Channel coding: finite blocklength results

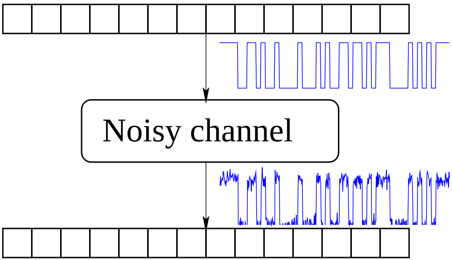
Plan:

1. Overview on the example of BSC

---

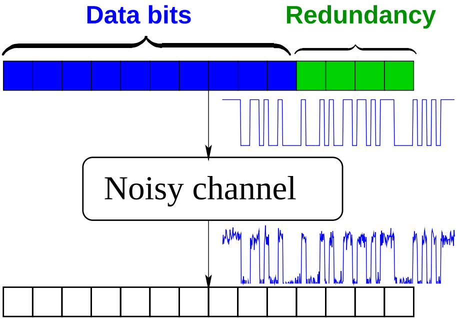
2. Converse bounds
3. Achievability bounds
4. Channel dispersion
5. Applications: performance of real-world codes
   Extensions: codes with feedback

Abstract communication problem



Goal: Decrease corruption of data caused by noise

Channel coding: principles

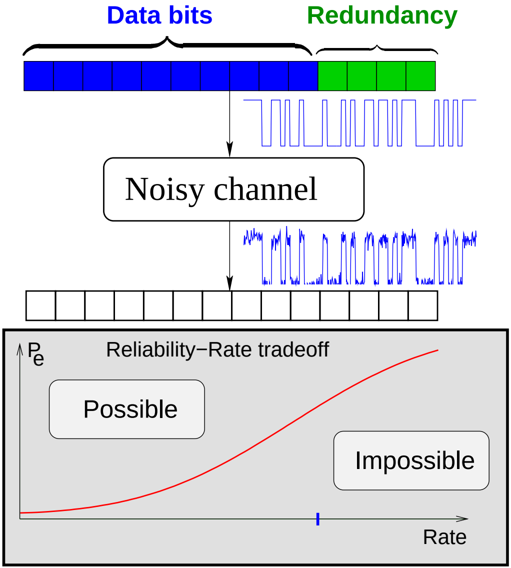


Goal: Decrease corruption of data caused by noise

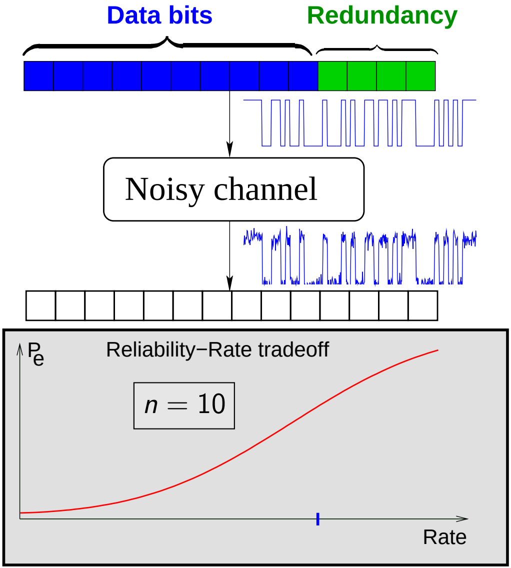
Solution: *Code* to diminish probability of error P_e.

Key metrics: Rate and P_e

Channel coding: principles



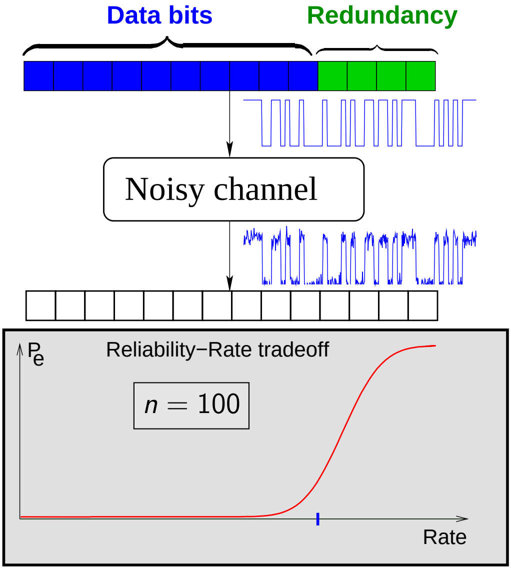
Channel coding: principles



Decreasing P_e further:

