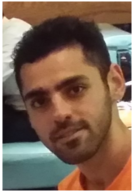
Sub-optimality of superposition coding for three or more receivers

Chandra Nair, &



Mehdi Yazdanpanah

The Chinese University of Hong Kong

CISS 2018

23 Mar, 2018

The **story** of superposition coding is the **story** of broadcast channels with **degradation** (receivers or message sets)

The **story** of superposition coding is the **story** of broadcast channels with **degradation** (receivers or message sets)

It is also the **story** of auxiliary random variables

In the beginning

- Cover (1972) proposed the superposition coding achievable region for degraded broadcast channels
  - He used an **auxiliary** variable to represent the message for the weaker of two receivers.
- Bergmans (1973, Gaussian) and Gallager (1974, discrete-memoryless) established the optimality of superposition coding for the degraded broadcast channel
  - Gallager's proof of optimality of superposition coding naturally extended to a sequence of degraded receivers

In the beginning

- Cover (1972) proposed the superposition coding achievable region for degraded broadcast channels
  - He used an **auxiliary** variable to represent the message for the weaker of two receivers.
- Bergmans (1973, Gaussian) and Gallager (1974, discrete-memoryless) established the optimality of superposition coding for the degraded broadcast channel
  - Gallager's proof of optimality of superposition coding naturally extended to a sequence of degraded receivers

The optimality of superposition coding region was then established for

- Weaker notions of **weaker** receiver in a two-receiver broadcast channel
  - Less Noisy (Korner-Marton 75), More Capable (El Gamal 79)
- Degraded message sets (Korner-Marton 77)
  - Comparison of the sizes of the images of a set via two channels

In the beginning

- Cover (1972) proposed the superposition coding achievable region for degraded broadcast channels
  - He used an **auxiliary** variable to represent the message for the weaker of two receivers.
- Bergmans (1973, Gaussian) and Gallager (1974, discrete-memoryless) established the optimality of superposition coding for the degraded broadcast channel
  - Gallager's proof of optimality of superposition coding naturally extended to a sequence of degraded receivers

The optimality of superposition coding region was then established for

- Weaker notions of **weaker** receiver in a two-receiver broadcast channel
  - Less Noisy (Korner-Marton 75), More Capable (El Gamal 79)
- Degraded message sets (Korner-Marton 77)
  - Comparison of the sizes of the images of a set via two channels

A (seemingly) natural "Implication": Auxiliaries (in superposition coding) **captured** the **coarser** message.

More recent times ..

The optimality of superposition coding was established for

- Three receiver less-noisy broadcast channel (**N**-Wang 2010)

More recent times ..

The optimality of superposition coding was established for

- Three receiver less-noisy broadcast channel (**N**-Wang 2010)

However for the following three-receiver broadcast channel setting:

- Receivers Y_2, Y_3 wish to decode message M_0
- Receiver Y_1 wishes to decode messages M_0, M_1

superposition coding region was not optimal (El Gamal-**N** 2009)

More recent times ..

The optimality of superposition coding was established for

- Three receiver less-noisy broadcast channel (**N**-Wang 2010)

However for the following three-receiver broadcast channel setting:

- Receivers Y_2, Y_3 wish to decode message M_0
- Receiver Y_1 wishes to decode messages M_0, M_1

superposition coding region was not optimal (El Gamal-**N** 2009)

Demonstrates that auxiliaries **do not** capture the coarser message.

- Associating an auxiliary with information decoded by groups of receivers improved the achievable region
  - U_{123}, U_{12}, U_{13}, and $U_1 = X$.
  - The achievable region was no longer a superposition coding region, it also involved the **other** (old) idea: **random binning**

A (seemingly) natural "Implication" (Take 2): Auxiliaries (~~in superposition coding~~) **captured** the information decoded by groups of receivers.

- Binning and Superposition both present

Consistency

The revised intuition about auxiliaries is **consistent** with Marton's achievable scheme for two-receiver broadcast channels with private messages

- The region employs three auxiliaries: U_1, U_2, U_{12}
- It is known that this region is strictly better than the one with only U_1, U_2 (even for private messages).

Consistency

The revised intuition about auxiliaries is **consistent** with Marton's achievable scheme for two-receiver broadcast channels with private messages

- The region employs three auxiliaries: U_1, U_2, U_{12}
- It is known that this region is strictly better than the one with only U_1, U_2 (even for private messages).

