal{A}$ attain the minimum in $N_\epsilon(\mathcal{A})$:
  1. Let $P_e(A) := \mathbf{E}_{\mathcal{C}}[P_e(A,\mathcal{C})]$; $\forall A \in \mathcal{A}$, let $\tilde{A} \in \tilde{\mathcal{A}}$ has $\sup_{x\in\mathcal{X}} \|A(\cdot|x) - \tilde{A}(\cdot|x)\|_{\mathrm{TV}} \le \epsilon$;
  2. Hence $|P_e(A,c) - P_e(\tilde{A},c)| \le \epsilon$ and $P_e(A,c) \le \sum_{\tilde{A}\in\tilde{\mathcal{A}}} P_e(\tilde{A},c) + \epsilon$, therefore
  $$\mathbf{E}_{\mathcal{C}}\Big[\sup_{A\in\mathcal{A}} P_e(A,\mathcal{C})\Big] \le \mathbf{E}_{\mathcal{C}}\Big[\sup_{\tilde{A}\in\tilde{\mathcal{A}}} P_e(A,\mathcal{C}) + \epsilon\Big] \le \sum_{\tilde{A}\in\tilde{\mathcal{A}}} P_e(\tilde{A}) + \epsilon \le |\tilde{\mathcal{A}}| \cdot \overline{P_e} + \epsilon$$
  3. The proof is completed by existing a codebook c s.t. $\sup_{A\in\mathcal{A}} P_e(A,c) \le |\tilde{\mathcal{A}}|\overline{P_e} + \epsilon$

Recovering the Asymptotic Result

Proposition: simple bound on the ϵ-covering number

If $\mathcal{X}$ and $\mathcal{Y}$ are discrete and finite, then

$$N_\epsilon(\mathcal{A}) \le \left(\frac{1}{2\epsilon} + \frac{|\mathcal{Y}|+1}{2}\right)^{|\mathcal{X}|\cdot|\mathcal{Y}|}. \tag{5}$$

Proof:

- Write $d(A, \tilde{A}) := \sup_{x\in\mathcal{X}} \|A_{Y|X}(\cdot|x) - \tilde{A}_{Y|X}(\cdot|x)\|_{\mathrm{TV}}$.
- Start with $\tilde{\mathcal{A}} = \emptyset$, add $A \in \mathcal{A}$ not currently covered by $\tilde{\mathcal{A}}$ to $\tilde{\mathcal{A}}$ one by one. The $(\epsilon/2)$-balls $\{A : d(A, \tilde{A}) \le \epsilon/2\}$ must be disjoint for $\tilde{A} \in \tilde{\mathcal{A}}$.
- Treat $A_{Y|X}$ as a transition probability matrix $A \in \mathbb{R}^{|\mathcal{Y}|\times|\mathcal{X}|}$.
- $d(A, \tilde{A}) = \frac{1}{2}\|A - \tilde{A}\|_1 = \frac{1}{2}\max_x \sum_y |A_{y,x} - \tilde{A}_{y,x}|$.
- Ball $\{A \in \mathbb{R}^{|\mathcal{Y}|\times|\mathcal{X}|} : d(A, \tilde{A}) \le \epsilon/2\}$ has volume $V_1 := ((2\epsilon)^{|\mathcal{Y}|}/(|\mathcal{Y}|!))^{|\mathcal{X}|}$; they are subsets of $\{A \in \mathbb{R}^{|\mathcal{Y}|\times|\mathcal{X}|} : \min_{x,y} A_{y,x} \ge -\epsilon, \max_x \sum_y A_{y,x} \le 1+\epsilon\}$, which has volume $V_2 := ((1 + (|\mathcal{Y}|+1)\epsilon)^{|\mathcal{Y}|}/(|\mathcal{Y}|!))^{|\mathcal{X}|}$.
- Hence $|\tilde{\mathcal{A}}|$ is upper bounded by V_2/V_1, giving (5).

Recovering the Asymptotic Result by Moulin and O'sullivan (2003)

- Consider sequences S^n, K^n, X^n, Y^n from discrete spaces, and $A_{Y|X}$ is memoryless and subject to a distortion constraint, the hiding capacity is:

$$C = \max_{P_{U,X|S,K}\ A_{Y|X}:\, \mathbf{E}[d_2(X,Y)]\le \mathsf{D}_2} \min \Big(I(U;Y|K) - I(U;S|K) \Big), \qquad (6)$$

  where the maximum is over $P_{U,X|S,K}$ with $\mathbf{E}[d_1(S,X)] \le \mathsf{D}_1$.
- For a input distribution P_X, the class of memoryless attackers is

$$\mathcal{A}_n(P_X) := \Big\{ A_{Y|X}^n :\ A_{Y|X} \text{ is subject to } \mathbf{E}_{(X,Y)\sim P_X A_{Y|X}}[d_2(X,Y)] \le \mathsf{D}_2 \Big\}$$

