Multi-View Channels

V. Arvind Rameshwar

India Urban Data Exchange

IIT Madras 2024

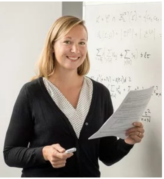
Joint work with

Prof. Nir Weinberger

(Technion, Israel)

Paraskevi (Vivian) Papadopoulou

(TU Münich, Germany)

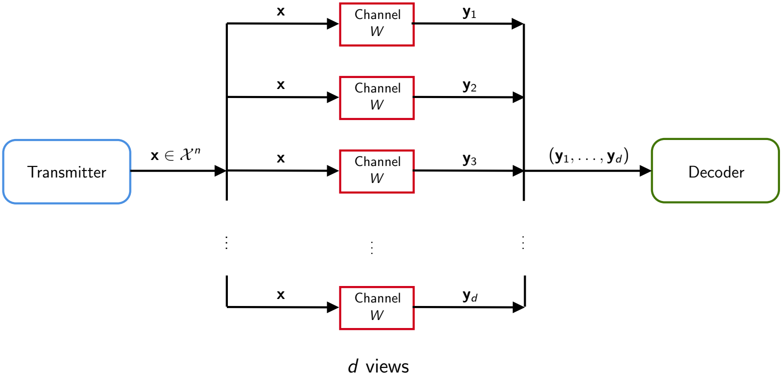


Prof. Dr.-Ing. Antonia Wachter-Zeh

(TU Münich, Germany)

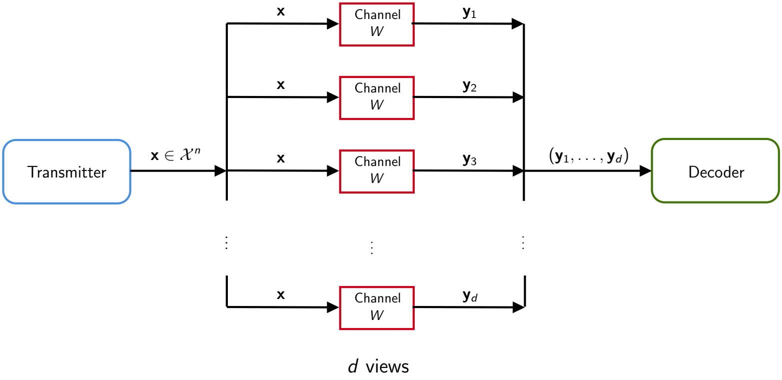
What is the talk about?

The decoder obtains d noisy views of an input sequence.



What is the talk about?

The decoder obtains d noisy views of an input sequence.

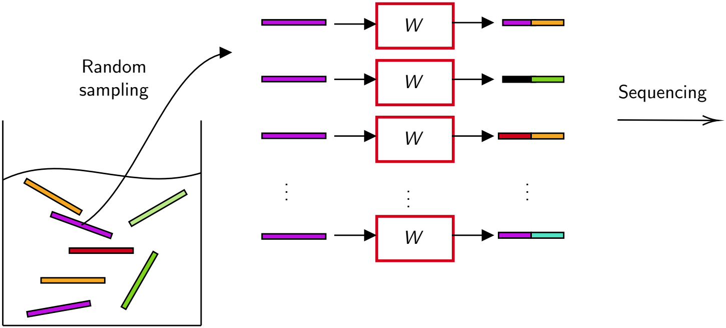


Q1: What rates are achievable, with vanishing reconstruction error probability?

Q2: How many views are required on average to exactly reconstruct a sequence?

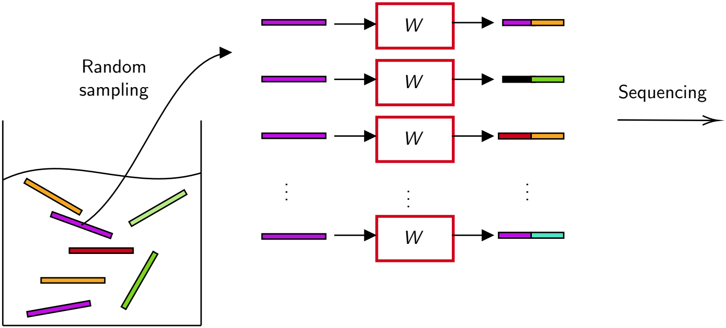
Whence does this problem arise? - I

- Short molecules in a DNA pool are “amplified” by Polymerase Chain Reaction (PCR) and sampled multiple times [Shomorony and Heckel (2022)].



Whence does this problem arise? - I

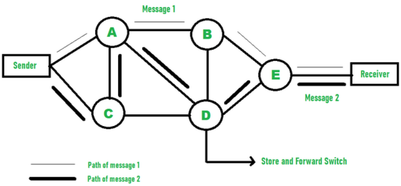
- Short molecules in a DNA pool are “amplified” by Polymerase Chain Reaction (PCR) and sampled multiple times [Shomorony and Heckel (2022)].

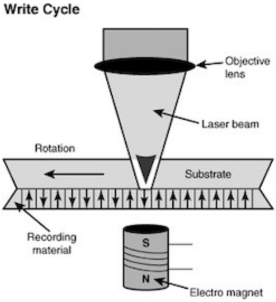


- The capacity is determined by mutual information terms corresponding to different # of views of the input sequence; see [Lenz et al. (2019, 2020), Shomorony and Heckel (2022), Weinberger and Merhav (2022)].

Whence does this problem arise? - II

- Errors due to synchronization in packet-switched communications and in reading magneto-optical media

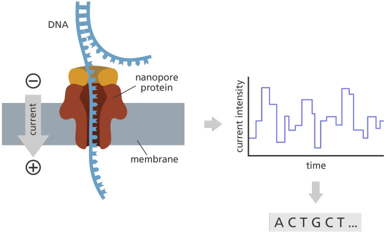


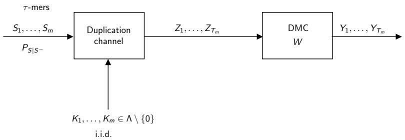


- Each output "run" at the end of (noisy) duplications corresponds to a multi-view channel [Mitzenmacher (2008), Cheraghchi and Ribeiro (2019)]

Whence does this problem arise? - III

- Experimentally accurate channel models for DNA nanopore sequencers [McBain and Viterbo (IEEE-BITS 2023), McBain, Viterbo, Saunderson (2024)]

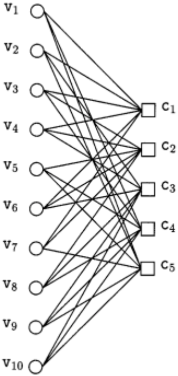


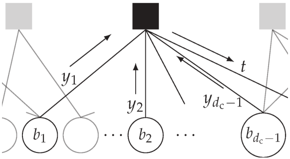


- The channel model is a noisy duplication channel (with a specific input process).