Open Question

Is Marton's achievable scheme for two-receiver broadcast channels optimal?

Three-or-more receivers

A three-receiver broadcast channel with private messages would have seven auxiliaries

$$U_{123}, U_{12}, U_{13}, U_{23}, U_1, U_2, U_3$$

and a natural extension of Marton's scheme would have two layers of superposition coding and binning between random variables in each layer.

A succinct clean representation of the rate constraints is not available

Consistency

The revised intuition about auxiliaries is **consistent** with Marton's achievable scheme for two-receiver broadcast channels with private messages

- The region employs three auxiliaries: U_1, U_2, U_{12}
- It is known that this region is strictly better than the one with only U_1, U_2 (even for private messages).

Open Question

Is Marton's achievable scheme for two-receiver broadcast channels optimal?

Three-or-more receivers

A three-receiver broadcast channel with private messages would have seven auxiliaries

$$U_{123}, U_{12}, U_{13}, U_{23}, U_1, U_2, U_3$$

and a natural extension of Marton's scheme would have two layers of superposition coding and binning between random variables in each layer.

A succinct clean representation of the rate constraints is not available

However: this region is not optimal (Padlakanda and Pradhan (2015))

Still one fundamental setting remained

Consider the setting:

- Receivers Y_3 wish to decode message M_0
- Receiver Y_1, Y_2 wishes to decode messages M_0, M_1

Interpretation of auxiliaries: U_{123} and $U_{12} = X$, and only superposition coding

Question: Is superposition coding optimal?

(Note: The first layer superposition coding of the previous private message setting)

Still one fundamental setting remained

Consider the setting:

- Receivers Y_3 wish to decode message M_0
- Receiver Y_1, Y_2 wishes to decode messages M_0, M_1

Interpretation of auxiliaries: U_{123} and $U_{12} = X$, and only superposition coding

Question: Is superposition coding optimal?

(Note: The first layer superposition coding of the previous private message setting)

- **Open problems:**

8.1. What is the capacity region of the general 3-receiver DM-BC with one common message to all three receivers and one private message to one receiver?

8.2. Is superposition coding optimal for the general 3-receiver DM-BC with one message to all three receivers and another message to two receivers?

8.3. What is the sum-capacity of the binary skew-symmetric broadcast channel?

8.4. Is Marton's inner bound tight in general?

Still one fundamental setting remained

Consider the setting:

- Receivers Y_3 wish to decode message M_0
- Receiver Y_1, Y_2 wishes to decode messages M_0, M_1

Interpretation of auxiliaries: U_{123} and $U_{12} = X$, and only superposition coding

Question: Is superposition coding optimal?

(Note: The first layer superposition coding of the previous private message setting)

3.1 SOME BASIC MATHEMATICAL PROBLEMS OF MULTIUSER SHANNON THEORY

I. Csiszár

Mathematical Institute of the
Hungarian Academy of Sciences
Budapest, Hungary

2. Image Size Characterization Problem.

The η-image size $g_W(A, \eta)$ of a set $A \subset X^n$ over a discrete memoryless channel (DMC) $\{W : X \to Y\}$ is the minimum cardinality of $B \subset Y^n$ such that $W^n(B \mid \mathbf{x}) \geq \eta$ for each $\mathbf{x} \in A$. The problem is to find, for a distribution P on X and DMCs $\{W_i : X \to Y_i\}$, $i = 1, \ldots, k$, a single-letter characterization of the limit of the sets of all $(k+1)$-dimensional vectors

$$\left[\frac{1}{n}\log|A|, \frac{1}{n}\log g_{W_1}(A, \eta), \ldots, \frac{1}{n}\log g_{W_k}(A, \eta)\right].$$

Here $A \subset X^n$ is any set of P-typical sequences, and $0 < \eta < 1$ is fixed (the result is independent of η).