- Let $P_{U,X|S,K}$ achieve the max of (6) subject to $\mathbf{E}[d_1(S,X)] \le \mathsf{D}_1'$, $\mathsf{D}_1' < \mathsf{D}_1$.
- Let $\hat{A}_{Y|X}$ be the minimizer of $\min_{A_{Y|X}:\, \mathbf{E}[d_2(X,Y)]\le \mathsf{D}_2} I(U;Y|K)$.
- Fix $R < \hat{I}(U;Y|K) - I(U;S|K)$ assuming $\hat{I}(U;Y|K), \hat{\imath}(U;Y|K)$ are calculated from $P_{U,X,S,K}\hat{A}_{Y|X}$. Let $\mathsf{L} = 2^{nR}$.
  1. $\mathsf{L} 2^{-\hat{\imath}(U^n;Y^n|K^n)+\imath(U^n;S^n|K^n)} \le 2^{nR-\sum_{i=1}^n (\hat{\imath}(U_i;Y_i|K_i)-\imath(U_i;S_i|K_i))} \overset{\mathrm{n}\to\infty}{\to} 0$
  2. Construct an ϵ-cover of $\mathcal{A}_n(P_X)$ using an (ϵ/n)-cover of $\mathcal{A}(P_X)$, we find $N_\epsilon(\mathcal{A}_n(P_X)) \le N_{\frac{\epsilon}{n}}(\mathcal{A}(P_X)) = O((\frac{\epsilon}{n})^{|\mathcal{X}|\cdot|\mathcal{Y}|})$, growing slower than above.
  3. Take $\epsilon = 1/n$, $P_e \to 0$ as $n \to \infty$. Taking $\mathsf{D}_1' \to \mathsf{D}_1$ completes the proof.

Summary

Summary

- We provide a one-shot analysis of the information hiding problem.
- Our results apply to arbitrary channels (not necessarily memoryless or subject to distortion constraints), any host distribution and any class of attackers (not necessarily finite).
- We assume the decoder is uninformed of the attack channel, which is more general and suitable to the one-shot scenario.
- We provide an alternative proof of the asymptotic hiding capacity which is probably simpler.

Outline

Background

- So far, based on the PFR and PML, we have shown:
  1. a multi-hop generalization of the PML on noisy networks;
  2. a covering argument together with the PML for channels with uncertainties.
- What if the channel is special and has some privacy properties?

Background

- In modern data science, large amounts of high-quality data generated with personal information (by edge devices) are susceptible to privacy breaches.
- Differential privacy (Warner, 1965; Dwork et al., 2006) is a powerful tool for safeguarding users' privacy by properly randomizing the local data.
- Communicating (high-dimensional) local data to the central server is often a bottleneck in the system pipeline, thus needs compression.

Research Question

How can we efficiently communicate private data?

Differential Privacy (DP)

Definition: Differentially Private Mechanisms

Given a mechanism $\mathcal{A}$ which induces distribution $P_{Z|X}$ of $Z = \mathcal{A}(X)$, we say that it satisfies (ϵ, δ)-DP if for any neighboring[a] $(x, x') \in \mathcal{N}$ and $\mathcal{S} \subseteq \mathcal{Z}$, it holds that

$$\mathbf{P}(Z \in \mathcal{S} \,|\, X = x) \leq e^{\epsilon} \mathbf{P}(Z \in \mathcal{S} \,|\, X = x') + \delta.$$

[a] We consider a symmetric neighbor relation $\mathcal{N} \subseteq \mathcal{X}^2$, and say x, x' are neighbors if $(x, x') \in \mathcal{N}$.
If a mechanism satisfies $(\epsilon, 0)$-DP, we write it as ϵ-DP.
If $\mathcal{N} = \mathcal{X}^2$, we say the mechanism satisfies (ϵ, δ)-local DP.

- For $\epsilon \leq 1$, Bassily and Smith (2015) showed that a single bit can simulate any local DP randomizer with a small degradation of utility.
- Bun et al. (2019) proposed a rejection-sampling-based compression technique, which compresses an ϵ-DP mechanism into a 10ϵ-DP mechanism.
- Feldman and Talwar (2021) also considered a rejection sampling scheme.
- In (Triastcyn et al., 2021; Shah et al., 2022), importance sampling (or more specifically, minimum random coding (Havasi et al., 2018)) was utilized.

These works are either approximate (the output distribution is distorted), or do not guarantee compression sizes close to $I(X; Z)$.

Channel Simulation

One-shot channel simulation (a compression task) aims to find the minimal needed communication over a noiseless channel to "simulate" another channel $P_{Z|X}$.