Whence does this problem arise? - IV

- More fundamentally, in the iterative decoding of LDPC codes [Gallager (1962), Richardson and Urbanke (2001)]





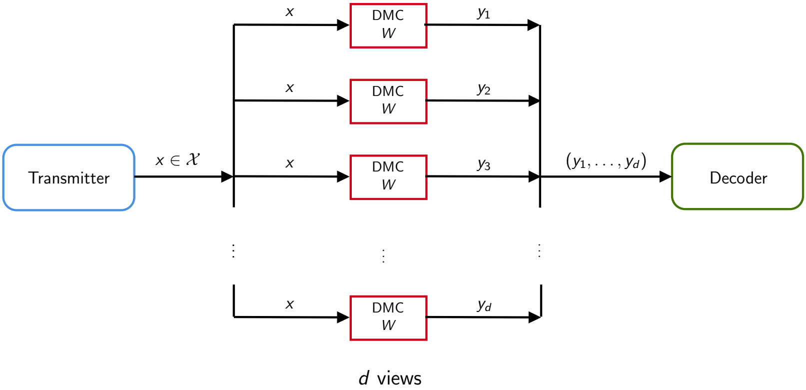
- A single variable node receives multiple "views"/estimates of its value, from different check nodes.

Answering **Q1** in the large view limit

A simpler setting: Multi-view DMCs

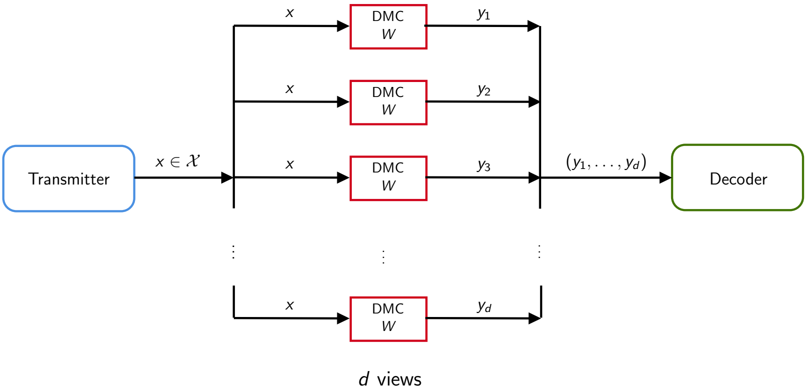
The setting, revisited

The decoder obtains d independent, noisy views of an input symbol.



The setting, revisited

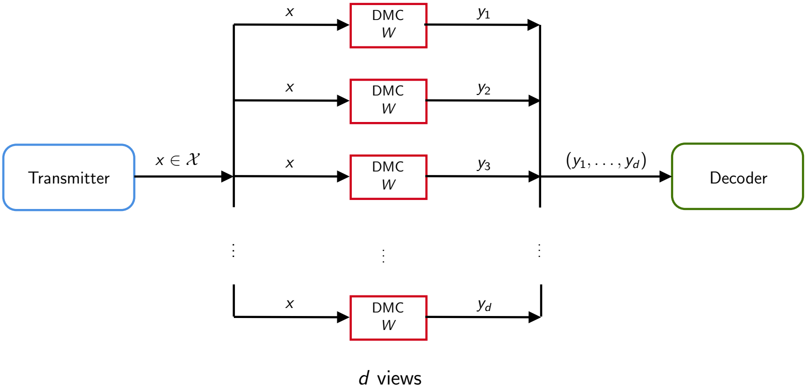
The decoder obtains d independent, noisy views of an input symbol.



Since a multi-view DMC is also a DMC, it suffices for us to focus on the transmission of a single symbol, for rate computations.

The setting, revisited

The decoder obtains d independent, noisy views of an input symbol.



Goal: Exact asymptotics of the mutual info. (+ capacity) and dispersion of such a multi-view channel, for arbitrary P_X.

What is known about this setting?

- Hellman and Raviv (1970), Kanaya and Han (1995): Exact asymptotics of information rates over DMCs with multiple views
- Levenshtein (2001): Characterization of # of views for exact reconstruction over comb. error channels and for reconstruction with decaying error prob. over multi-view DMCs

What is known about this setting?

- Hellman and Raviv (1970), Kanaya and Han (1995): Exact asymptotics of information rates over DMCs with multiple views
- Levenshtein (2001): Characterization of # of views for exact reconstruction over comb. error channels and for reconstruction with decaying error prob. over multi-view DMCs
- Land et al. (2005), Land and Huber (2006): Bounds on the mutual information rates over multi-view DMCs via information combining
- Mitzenmacher (2006): Calculation of the capacity of a multi-view binary symmetric channel (BSC)
- Haeupler and Mitzenmacher (2014): Information rates over multi-view deletion channels when Pr[deletion] $\to 0$.
- Batu et al. (2004), Cheraghchi et al. (2020), Peres and Zhai (2017): Trace reconstruction over the deletion channel

What is known about this setting?

- Hellman and Raviv (1970), Kanaya and Han (1995): Exact asymptotics of information rates over DMCs with multiple views
- Levenshtein (2001): Characterization of # of views for exact reconstruction over comb. error channels and for reconstruction with decaying error prob. over multi-view DMCs
- Land et al. (2005), Land and Huber (2006): Bounds on the mutual information rates over multi-view DMCs via information combining
- Mitzenmacher (2006): Calculation of the capacity of a multi-view binary symmetric channel (BSC)
- Haeupler and Mitzenmacher (2014): Information rates over multi-view deletion channels when $\Pr[\text{deletion}] \to 0$.
- Batu et al. (2004), Cheraghchi et al. (2020), Peres and Zhai (2017): Trace reconstruction over the deletion channel

Our contributions:

1. Unified treatment of info. rate and channel dispersion

$\Downarrow$

Finite-blocklength achievable rates with fixed error probability

What is known about this setting?

- Hellman and Raviv (1970), Kanaya and Han (1995): Exact asymptotics of information rates over DMCs with multiple views
- Levenshtein (2001): Characterization of # of views for exact reconstruction over comb. error channels and for reconstruction with decaying error prob. over multi-view DMCs
- Land et al. (2005), Land and Huber (2006): Bounds on the mutual information rates over multi-view DMCs via information combining
- Mitzenmacher (2006): Calculation of the capacity of a multi-view binary symmetric channel (BSC)
- Haeupler and Mitzenmacher (2014): Information rates over multi-view deletion channels when Pr[deletion] $\to 0$.
- Batu et al. (2004), Cheraghchi et al. (2020), Peres and Zhai (2017): Trace reconstruction over the deletion channel

Our contributions:

2. Directly extensible proofs for multi-letter channels

What is known about this setting?

- Hellman and Raviv (1970), Kanaya and Han (1995): Exact asymptotics of information rates over DMCs with multiple views
- Levenshtein (2001): Characterization of # of views for exact reconstruction over comb. error channels and for reconstruction with decaying error prob. over multi-view DMCs
- Land et al. (2005), Land and Huber (2006): Bounds on the mutual information rates over multi-view DMCs via information combining
- Mitzenmacher (2006): Calculation of the capacity of a multi-view binary symmetric channel (BSC)
- Haeupler and Mitzenmacher (2014): Information rates over multi-view deletion channels when $\Pr[\text{deletion}] \to 0$.
- Batu et al. (2004), Cheraghchi et al. (2020), Peres and Zhai (2017): Trace reconstruction over the deletion channel

Our contributions:

3. Non-asymptotic capacity bounds for multi-view BIMS channels

Background

Some formalism

- Consider a DMC W with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$, both finite. Assume that $|\mathcal{X}|$, $|\mathcal{Y}|$ do not depend on d.
- The d-view DMC $W^{(d)}$ obeys the channel law

$$W^{(d)}((y_1, \ldots, y_d) \mid x) = \prod_{i=1}^{d} W(y_i \mid x),$$

for $(y_1, \ldots, y_d) \in \mathcal{Y}^d$ and $x \in \mathcal{X}$.