1. More redundancy
   Bad: loses rate
2. Increase blocklength!

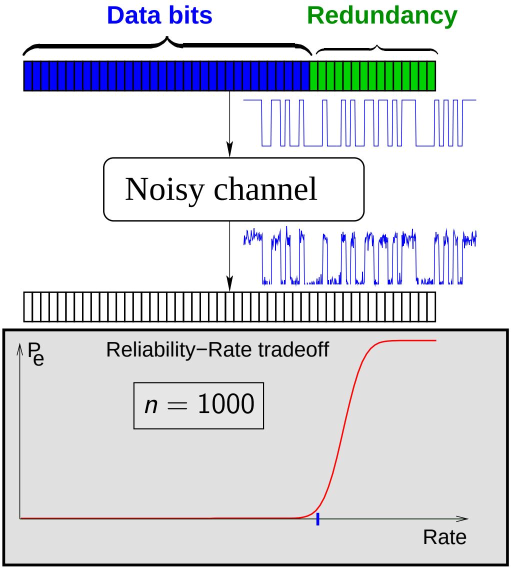
Channel coding: principles



Decreasing P_e further:

1. More redundancy
   Bad: loses rate
2. Increase blocklength!

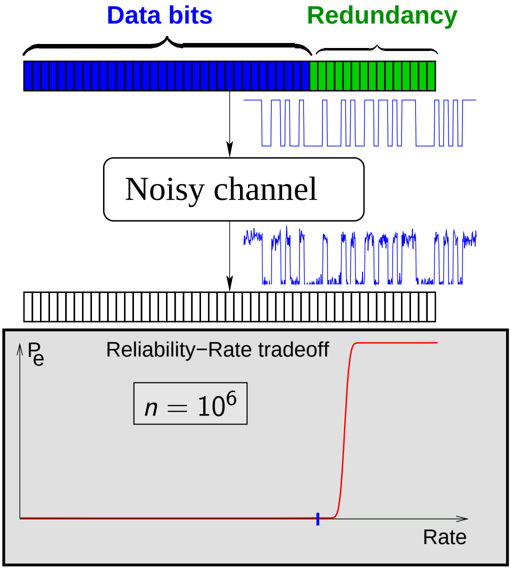
Channel coding: principles



Decreasing P_e further:

1. More redundancy
   Bad: loses rate
2. Increase blocklength!

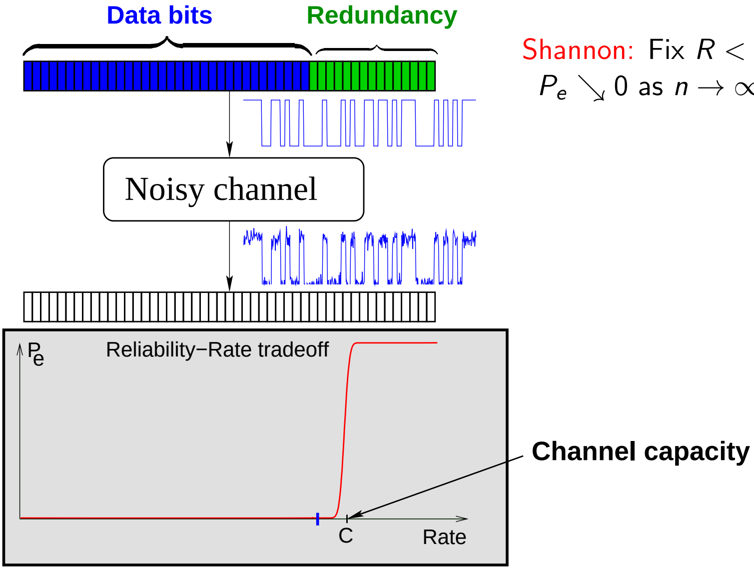
Channel coding: principles



Decreasing P_e further:

1. More redundancy
   Bad: loses rate
2. Increase blocklength!

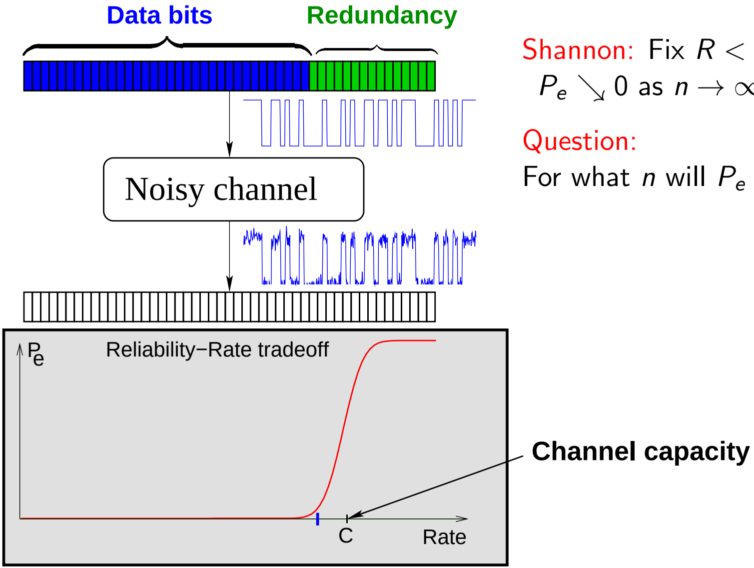
Channel coding: Shannon capacity



Shannon: Fix $R < C$

$P_e \searrow 0$ as $n \to \infty$

Channel coding: Shannon capacity



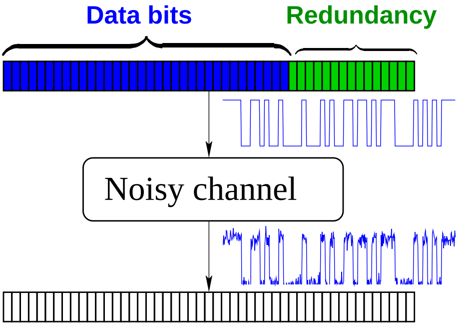
Shannon: Fix $R < C$

$P_e \searrow 0$ as $n \to \infty$

Question:

For what n will $P_e < 10^{-3}$?

Channel coding: Gaussian approximation

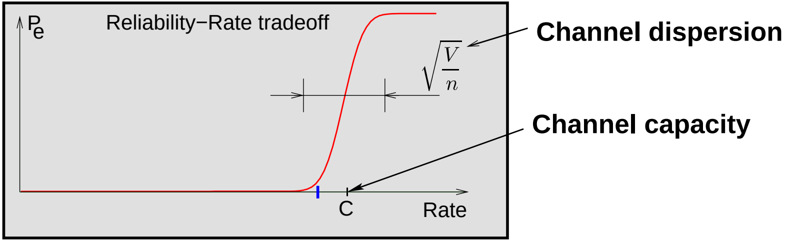


Shannon: Fix $R < C$

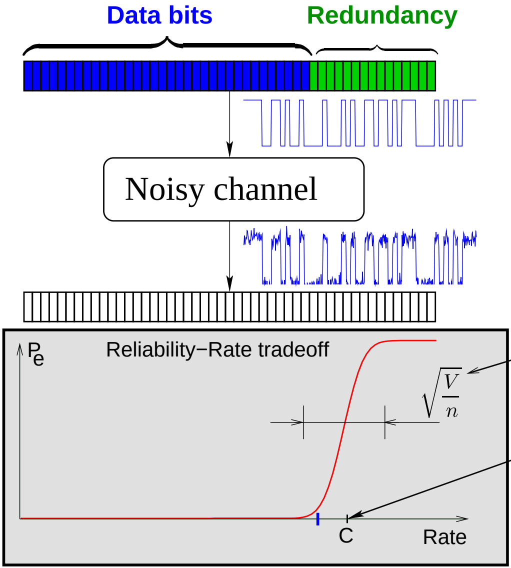
$P_e \searrow 0$ as $n \to \infty$

Question:

For what n will $P_e < 10^{-3}$?



Channel coding: Gaussian approximation



Shannon: Fix $R < C$

$P_e \searrow 0$ as $n \to \infty$

Question:

For what n will $P_e < 10^{-3}$?