One-Shot Channel Simulation

- Alice sees $X \sim P_X$ and sends description M to Bob noiselessly, so that Bob generates $Z \sim P_{Z|X}$. They share unlimited common randomness W.
- The goal is to find $\mathbf{E}[\mathrm{Len}(M)]$, minimum expected description length of M.
- **Converse**: by $X \leftrightarrow M \leftrightarrow Z$, we know

$$\mathbf{E}[\mathrm{Len}(M)] \geq H(M|W) \geq I(X;Z|W) = I(X;Z,W) - I(X;W) \geq I(X;Z)$$

- **Achievability**: The PFR (Li and El Gamal, 2018) promises:
  1. $\mathbf{E}[\log K] \leq I(X;Z) + e^{-1}\log e + 1$
  2. $H(K) \leq \mathbf{E}[\log K] + \log\Big(\mathbf{E}[\log K] + 1\Big) + 1.$
  3. $\mathbf{E}[\mathrm{Len}(M)] \leq I(X;Z) + \log\Big(I(X;Z) + 1\Big) + 5.$
- Applications of channel simulation:
  1. Neural network compression by (Havasi et al., 2018);
  2. Image compression via variational autoencoders by (Flamich et al., 2020);
  3. Differentially private federated learning by (Shah et al., 2022).

Poisson Private Representation

Our Objective

- We treat differential privacy mechanism $P_{Z|X}$ as a channel, and simulate it.
- Can we just use the PFR?
  - No! PFR selects K as a deterministic function of X and W.
  - Changes on input affect K deterministically, but privacy requires randomness.
- We look for a scheme that preserves local differential privacy, while maintaining the advantages of PFR.

Poisson Private Representation (PPR)

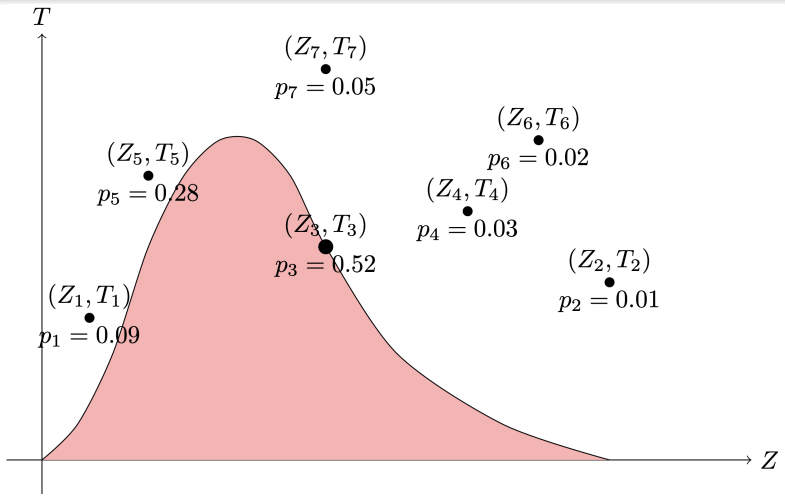
- We design an algorithm that compresses DP mechanism while ensuring:
  1. **Universality**: we simulate arbitrary DP mechanism with discrete or continuous input.
  2. **Exactness**: we ensure exact simulation where the reproduced distribution perfectly matches the original one.
  3. **Communication efficiency**: we compress the output to a size close to the theoretical lower bound $I(X;Z)$.
- Our algorithm is the first method that can achieve all three targets.

Poisson Private Representation

Poisson Private Representation: Construction

Input: x, (ϵ, δ)-DP mechanism $P_{Z|X}$, reference distribution Q, parameter $\alpha > 1$.

1. Generate shared randomness between user and server $(Z_i)_{i=1,2,\ldots} \overset{\text{iid}}{\sim} Q$.
2. The user knows $(Z_i)_i$, input x, $P_{Z|X}$ and performs:
   1. Generate the Poisson process $(T_i)_i$ with rate 1.
   2. Compute $\tilde{T}_i := T_i \cdot \left(\frac{dP}{dQ}(Z_i)\right)^{-1}$ where $P := P_{Z|X}(\cdot|x)$.
   3. Generate $K \in \mathbb{Z}_+$ with $\mathbf{P}(K = k) = \frac{\tilde{T}_k^{-\alpha}}{\sum_{i=1}^{\infty} \tilde{T}_i^{-\alpha}}$.
   4. Compress and send K.
3. The server, which observes $(Z_i)_i$ and K, outputs $Z = Z_K$.



Poisson Private Representation: Theory

Theorem: Communication efficiency

For PPR with parameter $\alpha > 1$, message K satisfies

$$\mathbf{E}[\log K] \leq D_{\mathsf{KL}}(P\|Q) + \frac{\log(3.56)}{\min\{\frac{\alpha-1}{2}, 1\}}.$$

As a result, when the input $X \sim P_X$ is random, taking $Q = P_Z$, we have

$$\mathbf{E}[\log K] \leq I(X;Z) + \frac{\log(3.56)}{\min\{\frac{\alpha-1}{2}, 1\}}.$$

Hence, K can be encoded into $I(X;Z) + \log(I(X;Z)+1) + O(1)$ bits, close to the theoretical lower bound $I(X;Z)$.

Poisson Private Representation: Theoretic Guarantees

Theorem: Exactness

The output Z of PPR follows $P_{Z|X}$ exactly.