Some formalism

- Consider a DMC W with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$, both finite. Assume that $|\mathcal{X}|$, $|\mathcal{Y}|$ do not depend on d.
- The d-view DMC $W^{(d)}$ obeys the channel law

$$W^{(d)}((y_1, \ldots, y_d) \mid x) = \prod_{i=1}^{d} W(y_i \mid x),$$

for $(y_1, \ldots, y_d) \in \mathcal{Y}^d$ and $x \in \mathcal{X}$.

- Fix a d-independent input distribution P_X. We are interested in:

$$I^{(d)} = I(X; Y^d) = \mathrm{H}(X) - \mathrm{H}(X \mid Y^d) \qquad \text{[Mutual info.]}$$

and

$$V^{(d)} = \mathbb{E}\left[(\iota(X; Y^d) - I^{(d)})^2\right], \qquad \text{[Channel dispersion]}$$

where

$$\iota(X; Y^d) = \log \frac{P(Y^d \mid X)}{P(Y^d)}. \qquad \text{[Info. spectrum]}$$

Building intuition: Multi-view BSC$^{(d)}$

Consider the d-view BSC(p), with binary input $X \in \{0, 1\}$:

$$Y = X + Z \pmod 2,$$

where $Z \sim \text{Ber}(p)$.

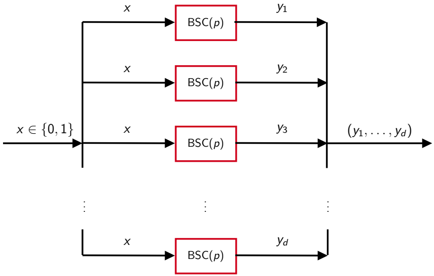
Building intuition: Multi-view BSC$^{(d)}$

Consider the d-view BSC(p), with binary input $X \in \{0, 1\}$:

$$Y = X + Z \ (\text{mod } 2),$$

where $Z \sim \text{Ber}(p)$. Suppose that $X \sim \text{Ber}(1/2)$.

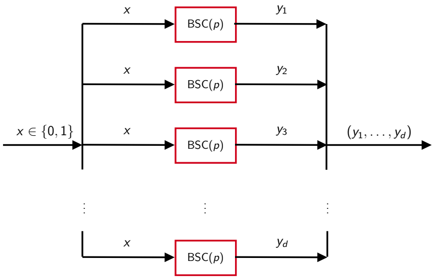
Building intuition: Multi-view BSC$^{(d)}$



- Given outputs $Y^d = (Y_1, \ldots, Y_d)$, compute

$$M = \texttt{Majority}(Y^d).$$

Building intuition: Multi-view BSC$^{(d)}$



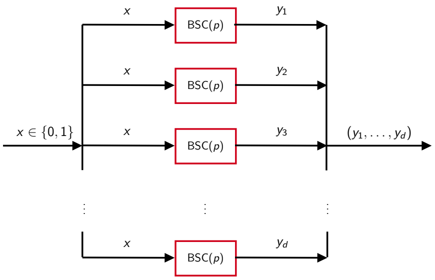
- Given outputs $Y^d = (Y_1, \dots, Y_d)$, compute

$$M = \texttt{Majority}(Y^d).$$

- Then,

$$I(X; Y^d) \geq I(X; M) \\ = \log 2 - h_b(\Pr[X \neq M]).$$

Building intuition: Multi-view $\mathrm{BSC}^{(d)}$



- Given outputs $Y^d = (Y_1, \ldots, Y_d)$, compute

$$M = \texttt{Majority}(Y^d).$$

- Then,

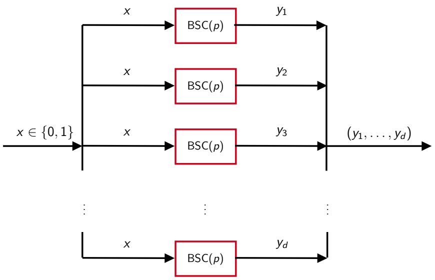
$$I(X; Y^d) \geq I(X; M) = \log 2 - h_b(\Pr[X \neq M]).$$

- By the Chernoff bound,

$$\Pr[X \neq M] \leq \exp(-dZ(p)),$$

where $Z(p) = 2\sqrt{p(1-p)}$.

Building intuition: Multi-view BSC$^{(d)}$



- Given outputs $Y^d = (Y_1, \ldots, Y_d)$, compute

$$M = \texttt{Majority}(Y^d).$$

- Then,

$$I(X; Y^d) \geq I(X; M) \\ = \log 2 - h_b(\Pr[X \neq M]).$$

- By the Chernoff bound,

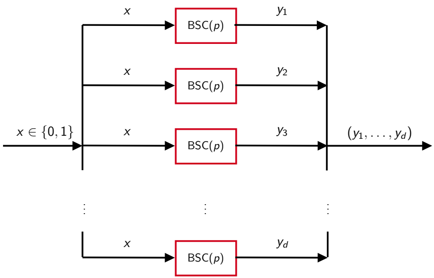
$$\Pr[X \neq M] \leq \exp(-dZ(p)),$$

where $Z(p) = 2\sqrt{p(1-p)}$.

Hence, for large d:

$$h_b(\Pr[X \neq M]) \leq -2\Pr[X \neq M] \cdot \log \Pr[X \neq M] \\ = 2dZ(p) \cdot \exp(-dZ(p)).$$

Building intuition: Multi-view BSC$^{(d)}$



- Given outputs $Y^d = (Y_1, \ldots, Y_d)$, compute
$$M = \texttt{Majority}(Y^d).$$
- Then,
$$I(X; Y^d) \geq I(X; M)$$
$$= \log 2 - h_b(\Pr[X \neq M]).$$
- By the Chernoff bound,
$$\Pr[X \neq M] \leq \exp(-dZ(p)),$$
where $Z(p) = 2\sqrt{p(1-p)}$.