Csiszar's open problem is very closely tied to finding the capacity region

A remark

Körner had proposed a region (1984) for the image size characterization over three channels

Theorem: For every RV's T, U, and V such that

$$TUV \to S \to XYZ,$$

nonnegative numbers t, t', and t'', the point (r_x, r_y, r_z) with coordinates

$$\begin{aligned} r_x &\triangleq \min\big[H(X), H(X|T)+t, H(X|TU)+t', \\ &\qquad H(X|TUV)+t''\big], \\ r_y &\triangleq \min\big[H(Y), H(Y|T)+t, H(Y|TU)+t', \\ &\qquad H(Y|TUV)+t''\big], \\ r_z &\triangleq \min\big[H(Z), H(Z|T)+t, H(Z|TU)+t', \\ &\qquad H(Z|TUV)+t''\big] \end{aligned} \tag{30}$$

is an element of $\mathscr{H}(X; Y; Z|S)$. □

Suboptimality of superposition coding

Superposition coding is sub-optimal for the setting (**N**-Yazdanpanah 2017)

- Constructed a channel whose 2-letter superposition-coding region was larger than the 1-letter one

Suboptimality of superposition coding

Superposition coding is sub-optimal for the setting (**N**-Yazdanpanah 2017)

- Constructed a channel whose 2-letter superposition-coding region was larger than the 1-letter one

Note: The same (counter)-example showed that Korner's region is a proper subset of $\mathcal{H}(X;Y;Z|S)$.

Suboptimality of superposition coding

Superposition coding is sub-optimal for the setting (**N**-Yazdanpanah 2017)

- Constructed a channel whose 2-letter superposition-coding region was larger than the 1-letter one

Note: The same (counter)-example showed that Korner's region is a proper subset of $\mathcal{H}(X;Y;Z|S)$.

Remarks

- It took us three years to get counterexamples
  - The optimization problems involved are non-convex
  - In small dimensions counter-examples lie in a set of very small "size" (random sampling does not work)
- Question: Why did we believe that superposition coding was sub-optimal?
- More generally, why do we believe certain regions are optimal while certain others are not?

Suboptimality of superposition coding

Superposition coding is sub-optimal for the setting (**N**-Yazdanpanah 2017)

- Constructed a channel whose 2-letter superposition-coding region was larger than the 1-letter one

Note: The same (counter)-example showed that Korner's region is a proper subset of $\mathcal{H}(X;Y;Z|S)$.

Remarks

- It took us three years to get counterexamples
  - The optimization problems involved are non-convex
  - In small dimensions counter-examples lie in a set of very small "size" (random sampling does not work)
- Question: Why did we believe that superposition coding was sub-optimal?
- More generally, why do we believe certain regions are optimal while certain others are not?
  - An unpublished conjecture: Local-tensorization **implies** global tensorization

Suboptimality of superposition coding

Superposition coding is sub-optimal for the setting (**N**-Yazdanpanah 2017)

- Constructed a channel whose 2-letter superposition-coding region was larger than the 1-letter one

Note: The same (counter)-example showed that Korner's region is a proper subset of $\mathcal{H}(X;Y;Z|S)$.

This talk: Focus on the (counter)-example

- Bounds on the capacity region
- Shed light to properties of good codes for this channel

Multilevel product broadcast erasure channel (ISIT '17)

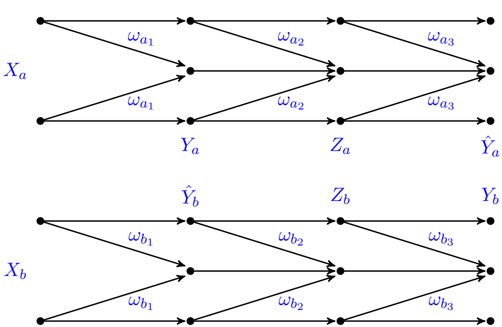
$X_a \to Y_a : BEC(e_a),\ X_b \to Y_b : BEC(e_b)$

$X_a \to \hat{Y}_a : BEC(\hat{e}_a),\ X_b \to \hat{Y}_b : BEC(\hat{e}_b)$

$X_a \to Z_a : BEC(f_a),\ X_b \to Z_b : BEC(f_b)$

$$\hat{e}_a \ge f_a \ge e_a \quad \& \quad e_b \ge f_b \ge \hat{e}_b$$

$$C_Z = (1 - f_a) + (1 - f_b)$$

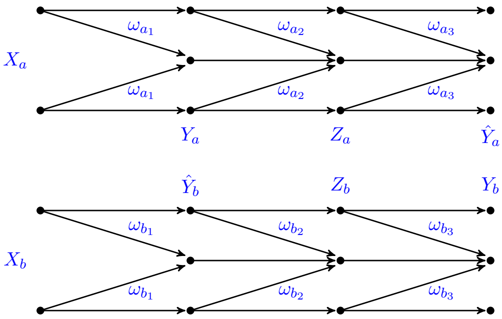


Multilevel product broadcast erasure channel (ISIT '17)