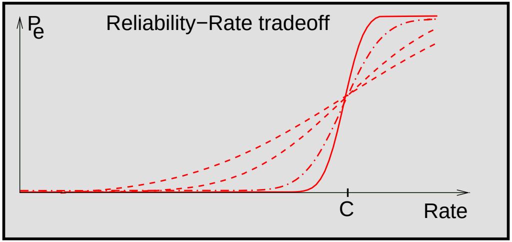
Answer:

$$n \gtrsim \mathrm{const} \cdot \frac{V}{C^2}$$

Channel dispersion

Channel capacity

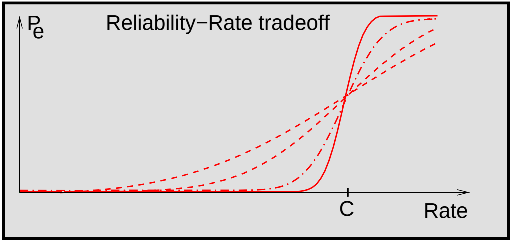
How to describe evolution of the boundary?



Classical results:

- **Vertical asymptotics:** fixed rate, reliability function
  Elias, Dobrushin, Fano, Shannon-Gallager-Berlekamp
- **Horizontal asymptotics:** fixed ϵ, strong converse, $\sqrt{n}$ terms
  Wolfowitz, Weiss, Dobrushin, Strassen, Kemperman

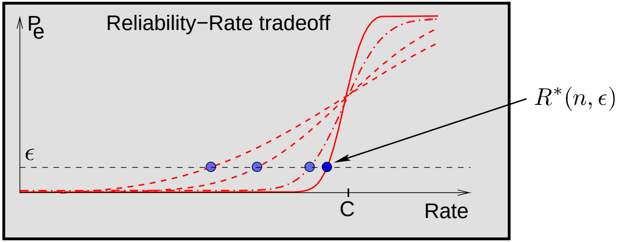
How to describe evolution of the boundary?



XXI century:

- Tight non-asymptotic bounds
- Remarkable precision of normal approximation
- Extended results on *horizontal* asymptotics
  AWGN, $O(\log n)$, cost constraints, feedback, *etc.*

Finite blocklength fundamental limit



Definition

$$R^*(n, \epsilon) = \max \left\{ \frac{1}{n} \log M : \exists (n, M, \epsilon)\text{-code} \right\}$$

(max. achievable rate for blocklength n and prob. of error ϵ)

Note: Exact value unknown (search is doubly exponential in n).

Minimal delay and error-exponents

Fix $R < C$. What is the smallest blocklength n^* needed to achieve

$$R^*(n, \epsilon) \geq R \quad ?$$

Minimal delay and error-exponents

Fix $R < C$. What is the smallest blocklength n^* needed to achieve

$$R^*(n, \epsilon) \geq R \quad ?$$

Classical answer: Approximation via reliability function [Shannon-Gallager-Berlekamp'67]:

$$n^* \approx \frac{1}{E(R)} \log \frac{1}{\epsilon}$$

E.g., take $BSC(0.11)$ and $R = 0.9C$, prob. of error $\epsilon = 10^{-3}$:

$$n^* \approx 5000 \qquad \text{(channel uses)}$$

Difficulty: How to verify accuracy of this estimate?

Bounds

- Bounds are implicit in Shannon's theorem

$$\lim_{n\to\infty} R^*(n,\epsilon) = C \quad \Longleftrightarrow \begin{cases} R^*(n,\epsilon) \le C + o(1)\,, \\ R^*(n,\epsilon) \ge C + o(1)\,. \end{cases}$$

(Feinstein'54, Shannon'57, Wolfowitz'57, Fano)

- Reliability function: even better bounds (Elias'55, Shannon'59, Gallager'65, SGB'67)
- Problems: derived for asymptotics (need "de-asymptotization")
  unexpected sensitivity:

$$\epsilon \le e^{-nE_r(R)} \qquad \text{[Gallager'65]}$$

$$\epsilon \le e^{-nE_r(R-o(1))+O(\log n)} \qquad \text{[Csiszár-Körner'81]}$$

Bounds

- Bounds are implicit in Shannon's theorem

$$\lim_{n\to\infty} R^*(n,\epsilon) = C \quad \Longleftrightarrow \begin{cases} R^*(n,\epsilon) \le C + o(1)\,, \\ R^*(n,\epsilon) \ge C + o(1)\,. \end{cases}$$