This gives us

$$I(X; Y^d) \geq \log 2 - 2dZ(p) \cdot \exp(-dZ(p)).$$

Exponentially fast convergence to $\mathrm{H}(X) = \log 2$

Our Results

Our main result

- Intuitively, as d becomes large, we expect $I^{(d)} \approx \mathsf{H}(X)$ and

$$V^{(d)} \approx \underbrace{\mathsf{V}(X)}_{\text{i/p varentropy}} = \mathbb{E}\left[(\log P(X) + \mathsf{H}(X))^2\right].$$

Our main result

- Intuitively, as d becomes large, we expect $I^{(d)} \approx \mathsf{H}(X)$ and

$$V^{(d)} \approx \underbrace{\mathsf{V}(X)}_{\text{i/p varentropy}} = \mathbb{E}\left[(\log P(X) + \mathsf{H}(X))^2\right].$$

- For distributions P, Q on $\mathcal{X}$, define the Chernoff information

$$\mathsf{C}(P, Q) = -\min_{\lambda \in [0,1]} \log \left(\sum_{x \in \mathcal{X}} P(x)^{1-\lambda} Q(x)^{\lambda} \right).$$

Our main result

- Intuitively, as d becomes large, we expect $I^{(d)} \approx \mathsf{H}(X)$ and

$$V^{(d)} \approx \underbrace{\mathsf{V}(X)}_{\text{i/p varentropy}} = \mathbb{E}\left[(\log P(X) + \mathsf{H}(X))^2\right].$$

- For distributions P, Q on $\mathcal{X}$, define the Chernoff information

$$\mathsf{C}(P, Q) = -\min_{\lambda \in [0,1]} \log\left(\sum_{x \in \mathcal{X}} P(x)^{1-\lambda} Q(x)^{\lambda}\right).$$

Theorem

When $\mathcal{X}, \mathcal{Y}, P_X$ do not depend on d,

$$I^{(d)} = \mathsf{H}(X) - \exp\left(-d\rho + \Theta(\log d|\mathcal{X}|)\right), \text{ and}$$

$$\left|V^{(d)} - \mathsf{V}(X)\right| = \exp\left(-d\rho + \Theta(\log d|\mathcal{X}|)\right),$$

where

$$\rho = \min_{x, x' :\, x \neq x'} \mathsf{C}(P_{Y|x}, P_{Y|x'}).$$

Interpreting the result

- The rate of convergence ρ of the mutual information and channel dispersion to their limits

$$\rho = \min_{x,x':\ x \neq x'} \mathsf{C}(P_{Y|x}, P_{Y|x'})$$

is independent of the input distribution (except via its support)!

Interpreting the result

- The rate of convergence ρ of the mutual information and channel dispersion to their limits

$$\rho = \min_{x,x' \colon x \neq x'} \mathsf{C}(P_{Y|x}, P_{Y|x'})$$

  is independent of the input distribution (except via its support)!

- For a binary-input memoryless symmetric (BIMS) channel W, the rate

$$\begin{aligned} \rho &= -\log \sum_{y \in \mathcal{Y}} \sqrt{P_{Y|0}(y \mid 0) P_{Y|1}(y \mid 1)} \\ &= -\log Z_b(W), \end{aligned}$$

  where $Z_b(W)$ is the Bhattacharya parameter of the BIMS W.

Interpreting the result

- The rate of convergence ρ of the mutual information and channel dispersion to their limits

$$\rho = \min_{x,x' \colon x \neq x'} \mathsf{C}(P_{Y|x}, P_{Y|x'})$$

is independent of the input distribution (except via its support)!

- For a binary-input memoryless symmetric (BIMS) channel W, the rate

$$\begin{aligned}\rho &= -\log \sum_{y \in \mathcal{Y}} \sqrt{P_{Y|0}(y \mid 0) P_{Y|1}(y \mid 1)} \\ &= -\log Z_b(W),\end{aligned}$$

where $Z_b(W)$ is the Bhattacharya parameter of the BIMS W.

Hence, our earlier speed of convergence for the $\mathsf{BSC}^{(d)}$ is tight!

A finite blocklength corollary

- A characterization of finite-blocklength rates achievable over $W^{(d)}$, thus follows.
- For a fixed $\epsilon \in (0,1)$ and blocklength $n \geq 1$, let

$$M^{\star}(n,\epsilon) \leftarrow \text{largest } M \text{ s.t. } \exists \text{ length-}n \text{ code over } W^{(d)}$$
$$\text{with max. error } \epsilon.$$

A finite blocklength corollary

- A characterization of finite-blocklength rates achievable over $W^{(d)}$, thus follows.
- For a fixed $\epsilon \in (0,1)$ and blocklength $n \geq 1$, let

$$M^\star(n,\epsilon) \leftarrow \text{largest } M \text{ s.t. } \exists \text{ length-}n \text{ code over } W^{(d)} \text{ with max. error } \epsilon.$$

Theorem

If W is "non-singular" (think unlike a BEC), we have

$$\log|\mathcal{X}| - \frac{\log M^\star(n,\epsilon)}{n} \leq e^{-d\rho+\Theta(\log d|\mathcal{X}|)} - \Phi^{-1}(\epsilon)\cdot\frac{e^{-d\rho/2+\Theta(\log d|\mathcal{X}|)}}{\sqrt{n}} + \Theta\left(\frac{\log n}{n}\right).$$

A finite blocklength corollary

- A characterization of finite-blocklength rates achievable over $W^{(d)}$, thus follows.
- For a fixed $\epsilon \in (0,1)$ and blocklength $n \geq 1$, let

$$M^\star(n,\epsilon) \leftarrow \text{largest } M \text{ s.t. } \exists \text{ length-}n \text{ code over } W^{(d)} \text{ with max. error } \epsilon.$$

Theorem

If W is "non-singular" (think unlike a BEC), we have

$$\log|\mathcal{X}| - \frac{\log M^\star(n,\epsilon)}{n} \leq e^{-d\rho+\Theta(\log d|\mathcal{X}|)} - \Phi^{-1}(\epsilon)\cdot\frac{e^{-d\rho/2+\Theta(\log d|\mathcal{X}|)}}{\sqrt{n}} + \Theta\left(\frac{\log n}{n}\right).$$

- In particular, choosing $d = \rho^{-1}\log n$, we can achieve rates

$$\boxed{R_{n,\epsilon} \geq \log|\mathcal{X}| - O_\epsilon\left(\frac{\log n}{n}\right).}$$

Proof Sketch I: $\underbrace{\text{BIMS}}_{\text{bin.-i/p, memoryless, symmetric}}$ channels

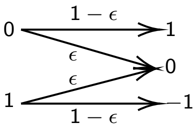
Background on BIMS channels

A BIMS channel obeys:

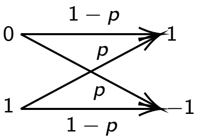
$$Y = (-1)^X \cdot Z,$$

for noise Z independent of X.

Examples:



Binary Erasure Channel (BEC)



Binary Symmetric Channel (BSC)

Proof for BIMS channels with uniform inputs

Consider $I^{(d)}$ for BIMS channels W with $P_X = \mathrm{Ber}(1/2)$.

Proof for BIMS channels with uniform inputs

Consider $I^{(d)}$ for BIMS channels W with $P_X = \mathrm{Ber}(1/2)$.

Lower bound: $\mathrm{H}(X \mid Y^d) \geq e^{-d\rho + \Theta(\log d)}$

We have

$$\mathrm{H}(X \mid Y^d) = \mathbb{E}\left[\log \frac{1}{P(X \mid Y^d)}\right].$$

Proof for BIMS channels with uniform inputs

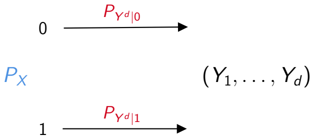
Consider $I^{(d)}$ for BIMS channels W with $P_X = \mathrm{Ber}(1/2)$.