$$X_a \to Y_a : BEC(e_a),\ X_b \to Y_b : BEC(e_b)$$
$$X_a \to \hat{Y}_a : BEC(\hat{e}_a),\ X_b \to \hat{Y}_b : BEC(\hat{e}_b)$$
$$X_a \to Z_a : BEC(f_a),\ X_b \to Z_b : BEC(f_b)$$

$$\hat{e}_a \geq f_a \geq e_a \quad \& \quad e_b \geq f_b \geq \hat{e}_b$$

$$C_Z = (1 - f_a) + (1 - f_b)$$



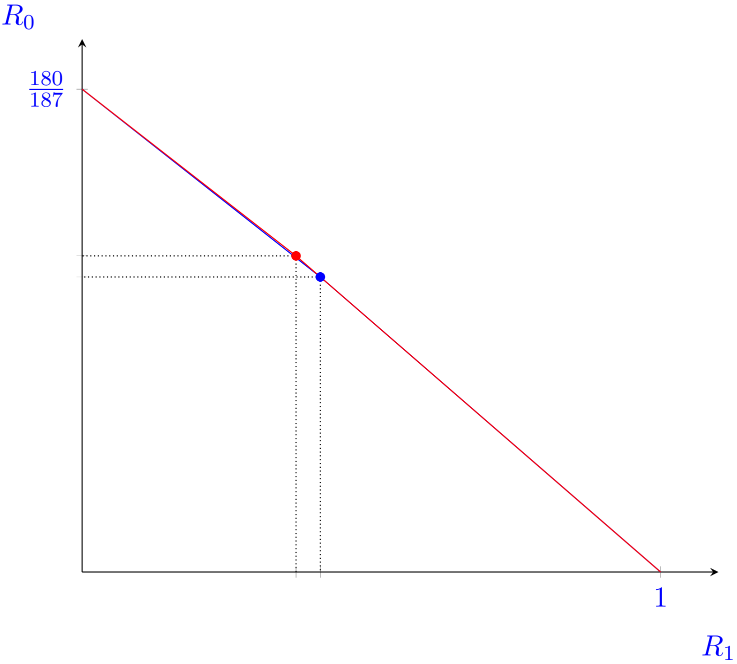
Theorem

For $e_a = 1/2$, $\hat{e}_a = 1$, $f_a = 17/22$
$e_b = 1/2$, $\hat{e}_b = 0$, $f_b = 9/34$

1-letter SC : $R_0 + R_1 \leq 1$ and $\frac{11}{10}R_0 + R_1 \leq \frac{18}{17} = \frac{11}{10}C_Z$

2-letter SC : $R_0 + R_1 \leq 1$ and $\frac{484}{435}R_0 + R_1 \leq \frac{528}{493} = \frac{484}{435}C_Z$

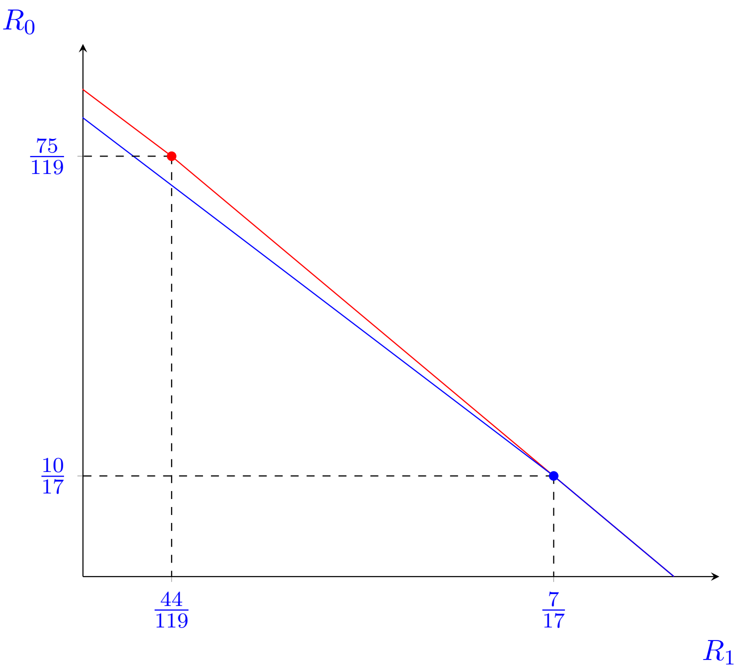
Plot



2-letter & 1-letter

SC regions

Plot



2-letter & 1-letter

SC regions

1-letter SC

The distribution that achieves the **corner-point**.