  (Feinstein'54, Shannon'57, Wolfowitz'57, Fano)

- Reliability function: even better bounds
  (Elias'55, Shannon'59, Gallager'65, SGB'67)
- Problems: derived for asymptotics (need "de-asymptotization")
  unexpected sensitivity:

$$\epsilon \le e^{-nE_r(R)} \qquad \text{[Gallager'65]}$$

$$\epsilon \le e^{-nE_r(R-o(1))+O(\log n)} \qquad \text{[Csiszár-Körner'81]}$$

For BSC($n = 10^3, 0.11$): $o(1) \approx 0.1$, $e^{O(\log n)} \approx 10^{24}$ (!)

Bounds

- Bounds are implicit in Shannon's theorem

$$\lim_{n\to\infty} R^*(n,\epsilon) = C \quad \iff \begin{cases} R^*(n,\epsilon) \le C + o(1)\,, \\ R^*(n,\epsilon) \ge C + o(1)\,. \end{cases}$$

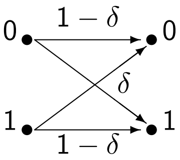
(Feinstein'54, Shannon'57, Wolfowitz'57, Fano)

- Reliability function: even better bounds
  (Elias'55, Shannon'59, Gallager'65, SGB'67)
- Problems: derived for asymptotics (need "de-asymptotization")
  unexpected sensitivity:

Strassen'62: Take $n > \frac{19600}{\epsilon^{16}}$ . . . (!)

- Solution: Derive bounds from scratch.

New achievability bound



BSC

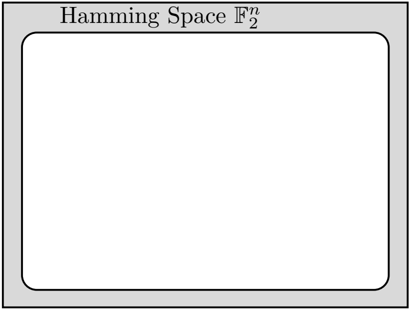
Theorem (RCU)

For a BSC(δ) there exists a code with rate R, blocklength n and

$$\epsilon \le \sum_{t=0}^{n} \binom{n}{t} \delta^t (1-\delta)^{n-t} \min\left\{1, \ \sum_{k=0}^{t} \binom{n}{k} 2^{-n+nR}\right\} .$$

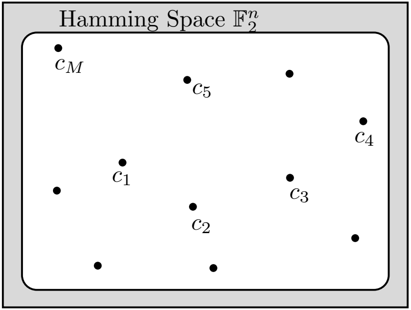
Proof of RCU bound for the BSC

- Input space: $A = \{0,1\}^n$



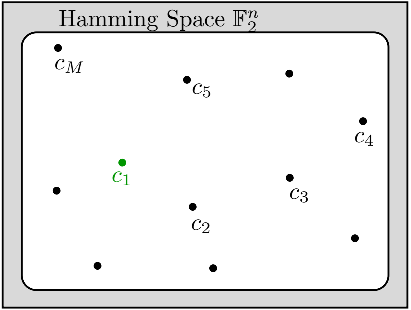
Proof of RCU bound for the BSC

- Input space: $A = \{0,1\}^n$
- Let $c_1, \ldots, c_M \sim Bern(\frac{1}{2})^n$ (random codebook)



Proof of RCU bound for the BSC

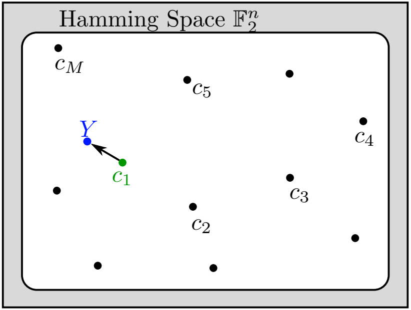
- Input space: $\mathsf{A} = \{0, 1\}^n$
- Let $c_1, \ldots, c_M \sim Bern(\frac{1}{2})^n$ (random codebook)
- Transmit c_1