Lower bound: $\mathrm{H}(X \mid Y^d) \geq e^{-d\rho + \Theta(\log d)}$

We have

$$\mathrm{H}(X \mid Y^d) = \mathbb{E}\left[\log \frac{1}{P(X \mid Y^d)}\right].$$

Note that in the binary hypothesis testing problem



the probability of correct decision is

$$\mathbb{E}\left[P(X \mid Y^d)\right] = 1 - e^{-\mathrm{C}\left(P_{Y^d|0}, P_{Y^d|1}\right) + \Theta(\log d)} = 1 - e^{-d\rho + \Theta(\log d)}.$$

Proof for BIMS channels with uniform inputs

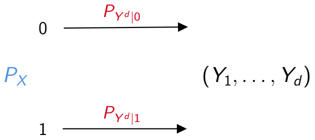
Consider $I^{(d)}$ for BIMS channels W with $P_X = \mathrm{Ber}(1/2)$.

Lower bound: $\mathrm{H}(X \mid Y^d) \geq e^{-d\rho + \Theta(\log d)}$

We have

$$\mathrm{H}(X \mid Y^d) = \mathbb{E}\left[\log \frac{1}{P(X \mid Y^d)}\right].$$

Note that in the binary hypothesis testing problem



the probability of correct decision is

$$\mathbb{E}\left[P(X \mid Y^d)\right] = 1 - e^{-\mathrm{C}\left(P_{Y^d|0}, P_{Y^d|1}\right) + \Theta(\log d)} = 1 - e^{-d\rho + \Theta(\log d)}.$$

Applying Jensen's inequality + a little algebra finishes the proof. □

Proof for BIMS channels with uniform inputs

Consider $I^{(d)}$ for BIMS channels W with $P_X = \mathrm{Ber}(1/2)$.

Upper bound: $\mathrm{H}(X \mid Y^d) \leq e^{-d\rho}$

A well-known lemma [e.g., Sasoglu (2012)]: for $P_X = \mathrm{Unif}(\{0,1\})$,

$$H(X \mid Y) \leq Z(W).$$

Proof for BIMS channels with uniform inputs

Consider $I^{(d)}$ for BIMS channels W with $P_X = \mathrm{Ber}(1/2)$.

Upper bound: $\mathrm{H}(X \mid Y^d) \leq e^{-d\rho}$

A well-known lemma [e.g., Sasoglu (2012)]: for $P_X = \mathrm{Unif}(\{0,1\})$,

$$H(X \mid Y) \leq Z(W).$$

Here, $W^{(d)}$ is also a BIMS channel, with

$$H(X \mid Y^d) \leq Z(W^{(d)}) = Z(W)^d = e^{-d\rho}. \quad \square$$

Proof for BIMS channels with uniform inputs

Consider $I^{(d)}$ for BIMS channels W with $P_X = \mathrm{Ber}(1/2)$.

Upper bound: $\mathrm{H}(X \mid Y^d) \leq e^{-d\rho}$

A well-known lemma [e.g., Sasoglu (2012)]: for $P_X = \mathrm{Unif}(\{0,1\})$,

$$H(X \mid Y) \leq Z(W).$$

Here, $W^{(d)}$ is also a BIMS channel, with

$$H(X \mid Y^d) \leq Z(W^{(d)}) = Z(W)^d = e^{-d\rho}. \quad \square$$

- [Hellman & Raviv (1970), Kanaya & Han (1995)] (or, our lower bound+[Levenshtein (2001)] and [Shannon, Gallager, Berlekamp (1967)]) allows us to complete a proof for general DMCs W.
- Our approach: a unified proof for $I^{(d)}$ and $V^{(d)}$ that allows for finite-blocklength results+multi-letter extensions.

Proof Sketch II: General DMCs

Sketch of proof strategy for general DMCs

Consider the mutual information $I^{(d)}$ for fixed P_X.

- We write

$$
\begin{aligned}
\mathrm{H}(X \mid Y^d) &= \mathbb{E}\left[\log \frac{1}{P(X \mid Y^d)}\right] \\
&= \sum_{x \in \mathcal{X}} P_X(x) \mathbb{E}\left[\log \frac{1}{P(X \mid Y^d)} \middle| X = x\right].
\end{aligned}
$$

Sketch of proof strategy for general DMCs

Consider the mutual information $I^{(d)}$ for fixed P_X.

- We write

$$\mathrm{H}(X \mid Y^d) = \mathbb{E}\left[\log \frac{1}{P(X \mid Y^d)}\right] = \sum_{x \in \mathcal{X}} P_X(x) \mathbb{E}\left[\log \frac{1}{P(X \mid Y^d)} \middle| X = x\right].$$

- Fix an $x \in \mathcal{X}$ and focus on the inner term:

$$\mathbb{E}\left[\log \frac{1}{P(X \mid Y^d)} \middle| X = x\right] = \int_0^\infty \Pr\left[\log \frac{1}{P(x \mid Y^d)} \geq t \middle| X = x\right] \mathrm{d}t.$$

Sketch of proof strategy for general DMCs

Consider the mutual information $I^{(d)}$ for fixed P_X.

- We write

$$\begin{aligned} \mathrm{H}(X \mid Y^d) &= \mathbb{E}\left[\log \frac{1}{P(X \mid Y^d)}\right] \\ &= \sum_{x \in \mathcal{X}} P_X(x) \mathbb{E}\left[\log \frac{1}{P(X \mid Y^d)} \middle| X = x\right]. \end{aligned}$$

- Fix an $x \in \mathcal{X}$ and focus on the inner term:

$$\mathbb{E}\left[\log \frac{1}{P(X \mid Y^d)} \middle| X = x\right] = \int_0^\infty \Pr\left[\log \frac{1}{P(x \mid Y^d)} \geq t \middle| X = x\right] \mathrm{d}t.$$

Let $p_x(t) := \Pr\left[-\log P_{X|Y^d}(x \mid Y^d) \geq t | X = x\right]$.

Bounding $p_x(t)$

Proposition

We have that

$$\frac{p_x(t)}{(|\mathcal{X}|-1)} \leq \max_{\tilde{x} \neq x} \Pr\left[\frac{P(x)P(Y^d \mid x)}{P(\tilde{x})P(Y^d \mid \tilde{x})} \leq \frac{e^{-(t-\log(|\mathcal{X}|-1))}}{1-e^{-t}} \middle| X = x \right]$$

and

$$p_x(t) \geq \max_{\tilde{x} \neq x} \Pr\left[\frac{P(x)P(Y^d \mid x)}{P(\tilde{x})P(Y^d \mid \tilde{x})} \leq \frac{e^{-t}}{1-e^{-t}} \middle| X = x \right].$$

Bounding $p_x(t)$

Proposition

We have that

$$\frac{p_x(t)}{(|\mathcal{X}|-1)} \le \max_{\tilde{x}\neq x} \Pr\left[\frac{P(x)P(Y^d \mid x)}{P(\tilde{x})P(Y^d \mid \tilde{x})} \le \frac{e^{-(t-\log(|\mathcal{X}|-1))}}{1-e^{-t}} \middle| X = x\right]$$

and

$$p_x(t) \ge \max_{\tilde{x}\neq x} \Pr\left[\frac{P(x)P(Y^d \mid x)}{P(\tilde{x})P(Y^d \mid \tilde{x})} \le \frac{e^{-t}}{1-e^{-t}} \middle| X = x\right].$$

Hence, let us concentrate on terms of the form

$$\Gamma_{x,\tilde{x}}(z) := \Pr\left[\frac{P(x)P(Y^d \mid x)}{P(\tilde{x})P(Y^d \mid \tilde{x})} \le \frac{ce^{-z}}{1-e^{-z}} \middle| X = x\right],$$

for $c > 0$.