- Let U be a ternary random variable

$$\begin{cases} P(U=0)=\frac{13}{34} & (X_a,X_b)|\{U=0\}=(0,0) \\ P(U=1)=\frac{7}{34} & (X_a,X_b)|\{U=1\}=(M,0) \\ P(U=2)=\frac{14}{34} & (X_a,X_b)|\{U=2\}=(M,M) \end{cases}$$

where M is an unbiased binary random variable

- Let Q be the random variable that symmetrizes the distribution of X
- Let $\tilde{U}=(U,Q)$ and substitute $(\tilde{U},X)$ into SC

$$R_0 \le I(\tilde{U};Z)=\frac{10}{17}$$

$$R_0+R_1 \le I(\tilde{U};Z)+I(X;Y|\tilde{U})=1 \qquad R_0+R_1 \le I(X;Y)=1$$

$$R_0+R_1 \le I(\tilde{U};Z)+I(X;\hat{Y}|\tilde{U})=1 \qquad R_0+R_1 \le I(X;\hat{Y})=1$$

The distribution that achieves the **corner-point**.

M_1 & M_2 two independent unbiased binary r.v.

$$\begin{cases} P(U=0)=\frac{20}{119} & (X_{a1},X_{b1},X_{a2},X_{b2})|\{U=0\}=(0,0,0,0) \\ P(U=1)=\frac{11}{119} & (X_{a1},X_{b1},X_{a2},X_{b2})|\{U=1\}=(M_1,0,M_2,0) \\ P(U=2)=\frac{88}{119} & (X_{a1},X_{b1},X_{a2},X_{b2})|\{U=2\}=(M_1,M_1,M_1,0) \end{cases}$$

- Let Q be the random variable that symmetrizes the distribution of X
- Let $\tilde{U}=(U,Q)$ and substitute $(\tilde{U},X)$ into SC $\Rightarrow (R_0,R_1)=(\frac{75}{119},\frac{44}{119})$

The distribution that achieves the **corner-point**.

M_1 & M_2 two independent unbiased binary r.v.

$$\begin{cases} P(U=0)=\frac{20}{119} & (X_{a1},X_{b1},X_{a2},X_{b2})|\{U=0\}=(0,0,0,0) \\ P(U=1)=\frac{11}{119} & (X_{a1},X_{b1},X_{a2},X_{b2})|\{U=1\}=(M_1,0,M_2,0) \\ P(U=2)=\frac{88}{119} & (X_{a1},X_{b1},X_{a2},X_{b2})|\{U=2\}=(M_1,M_1,M_1,0) \end{cases}$$

- Let Q be the random variable that symmetrizes the distribution of X
- Let $\tilde{U}=(U,Q)$ and substitute $(\tilde{U},X)$ into SC $\Rightarrow (R_0,R_1)=(\frac{75}{119},\frac{44}{119})$

Observation

- A linear code achieves the 2-letter region

A natural question

Let $\mathbf{M}=(M_1,...,M_m)$ be mutually independent unbiased bits.
Let $X_a^n=\mathbf{AM}$ and $X_b^n=\mathbf{BM}$, where $\mathbf{A},\mathbf{B}$ are $n\times m$ matrices.
What is the rate region achieved by such a linear coding scheme.
(variables are m,$\mathbf{A},\mathbf{B}$).

Routine Idea: Intersection of the two capacity regions (ignoring one of the users)