Remarks

- Due to the exactness of PPR, it preserves all desirable statistical properties (e.g., unbiasedness and Gaussianity).
  1. If we only want a stand-alone privacy mechanism, we can just focus on the privacy and utility.
  2. However, if the output is used for downstream tasks (e.g., the server sends aggregated mean from clients to data analysts), exactness ensures more precise (central) privacy and utility guarantees.
- Dithered-quantization schemes only work for additive-noise mechanisms.

Poisson Private Representation: Theoretic Guarantees

Theorem: Privacy Guarantee on ϵ-DP

If the mechanism $P_{Z|X}$ is ϵ-DP, then PPR $P_{(Z_i)_i,K|X}$ with $\alpha > 1$ is $2\alpha\epsilon$-DP.

1. Consider neighbors x_1, x_2, let $P_j := P_{Z|X}(\cdot|x_j)$ and $\tilde{T}_{j,i} := T_i \big/ (\frac{dP_j}{dQ}(Z_i))$. Since $P_{Z|X}$ is ϵ-DP, we know $e^{-\epsilon}\frac{dP_2}{dQ}(z) \le \frac{dP_1}{dQ}(z) \le e^{\epsilon}\frac{dP_2}{dQ}(z)$, and hence
$$e^{-\epsilon}\tilde{T}_{2,i} \le \tilde{T}_{1,i} \le e^{\epsilon}\tilde{T}_{2,i}.$$
2. Let K_j be PPR's output applied on P_j and $A := (Z_i, T_i)_i$. Almost surely,
$$\mathbf{P}(K_1 = k) = \frac{\tilde{T}_{1,k}^{-\alpha}}{\sum_{i=1}^{\infty}\tilde{T}_{1,i}^{-\alpha}} \le \frac{e^{\alpha\epsilon}\tilde{T}_{2,k}^{-\alpha}}{\sum_{i=1}^{\infty}e^{-\alpha\epsilon}\tilde{T}_{1,i}^{-\alpha}} = e^{2\alpha\epsilon}\mathbf{P}(K_2 = k|(Z_i, T_i)_i).$$
3. For any measurable $\mathcal{S} \subseteq \mathcal{Z}^{\infty} \times \mathbb{Z}_{\ge 0}$,
$$\begin{aligned}\mathbf{P}(((Z_i)_i, K_1) \in \mathcal{S}) &= \mathbf{E}[\mathbf{P}(((Z_i)_i, K_1) \in \mathcal{S}|(Z_i, T_i)_i))] \\ &= \mathbf{E}\Big[\sum_{k:((Z_i)_i,k)\in\mathcal{S}} \mathbf{P}(K_1 = k|(Z_i, T_i)_i))\Big] \\ &\le e^{2\alpha\epsilon}\mathbf{E}\Big[\sum_{k:((Z_i)_i,k)\in\mathcal{S}} \mathbf{P}(K_2 = k|(Z_i, T_i)_i))\Big] \\ &= e^{2\alpha\epsilon}\mathbf{P}(((Z_i)_i, K_2) \in \mathcal{S})\end{aligned}$$

Poisson Private Representation: Theoretic Guarantees

Theorem: Privacy Guarantee on (ϵ, δ)-DP

If $P_{Z|X}$ is (ϵ, δ)-DP, then PPR $P_{(Z_i)_i, K|X}$ with parameter $\alpha > 1$ is $(2\alpha\epsilon, 2\delta)$-DP.

Theorem: Tighter Privacy Guarantee on (ϵ, δ)-DP

If $P_{Z|X}$ is (ϵ, δ)-DP, then PPR $P_{(Z_i)_i, K|X}$ with parameter $\alpha > 1$ is $(\alpha\epsilon + \tilde{\epsilon}, 2(\delta + \tilde{\delta}))$-DP, for every $\tilde{\epsilon} \in (0, 1]$ and $\tilde{\delta} \in (0, 1/3]$ that satisfy $\alpha \leq e^{-4.2}\tilde{\delta}\tilde{\epsilon}^2/(-\ln \tilde{\delta}) + 1$.

Running Time

Running Time

- Note since $\mathbf{E}[\log K] \approx I(X;Z)$, K (and hence the running time) is at least exponential in $I(X;Z)$.
- However, an exponential complexity is also needed in sampling methods without privacy guarantee, e.g., (Havasi et al., 2018) and (Maddison, 2016).
- By Agustsson and Theis (2020), no polynomial time general sampling-based method exists (even without privacy constraint), if $RP \neq NP$.
- Nevertheless, this is not an obstacle: $I(X;Z)$ for a good local DP mechanism must be small, or the leakage of X in Z will be too large.
  - By Cuff and Yu (2016), for an ϵ-local DP mechanism, $I(X;Z) \leq \min\{\epsilon, \epsilon^2\}$
- Another way to reduce the running time is to divide the data into small chunks and apply the mechanism to each chunk separately.