Controlling the integral

Recall that we are primarily interested in

$$p_x(t) \approx e^{\Theta(\log |\mathcal{X}|)} \cdot \max_{\tilde{x} \neq x} \Gamma_{x,\tilde{x}}(z),$$

for suitable values of c, z, via $\int_0^\infty p_x(t) \mathrm{d}t$.

The following observation then holds:

Lemma

We have that

$$\int_0^\infty \max_{\tilde{x} \neq x} \Gamma_{x,\tilde{x}}(z) \mathrm{d}z = e^{\Theta(\log |\mathcal{X}|)} \cdot \max_{\tilde{x} \neq x} \int_0^\infty \Gamma_{x,\tilde{x}}(z) \mathrm{d}z.$$

Controlling the integral

Recall that we are primarily interested in

$$p_x(t) \approx e^{\Theta(\log|\mathcal{X}|)} \cdot \max_{\tilde{x} \neq x} \Gamma_{x,\tilde{x}}(z),$$

for suitable values of c, z, via $\int_0^\infty p_x(t) \mathrm{d}t$.

The following observation then holds:

Lemma

We have that

$$\int_0^\infty \max_{\tilde{x} \neq x} \Gamma_{x,\tilde{x}}(z) \mathrm{d}z = e^{\Theta(\log|\mathcal{X}|)} \cdot \max_{\tilde{x} \neq x} \int_0^\infty \Gamma_{x,\tilde{x}}(z) \mathrm{d}z.$$

Via Sanov's theorem+Laplace's method, it is possible to show that

$$\int_0^\infty \Gamma_{x,\tilde{x}}(z) \mathrm{d}z = e^{\Theta(\log d|\mathcal{X}|)} \cdot e^{-d \cdot \min_{\tilde{x} \neq x} \mathsf{C}(P_{Y|x}, P_{Y|\tilde{x}})}.$$

Putting everything together

Recall that we were ultimately interested in

$$\mathbb{E}\left[\log \frac{1}{P(X \mid Y^d)} \middle| X = x\right] = \int_0^\infty \Pr\left[\log \frac{1}{P(x \mid Y^d)} \geq t \middle| X = x\right] \mathrm{d}t$$
$$= \int_0^\infty p_x(t)\mathrm{d}t$$
$$\approx e^{\Theta(\log |\mathcal{X}|)} \cdot \int_0^\infty \max_{\tilde{x} \neq x} \Gamma_{x,\tilde{x}}(z)\mathrm{d}z$$
$$= e^{\Theta(\log d|\mathcal{X}|)} \cdot e^{-d \cdot \min_{\tilde{x} \neq x} \mathsf{C}(P_{Y|x}, P_{Y|\tilde{x}})}.$$

Putting everything together

Recall that we were ultimately interested in

$$\mathbb{E}\left[\log \frac{1}{P(X \mid Y^d)} \middle| X = x\right] = \int_0^\infty \Pr\left[\log \frac{1}{P(x \mid Y^d)} \geq t \middle| X = x\right] \mathrm{d}t$$
$$= \int_0^\infty p_x(t)\mathrm{d}t$$
$$\approx e^{\Theta(\log|\mathcal{X}|)} \cdot \int_0^\infty \max_{\tilde{x} \neq x} \Gamma_{x,\tilde{x}}(z)\mathrm{d}z$$
$$= e^{\Theta(\log d|\mathcal{X}|)} \cdot e^{-d \cdot \min_{\tilde{x} \neq x} \mathsf{C}(P_{Y|x}, P_{Y|\tilde{x}})}.$$

Hence, we have shown that

$$\mathsf{H}(X \mid Y^d) \approx e^{\Theta(\log d|\mathcal{X}|)} \cdot e^{-d \cdot \min_{\tilde{x} \neq x} \mathsf{C}(P_{Y|x}, P_{Y|\tilde{x}})},$$

with $\rho = \min_{\tilde{x} \neq x} \mathsf{C}(P_{Y|x}, P_{Y|\tilde{x}})$.

Proof sktech for channel dispersion

Consider the channel dispersion $V^{(d)}$ for fixed P_X.

- We write

$$V^{(d)} = \mathbb{E}\left[(\iota(X; Y^d) - I^{(d)})^2\right]$$
$$= \mathsf{V}(X) + \left(\mathbb{E}\left[\left(\log P(X \mid Y^d)\right)^2\right] - H(X \mid Y^d)^2\right) + \theta_d,$$

where

$$\theta_d = 2 \cdot \mathbb{E}\left[(\log P(X|Y^d) + \mathsf{H}(X|Y^d)) \cdot \left(\log \frac{1}{P(X)} - \mathsf{H}(X)\right)\right].$$

Proof sktech for channel dispersion

Consider the channel dispersion $V^{(d)}$ for fixed P_X.

- We write

$$V^{(d)} = \mathbb{E}\left[(\iota(X;Y^d) - I^{(d)})^2\right]$$
$$= \mathrm{V}(X) + \left(\mathbb{E}\left[\left(\log P(X \mid Y^d)\right)^2\right] - H(X \mid Y^d)^2\right) + \theta_d,$$

where

$$\theta_d = 2 \cdot \mathbb{E}\left[(\log P(X|Y^d) + \mathrm{H}(X|Y^d)) \cdot \left(\log \frac{1}{P(X)} - \mathrm{H}(X)\right)\right].$$

- Using an integral-based approach (as earlier), we can obtain that

$$\mathbb{E}\left[\left(\log P(X \mid Y^d)\right)^2\right] = e^{-d\rho + \Theta(\log d|\mathcal{X}|)}.$$

- From the earlier proof, we know that $H(X \mid Y^d) = e^{-d\rho+\Theta(\log d|\mathcal{X}|)}$.
- The cross-term θ_d can be handled by "averaging" over $x \in \mathcal{X}$.

Multi-view multi-letter channels

Extensions to multi-letter channels

- Our main result can be extended to multi-letter channels, including synchronization channels with equal output length across views.
- Consider such a multi-letter channel W_n of input length n, with

$$W_n^{(d)}(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_d \mid x^n) = \prod_{i=1}^{d} W_n(\boldsymbol{y}_i \mid x^n)$$

- Multi-letter variants of the mutual information $I_n^{(d)}$ and dispersion $V_n^{(d)}$ can be defined similar to earlier.