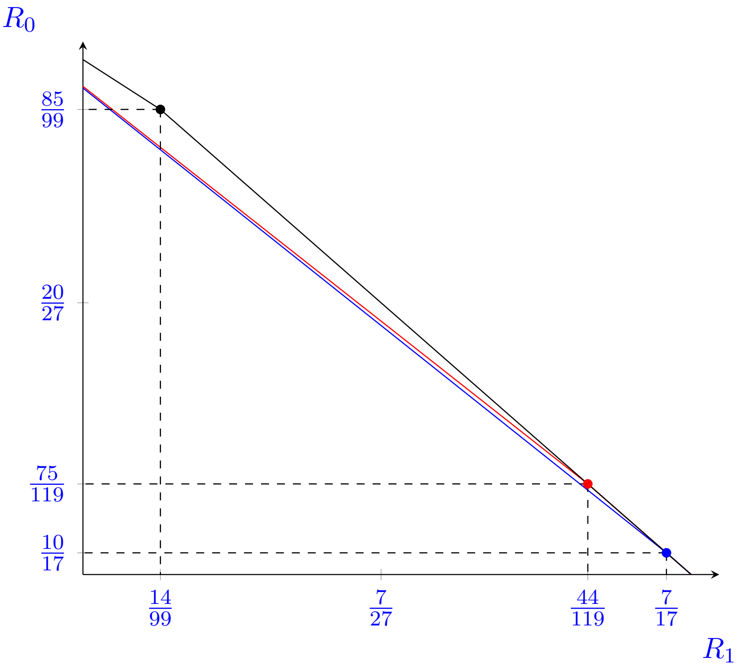
Theorem

For
$$e_a = 1/2 \qquad \hat{e}_a = 1 \qquad f_a = 17/22$$
$$e_b = 1/2 \qquad \hat{e}_b = 0 \qquad f_b = 9/34$$

ignoring $\hat{Y}$:
$$R_0 + R_1 \leq 1 \quad \text{and} \quad \frac{11}{5}R_0 + R_1 \leq \frac{11}{5}C_Z$$

ignoring Y :
$$R_0 + R_1 \leq 1 \quad \text{and} \quad \frac{34}{25}R_0 + R_1 \leq \frac{34}{25}C_Z$$

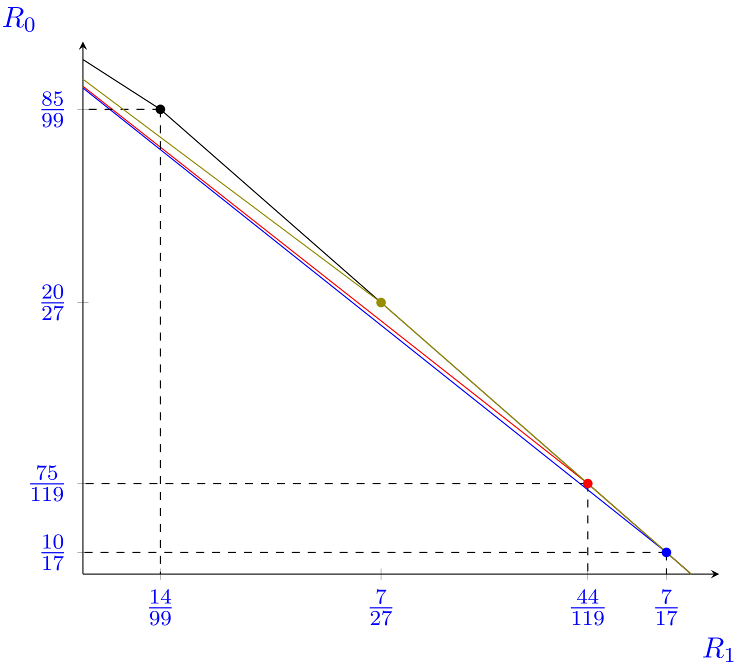
Plot



2-letter & 1-letter SC

Trad. Outer Bound

Plot



2-letter & 1-letter SC

Trad. Outer Bound

New Outer Bound

Idea for new outer-bound

From limiting n-letter **inner bound** that goes to capacity:

Idea for new outer-bound

From limiting n-letter **inner bound** that goes to capacity:

Theorem (Concentration of mutual information over memoryless product erasure channel)

Consider a product erasure channel, $W_a(y_a|x_a) \otimes W_b(y_b|x_b)$, mapping X_a, X_b to Y_a, Y_b with erasure probabilities ϵ_a, ϵ_b, respectively. Then

$$I(X_a^n, X_b^n; Y_a^n, Y_b^n) = \mathcal{H}(\lfloor n(1-\epsilon_a) \rfloor, \lfloor n(1-\epsilon_b) \rfloor) + O\left(\sqrt{n \log n}\right),$$

where

$$\mathcal{H}_n(k,l) = \frac{1}{\binom{n}{k}\binom{n}{l}} \sum_{S,T \subseteq [n]: |S|=k, |T|=l} H(X_{aS}, X_{bT}).$$

Idea for new outer-bound

From limiting n-letter **inner bound** that goes to capacity:

Theorem (Concentration of mutual information over memoryless product erasure channel)