Application: Distributed Mean Estimation

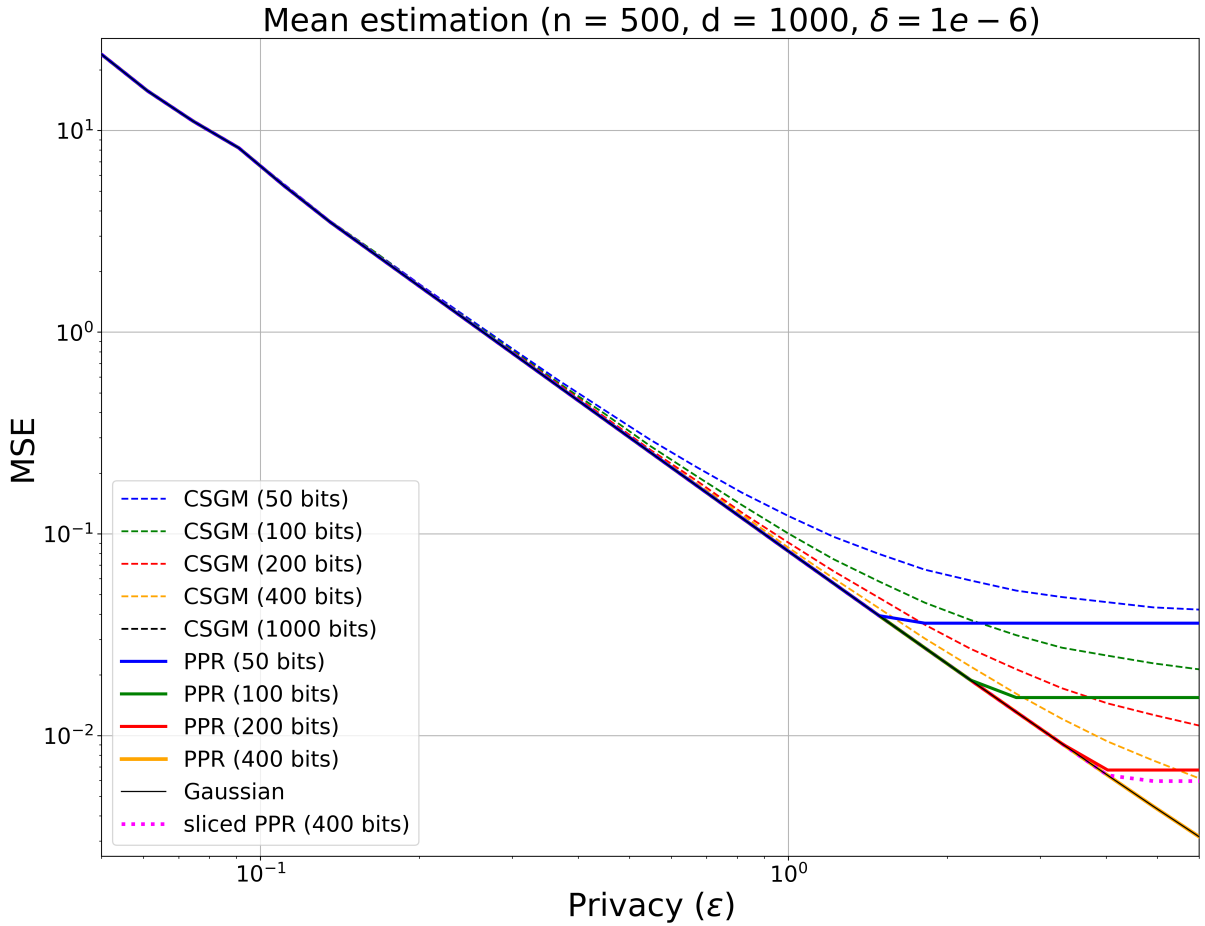
Distributed Mean Estimation

- We compare PPR with (Chen et al., 2024) on distributed mean estimation, which is the core sub-routine in federated optimization (Abadi et al., 2016).
- Each of n clients has $X_i \in \mathbb{R}^d$ and sends Z_i; server estimates $\mu = \frac{1}{n}\sum_{i=1}^n X_i$.
- For Gaussian mechanism $P_{Z|X}(\cdot|x) = \mathcal{N}(x, \frac{\sigma^2}{n}\mathbb{I}_d)$ and proposal distribution $Q = \mathcal{N}(0, (\frac{C^2}{d} + \frac{\sigma^2}{n})\mathbb{I}_d)$, for each client i, the output of PPR is Z_i, and:
  - $\hat{\mu} = \frac{1}{n}\sum_i Z_i$ is an **unbiased** estimator of μ, satisfying (ϵ, δ)-central DP and has mean squared error (MSE) $\mathbf{E}[\|\mu - \hat{\mu}\|_2^2] = \sigma^2 d/n^2$.
  - For $\epsilon < 1/\sqrt{n}$, PPR satisfies $(2\alpha\sqrt{n}\epsilon, 2\delta)$-local DP.
  - The average per-client communication cost $\leq \ell + \log(\ell+1) + 2$ bits, where

$$\ell \leq \frac{d}{2}\log\left(\frac{n\varepsilon^2}{2d\ln(1.25/\delta)} + 1\right) + \frac{\log(3.56)}{\min\{(\alpha-1)/2,\, 1\}}.$$

- For a fixed α, the communication cost is as good as (Suresh et al., 2017; Chen et al., 2024), and is better when $n \gg d$.