Extensions to multi-letter channels

- Our main result can be extended to multi-letter channels, including synchronization channels with equal output length across views.
- Consider such a multi-letter channel W_n of input length n, with

$$W_n^{(d)}(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_d \mid x^n) = \prod_{i=1}^{d} W_n(\boldsymbol{y}_i \mid x^n)$$

- Multi-letter variants of the mutual information $I_n^{(d)}$ and dispersion $V_n^{(d)}$ can be defined similar to earlier.

Theorem

We have that

$$I_n^{(d)} = \mathrm{H}(X^n) - \exp\left(-d\rho_n + \Theta(n \log d|\mathcal{X}|)\right), \text{ and}$$

$$\left|V_n^{(d)} - \mathrm{V}(X^n)\right| = \exp\left(-d\rho_n + \Theta(n \log d|\mathcal{X}|)\right),$$

where

$$\rho_n = \min_{u^n, \tilde{u}^n : u^n \neq \tilde{u}^n} \mathrm{C}(P_{\boldsymbol{Y}|u^n}, P_{\boldsymbol{Y}|\tilde{u}^n}).$$

Computations for the deletion channel

Consider the deletion channel Del(δ) that, on input $u^n \in \{0,1\}^n$, deletes each symbol u_i independently w.p. $\delta \in (0,1)$.

$$0\ 1\ 0\ 0\ 1\ 0\ 1\ 1 \longrightarrow 0\ 0\ 0\ 1\ 1$$

Theorem

For Del(δ), where $\delta \in (0,1)$, we have that

$$\limsup_{n\to\infty} \rho_n \leq -\log \delta.$$

Computations for the deletion channel

Consider the deletion channel $\mathrm{Del}(\delta)$ that, on input $u^n \in \{0,1\}^n$, deletes each symbol u_i independently w.p. $\delta \in (0,1)$.

$$0\ 1\ 0\ 0\ 1\ 0\ 1\ 1 \longrightarrow 0\ 0\ 0\ 1\ 1$$

Theorem

For $\mathrm{Del}(\delta)$, where $\delta \in (0,1)$, we have that

$$\limsup_{n\to\infty} \rho_n \leq -\log \delta.$$

- The proof follows by obtaining upper bounds on ρ_n by choosing special sequences $x^n, \tilde{x}^n$.
- The rate of convergence ρ_n for $\mathrm{Del}(\delta)$, for large enough n, is hence much slower than that of an n-letter DMC W_n, for which

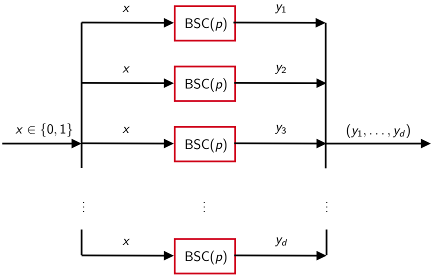
$$\rho_n(W_n) = n \cdot \rho(W).$$

Non-Asymptotic Bounds on Capacity

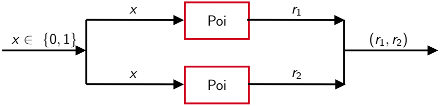
Non-asymptotic bounds for the $\mathrm{BSC}^{(d)}$

Consider the $\mathrm{BSC}^{(d)}(p)$. We introduce a closely related DMC $\mathrm{Poi}_d(p)$.

$\mathrm{BSC}^{(d)}(p)$



$\mathrm{Poi}_d(p)$

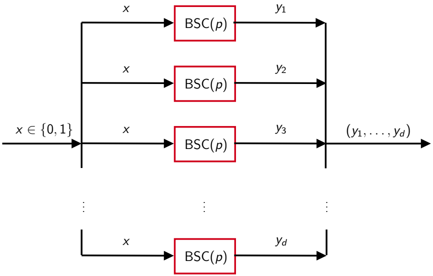


$$P_{R_1|0} = \mathrm{Poi}(d(1-p)),\ P_{R_1|1} = \mathrm{Poi}(dp)$$
$$P_{R_2|1} = \mathrm{Poi}(d(1-p)),\ P_{R_2|0} = \mathrm{Poi}(dp)$$

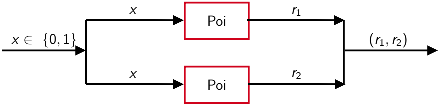
Non-asymptotic bounds for the BSC$^{(d)}$

Consider the BSC$^{(d)}(p)$. We introduce a closely related DMC Poi$_d(p)$.

BSC$^{(d)}(p)$



Poi$_d(p)$



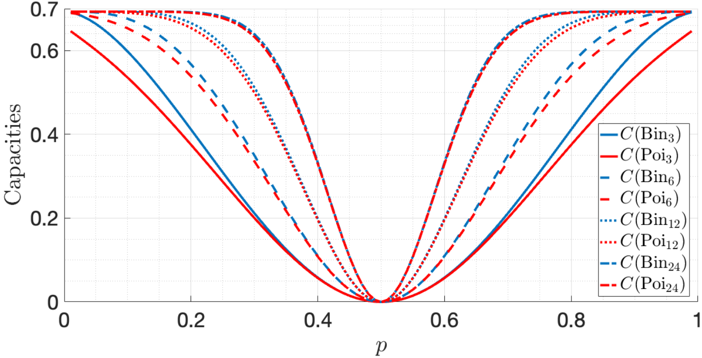
$$P_{R_1|0} = \text{Poi}(d(1-p)),\ P_{R_1|1} = \text{Poi}(dp)$$
$$P_{R_2|1} = \text{Poi}(d(1-p)),\ P_{R_2|0} = \text{Poi}(dp)$$

Theorem

$$C(\text{Poi}_d(p)) \le C(\text{Bin}_d(p)) \le C(\text{Poi}_d(p)) + e^{-d(1-Z(p))} - Z(p)^{2d},$$

where $Z(p) = 2\sqrt{p(1-p)}$.

Plots and comparisons



Plots comparing $C(\mathrm{Bin}_d(p))$ and $C(\mathrm{Poi}_d(p))$, for varying d.

A corollary for BIMS channels

- Any BIMS channel W with a finite output alphabet can be decomposed as
$$W_{Y|X=x} = \sum_{i=1}^{K} \epsilon_i \cdot W^{(i)}_{Y|X=x},$$
where $W^{(i)}_{Y|X=x}$ is the channel law of a BSC with crossover probability p_i, and $\epsilon_i > 0$, with $\sum_i \epsilon_i = 1$.
- Let P_s be the distribution on $\{1, \ldots, K\}$ with mass ϵ_i at point i.
$$C^{(d)}(W) = \mathbb{E}\left[C(J_1, \ldots, J_d)\right],$$
where $J_i \stackrel{\text{i.i.d.}}{\sim} P_s$, $1 \le i \le K$.