Consider a product erasure channel, $W_a(y_a|x_a) \otimes W_b(y_b|x_b)$, *mapping* X_a, X_b *to* Y_a, Y_b *with erasure probabilities* ϵ_a, ϵ_b, *respectively. Then*

$$I(X_a^n, X_b^n; Y_a^n, Y_b^n) = \mathcal{H}(\lfloor n(1-\epsilon_a) \rfloor, \lfloor n(1-\epsilon_b) \rfloor) + O\left(\sqrt{n \log n}\right),$$

where

$$\mathcal{H}_n(k,l) = \frac{1}{\binom{n}{k}\binom{n}{l}} \sum_{S,T \subseteq [n]: |S|=k, |T|=l} H(X_{aS}, X_{bT}).$$

Using (essentially) sub-modularity of entropy, we can establish that

$$\limsup_n \max_{p(x_a^n, x_b^n)} \frac{1}{n} \left(\frac{85}{160} \mathcal{H}(\frac{n}{2}, \frac{n}{2}) + \frac{75}{160} \mathcal{H}(0, n) - \frac{187}{160} \mathcal{H}(\frac{5n}{22}, \frac{25n}{34}) \right) \leq 0.$$

Outer bound continued

Theorem (Outer bound)

Any achievable rate pair (R_0, R_1) *must satisfy the constraints.*

$$R_0 + R_1 \leq 1 \;\; \textit{and} \;\; \frac{187}{160} R_0 + R_1 \leq \frac{18}{16}.$$

Outer bound continued

Theorem (Outer bound)

Any achievable rate pair (R_0, R_1) must satisfy the constraints.

$$R_0 + R_1 \leq 1 \ \ \textit{and} \ \ \frac{187}{160}R_0 + R_1 \leq \frac{18}{16}.$$

Achievability

If there is a non-trivial collection (X_a^n, X_b^n) such that

$$\begin{aligned}
&\mathcal{H}_n(\frac{n}{2}, \frac{n}{2}) = \mathcal{H}_n(\frac{n}{2}, \frac{25n}{34}) + o(n),\\
&\frac{5}{11}\mathcal{H}_n(\frac{n}{2}, \frac{25n}{34}) + \frac{6}{11}\mathcal{H}_n(0, \frac{25n}{34}) = \mathcal{H}(\frac{5n}{22}, \frac{25n}{34}) + o(n),\\
&\frac{8}{17}\mathcal{H}_n(0, n) + \frac{9}{17}\mathcal{H}_n(0, \frac{n}{2}) = \mathcal{H}_n(0, \frac{25n}{34}) + o(n),\\
&\frac{17}{25}\mathcal{H}_n(0, \frac{25n}{34}) = \mathcal{H}_n(0, \frac{n}{2}) + o(n),
\end{aligned}$$

then there are non-trivial points of the outer bound that are achievable.

Outer bound continued

Theorem (Outer bound)

Any achievable rate pair (R_0, R_1) *must satisfy the constraints.*

$$R_0 + R_1 \leq 1 \;\; \textit{and} \;\; \frac{187}{160} R_0 + R_1 \leq \frac{18}{16}.$$

Achievability

If there is a non-trivial collection (X_a^n, X_b^n) such that

$$\mathcal{H}_n(\frac{n}{2}, \frac{n}{2}) = \mathcal{H}_n(\frac{n}{2}, \frac{25n}{34}) + o(n),$$
$$\frac{5}{11}\mathcal{H}_n(\frac{n}{2}, \frac{25n}{34}) + \frac{6}{11}\mathcal{H}_n(0, \frac{25n}{34}) = \mathcal{H}(\frac{5n}{22}, \frac{25n}{34}) + o(n),$$
$$\frac{8}{17}\mathcal{H}_n(0, n) + \frac{9}{17}\mathcal{H}_n(0, \frac{n}{2}) = \mathcal{H}_n(0, \frac{25n}{34}) + o(n),$$
$$\frac{17}{25}\mathcal{H}_n(0, \frac{25n}{34}) = \mathcal{H}_n(0, \frac{n}{2}) + o(n),$$

then there are non-trivial points of the outer bound that are achievable.

Suggests: "MDS-like" (linear increase followed by flat region) code-construction for X_a^n, X_b^n

Conclusion

Observations

- Sub-optimality of superposition coding region
- Sub-optimality of Korner's image-size characterization
- Linear code achieves 2-letter inner bound
- A new (explicit) outer bound from limiting n-letter inner bound
- Outer bound yields insights into structure of good codes