Application: Distributed Mean Estimation



Compare to the Coordinate Subsampled Gaussian Mechanism (CSGM) by (Chen et al., 2024).

- Compared to CSGM, PPR consistently achieves smaller MSE:
  - For $\epsilon = 1$ and compressing to 50 bits, we give a 33.61% MSE reduction;
  - For $\epsilon = 0.5$ and compressing to 25 bits, we give a 22.33% MSE reduction;
  - Both schemes are asymptotically optimal, hence the reduction is significant.
- Running time can be greatly reduced by breaking the vector into chunks.

Application: Metric Privacy and Laplace Mechanism

Metric Privacy of PPR

- ϵ-DP can be extended to *metric privacy* by using a metric $d_{\mathcal{X}}(x, x')$ over $\mathcal{X}$.
- For a mechanism $\mathcal{A} := P_{Z|X}$ and a metric $d_{\mathcal{X}}$, it has $\epsilon \cdot d_{\mathcal{X}}$-privacy (Andrés et al., 2013; Chatzikokolakis et al., 2013) if for any $x, x' \in \mathcal{X}$, $\mathcal{S} \subseteq \mathcal{Z}$,

$$\mathbf{P}(Z \in \mathcal{S} \,|\, X = x) \leq e^{\epsilon \cdot d_{\mathcal{X}}(x,x')} \mathbf{P}(Z \in \mathcal{S} \,|\, X = x').$$

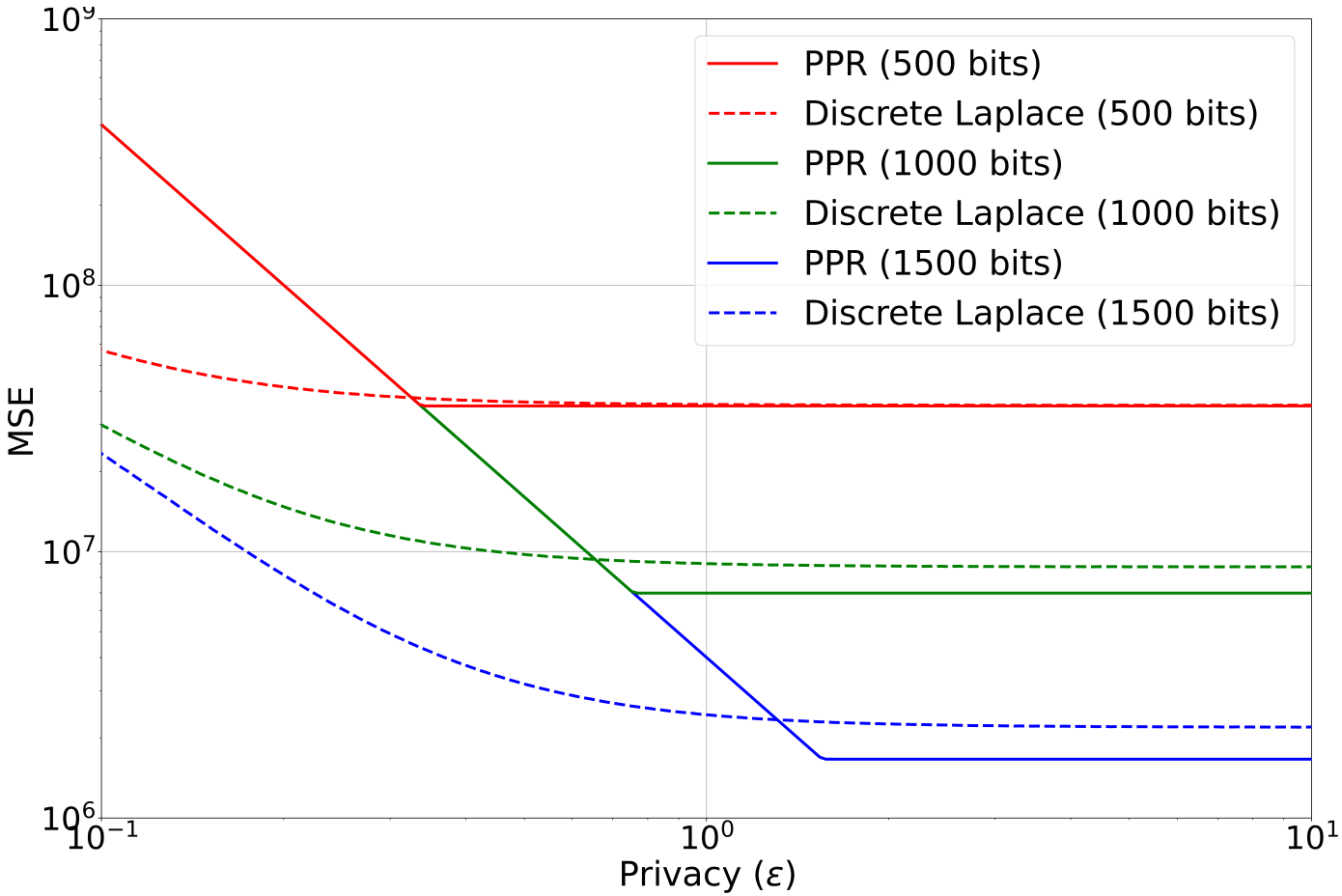
- It recovers ϵ-central DP by letting $d_{\mathcal{X}}$ be the Hamming distance, and ϵ-local DP by letting $d_{\mathcal{X}}$ be the discrete metric (Chatzikokolakis et al., 2013).
- **Metric privacy of PPR**: If the mechanism $P_{Z|X}$ satisfies $\epsilon \cdot d_{\mathcal{X}}$-privacy, then PPR $P_{(Z_i)_i, K|X}$ with $\alpha > 1$ satisfies $2\alpha\epsilon \cdot d_{\mathcal{X}}$-privacy.