A corollary for BIMS channels

- Any BIMS channel W with a finite output alphabet can be decomposed as

$$W_{Y|X=x} = \sum_{i=1}^{K} \epsilon_i \cdot W^{(i)}_{Y|X=x},$$

  where $W^{(i)}_{Y|X=x}$ is the channel law of a BSC with crossover probability p_i, and $\epsilon_i > 0$, with $\sum_i \epsilon_i = 1$.
- Let P_s be the distribution on $\{1, \ldots, K\}$ with mass ϵ_i at point i.

$$C^{(d)}(W) = \mathbb{E}\left[C(J_1, \ldots, J_d)\right],$$

  where $J_i \overset{\text{i.i.d.}}{\sim} P_s$, $1 \le i \le K$.
- Now, suppose that $0 < p_1 \le p_2 \le \ldots \le p_K < 1/2$. Then,

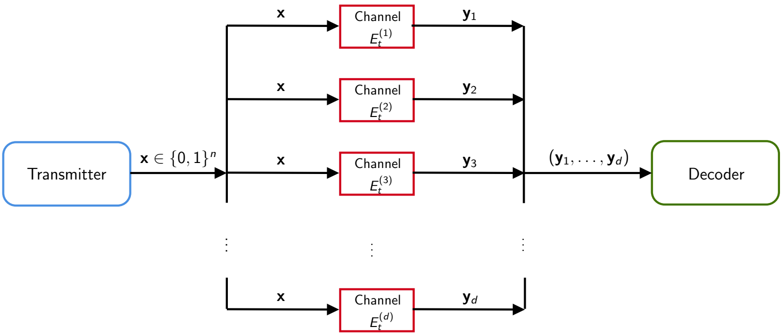
Corollary

$$C(\mathrm{Poi}_d(p_K)) \le C^{(d)}(W)$$
$$\le C(\mathrm{Poi}_d(p_1)) + \exp(-d(1 - Z(p_1))) - Z(p_1)^{2d}.$$

Towards **Q2**: Extending Levenshtein's results to the average case

The setting

Consider the following special setting:



d **distinct** views

- The views $\mathbf{y}_1, \ldots, \mathbf{y}_d$ are distinct and are drawn uniformly, without replacement, from a t-substitution error sphere around $\mathbf{x}$, i.e.,

$$\mathbf{y}_i = \mathbf{x} + \mathbf{e}_i, \text{ where } w_H(\mathbf{e}_i) = t \text{ and } \mathbf{e}_1, \ldots, \mathbf{e}_d \text{ are distinct.}$$

Some comments

- This setting is a generalization of the classical problem in [Levenshtein (2001)], wherein the (distinct) errors $\mathbf{e}_1, \ldots, \mathbf{e}_d$ were adversarially drawn from the Hamming sphere of radius t.

Some comments

- This setting is a generalization of the classical problem in [Levenshtein (2001)], wherein the (distinct) errors $\mathbf{e}_1, \ldots, \mathbf{e}_d$ were adversarially drawn from the Hamming sphere of radius t.

- From [Levenshtein (2001)], it can be argued that

$$\mathrm{Dec}(\mathbf{y}_1, \ldots, \mathbf{y}_d) = \mathbf{x},$$

  for all $\mathbf{x} \in \{0, 1\}^n$, only if $d > N_{\mathrm{Lev}} := 2\binom{n-1}{t-1}$, for any decoder Dec.

Some comments

- This setting is a generalization of the classical problem in [Levenshtein (2001)], wherein the (distinct) errors $\mathbf{e}_1, \ldots, \mathbf{e}_d$ were adversarially drawn from the Hamming sphere of radius t.

- From [Levenshtein (2001)], it can be argued that

$$\mathrm{Dec}(\mathbf{y}_1, \ldots, \mathbf{y}_d) = \mathbf{x},$$

for all $\mathbf{x} \in \{0,1\}^n$, only if $d > N_{\mathrm{Lev}} := 2\binom{n-1}{t-1}$, for any decoder Dec.

- Moreover, Levenshtein argued that the bit-wise majority decoder

$$\mathrm{Dec}^{\mathrm{Maj}}(\mathbf{y}_1, \ldots, \mathbf{y}_d) = (\mathrm{Maj}(y_{1,1}, \ldots, y_{d,1}), \ldots, \mathrm{Maj}(y_{1,n}, \ldots, y_{d,n}))$$

is "optimal", in that

$$\mathrm{Dec}^{\mathrm{Maj}}(\mathbf{y}_1, \ldots, \mathbf{y}_d) = \mathbf{x},$$

for all $\mathbf{x} \in \{0,1\}^n$, whenever $d > N_{\mathrm{Lev}}$.

How does Dec^{Maj} perform in the average case?

Fix the bit-wise majority decoder Dec^{Maj}. Given views $\mathbf{y}_1, \ldots, \mathbf{y}_d$, construct the array

$M(\mathbf{y}_1, \ldots, \mathbf{y}_d) =$

$y_{1,1}$	$y_{1,2}$					$y_{1,n}$	$\mathbf{y}_1$
$y_{2,1}$							$\mathbf{y}_2$
							$\mathbf{y}_3$
							$\vdots$
$y_{d,1}$						$y_{d,n}$	$\mathbf{y}_d$

Let $m(n, t, d)$ denote the number of distinct arrays $M(\mathbf{y}_1, \ldots, \mathbf{y}_d)$ that lead to correct reconstruction of $\mathbf{x}$ via Dec^{Maj}.

A recurrence relation for estimating $m(n, t, d)$

From the structure of the $\mathrm{Dec}^{\mathrm{Maj}}$ decoder, the following lemma holds:

Lemma
We have that

$$m(n,t,d) \geq m(n-3,t,d) + m(n-3,t-1,d) \cdot a_d,$$

where $a_d = \sum_{\substack{k_1+k_2+k_3=N \\ k_1,k_2,k_3 \leq \lfloor \frac{d-1}{2} \rfloor}} \binom{d}{k_1,k_2,k_3}$.

- With the aid of suitable initial conditions

$$m(n,t,d) = 0, \text{ if } d < 3 \text{ or } d > n-3(t-1),$$

the above recurrence relation can be explicitly solved, giving a lower bound on $m(n, t, d)$.

Estimating the average # of views for reconstruction

- Via the previous arguments, we can derive a lower bound on

$$P_{\mathsf{c}}(n,t,d) := \frac{m(n,t,d)}{\binom{\binom{n}{t}}{d}}.$$

- This allows us to obtain an upper bound on $\mathbb{E}[\#\text{ views for reconst.}]$ via a union-bound style argument.

- Moreover, we have the following lemma:

> **Lemma**
> *When $t = 1$, we have*
>
> $$\mathbb{E}[\#\text{ views for reconst.}] = N_{\mathsf{Lev}} = 3.$$

Future work

- Bounds/exact computation of ρ_n for general synchronization channels and other channels with memory

- Relating $\mathrm{Dec}^{\mathrm{Maj}}$ to the "optimal" decoder in terms of $\mathbb{E}[\#\text{ views for reconst.}]$.

References

1. V. Arvind Rameshwar and Nir Weinberger, "**Information rates over multi-view channels**," under review at the IEEE Transactions on Information Theory; presented at the IEEE ISIT 2024.

2. Vivian Papadopoulou, V. Arvind Rameshwar, Antonia Wachter-Zeh, "**How many noisy copies of a sequence are needed on average for exact reconstruction?**," accepted to the ITW 2024.

Thank You!