Laplace Mechanism

- For Laplace mechanism $f_{Z|X} \propto e^{-\epsilon \cdot d_{\mathcal{X}}(x,z)}$, $d_{\mathcal{X}}(x, z) = \|x - z\|_2$, with $X \in \{x \in \mathbb{R}^d | \|x\|_2 \leq C\}$ and $Q = \mathcal{N}(0, (\frac{C^2}{d^2} + \frac{d+1}{\epsilon^2}))$, PPR's output has MSE $\frac{d(d+1)}{\epsilon^2}$, $2\alpha\epsilon \cdot d_{\mathcal{X}}$-privacy, and compression size $\leq \ell + \log(\ell + 1) + 2$ bits,

$$\ell := \frac{d}{2} \log \left(\frac{2}{e} \left(\frac{C^2 \epsilon^2}{d} + d + 1 \right) \right) - \log \left(\frac{\Gamma(d+1)}{\Gamma(\frac{d}{2} + 1)} \right) + \frac{\log(3.56)}{\min\{(\alpha - 1)/2\}}.$$

Application: Metric Privacy and Laplace Mechanism



Compare to the discrete Laplace mechanism by quantization (Andrés et al., 2013). $C = 10000$, $d = 500$ and $\alpha = 2$.

- Application: users send privatized Z to an untrusted server for some services.
- PPR performs better for large ϵ or small MSE, and preserve $f_{Z|X}$ exactly.

Summary

Summary

- We propose a novel scheme for compressing differential privacy mechanisms, called the **Poisson private representation**.
- Unlike previous schemes which are either constrained on special classes of DP mechanisms or introducing additional distortions on the output, our scheme can compress and **exactly** simulate **arbitrary** mechanisms while providing privacy guarantees.
- PPR provides a compression size that is close to the theoretic lower bound.
- PPR is the first scheme that achieves universality, exactness and near-optimal compression at the same time.

References

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., and Zhang, L. (2016). Deep learning with differential privacy. In *CCS'16*, pages 308–318.

Agustsson, E. and Theis, L. (2020). Universally quantized neural compression. *Advances in neural information processing systems*, 33:12367–12376.

Andrés, M. E., Bordenabe, N. E., Chatzikokolakis, K., and Palamidessi, C. (2013). Geo-indistinguishability: Differential privacy for location-based systems. In *CCC'13*, pages 901–914.

Bassily, R. and Smith, A. (2015). Local, private, efficient protocols for succinct histograms. In *STOC*, pages 127–135.

Blackwell, D., Breiman, L., and Thomasian, A. (1959). The capacity of a class of channels. *The Annals of Mathematical Statistics*, pages 1229–1241.

Bun, M., Nelson, J., and Stemmer, U. (2019). Heavy hitters and the structure of local privacy. *ACM Transactions on Algorithms (TALG)*, 15(4):1–40.

Chatzikokolakis, K., Andrés, M. E., Bordenabe, N. E., and Palamidessi, C. (2013). Broadening the scope of differential privacy using metrics. In *PETS'13*, pages 82–102. Springer.

Chen, W.-N., Song, D., Ozgur, A., and Kairouz, P. (2024). Privacy amplification via compression: Achieving the optimal privacy-accuracy-communication trade-off in distributed mean estimation. *Advances in Neural Information Processing Systems*, 36.

Cover, T. and Gamal, A. E. (1979). Capacity theorems for the relay channel. *IEEE Transactions on information theory*, 25(5):572–584.

Cuff, P. (2013). Distributed channel synthesis. *IEEE Transactions on Information Theory*, 59(11):7071–7096.

Cuff, P., Permuter, H., and Cover, T. M. (2010). Coordination capacity. *IEEE Trans. Inf. Theory*, 56(9):4181–4206.

Cuff, P. and Yu, L. (2016). Differential privacy as a mutual information constraint. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *TCC 2006*, pages 265–284. Springer.

El Gamal, A., Hassanpour, N., and Mammen, J. (2007). Relay networks with delays. *IEEE Transactions on Information Theory*, 53(10).

Feinstein, A. (1954). A new basic theorem of information theory. *IRE Trans. Inf. Theory*, (4):2–22.

Feldman, V. and Talwar, K. (2021). Lossless compression of efficient private local randomizers. In *International Conference on Machine Learning*, pages 3208–3219. PMLR.

Flamich, G., Havasi, M., and Hernández-Lobato, J. M. (2020). Compressing images by encoding their latent representations with relative entropy coding. *NeurIPS*, 33:16131–16141.

Guo, Y., Salehkalaibar, S., Draper, S. C., and Yu, W. (2024). One-shot achievability region for hypothesis testing with communication constraint. In *ITW 2024*, pages 55–60. IEEE.

Havasi, M., Peharz, R., and Hernández-Lobato, J. M. (2018). Minimal random code learning: Getting bits back from compressed model parameters. *arXiv preprint arXiv:1810.00440*.

Hayashi, M. (2009). Information spectrum approach to second-order coding rate in channel coding. *IEEE Transactions on Information Theory*, 55(11):4947–4966.

Hentilä, H., Shkel, Y. Y., and Koivunen, V. (2024). Communication-constrained secret key generation: Second-order bounds. *IEEE Transactions on Information Theory*.

Khisti, A., Behboodi, A., Cesa, G., and Kumar, P. (2024). Unequal message protection: One-shot analysis via poisson matching lemma. In *ISIT 2024*, pages 629–634. IEEE.

Kim, Y.-H. (2007). Coding techniques for primitive relay channels. In *Proc. 45th Annual Allerton Conf. Commun., Contr. Comput*, page 2007.

Kobus, S. and Gündüz, D. (2025). Remote reinforcement learning with communication constraints.

Lee, S.-H. and Chung, S.-Y. (2018). A unified random coding bound. *IEEE Transactions on Information Theory*, 64(10):6779–6802.

Lei, E., Hassani, H., and Bidokhti, S. S. (2023). Neural estimation of the rate-distortion function with applications to operational source coding. *IEEE JSAIT*, 3(4):674–686.

Li, C. T. and Anantharam, V. (2021). A unified framework for one-shot achievability via the poisson matching lemma. *IEEE Transactions on Information Theory*, 67(5):2624–2651.

Li, C. T. and El Gamal, A. (2018). Strong functional representation lemma and applications to coding theorems. *IEEE Transactions on Information Theory*, 64(11):6967–6978.

Li, C. T., Wu, X., Özgür, A., and El Gamal, A. (2020). Minimax learning for distributed inference. *IEEE Transactions on Information Theory*, 66(12):7929–7938.

Lim, S. H., Kim, Y.-H., El Gamal, A., and Chung, S.-Y. (2011). Noisy network coding. *IEEE Transactions on Information Theory*, 57(5).

Maddison, C. J. (2016). A poisson process model for monte carlo. *Perturbation, Optimization, and Statistics*, pages 193–232.

Moulin, P. and O'Sullivan, J. A. (2003). Information-theoretic analysis of information hiding. *IEEE Transactions on information theory*.

Polyanskiy, Y., Poor, H. V., and Verdú, S. (2010). Channel coding rate in the finite blocklength regime. *IEEE Transactions on Information Theory*, 56(5):2307–2359.

Scarlett, J. (2015). On the dispersions of the gel'fand–pinsker channel and dirty paper coding. *IEEE Transactions on Information Theory*, 61(9):4569–4586.

Shah, A., Chen, W.-N., Balle, J., Kairouz, P., and Theis, L. (2022). Optimal compression of locally differentially private mechanisms. In *AISTATS*, pages 7680–7723. PMLR.

Shannon, C. E. (1957). Certain results in coding theory for noisy channels. *Information and control*, 1(1):6–25.

Somekh-Baruch, A. and Merhav, N. (2004). On the capacity game of public watermarking systems. *IEEE Transactions on Information Theory*, 50(3):511–524.

Suresh, A. T., Felix, X. Y., Kumar, S., and McMahan, H. B. (2017). Distributed mean estimation with limited communication. In *ICML*, pages 3329–3337. PMLR.

Triastcyn, A., Reisser, M., and Louizos, C. (2021). Dp-rec: Private & communication-efficient federated learning. *arXiv:2111.05454*.

Van Der Meulen, E. C. (1971). Three-terminal communication channels. *Advances in applied Probability*, 3(1):120–154.

Verdú, S. (2012). Non-asymptotic achievability bounds in multiuser information theory. In *2012 50th Annual Allerton Conference on Communication, Control, and Computing*.

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American statistical association*, 60(309):63–69.

Watanabe, S., Kuzuoka, S., and Tan, V. Y. (2015). Nonasymptotic and second-order achievability bounds for coding with side-information. *IEEE Transactions on Information Theory*.

Yamamoto, H. (1982). Wyner-ziv theory for a general function of the correlated sources (corresp.). *IEEE Transactions on Information Theory*, 28(5):803–807.

Yassaee, M. H., Aref, M. R., and Gohari, A. (2013). A technique for deriving one-shot achievability results in network information theory. In *ISIT 2013*, pages 1287–1291. IEEE.