Proof of RCU bound for the BSC

- Input space: $A = \{0, 1\}^n$
- Let $c_1, \ldots, c_M \sim Bern(\frac{1}{2})^n$ (random codebook)
- Transmit c_1
- Noise displaces $c_1 \rightarrow Y$

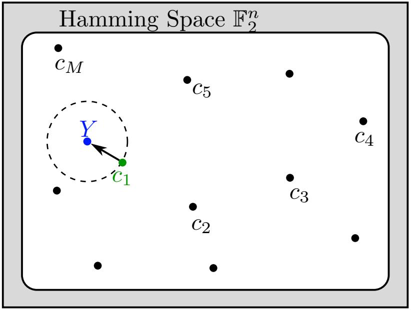
$Y = c_1 + Z, \quad Z \sim Bern(\delta)^n$



Proof of RCU bound for the BSC

- Input space: $A = \{0,1\}^n$
- Let $c_1, \ldots, c_M \sim Bern(\frac{1}{2})^n$ (random codebook)
- Transmit c_1
- Noise displaces $c_1 \to Y$

$Y = c_1 + Z, \quad Z \sim Bern(\delta)^n$



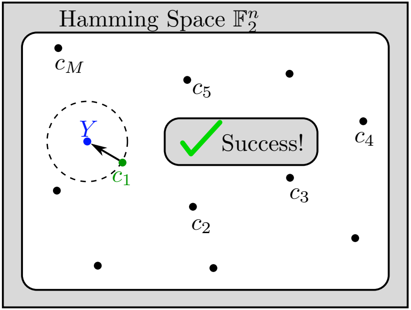
- Decoder: find closest codeword to Y

Proof of RCU bound for the BSC

- Input space: $A = \{0,1\}^n$
- Let $c_1, \ldots, c_M \sim Bern(\frac{1}{2})^n$ (random codebook)
- Transmit c_1
- Noise displaces $c_1 \to Y$

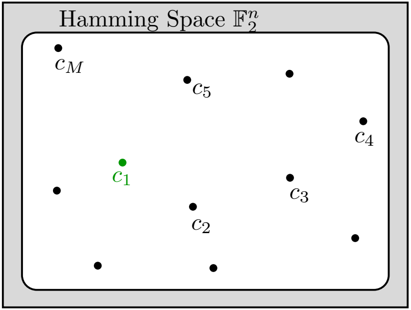
$Y = c_1 + Z, \quad Z \sim Bern(\delta)^n$

- Decoder: find closest codeword to Y



Proof of RCU bound for the BSC

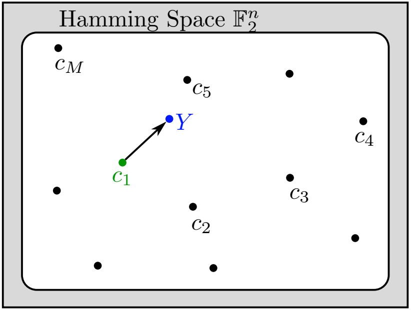
- Input space: $\mathsf{A} = \{0,1\}^n$
- Let $c_1, \ldots, c_M \sim Bern(\frac{1}{2})^n$ (random codebook)
- Transmit c_1



Proof of RCU bound for the BSC

- Input space: $A = \{0,1\}^n$
- Let $c_1, \ldots, c_M \sim Bern(\frac{1}{2})^n$ (random codebook)
- Transmit c_1
- Noise displaces $c_1 \to Y$

$$Y = c_1 + Z, \quad Z \sim Bern(\delta)^n$$

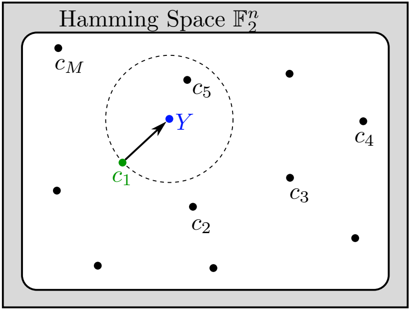


Proof of RCU bound for the BSC

- Input space: $\mathsf{A} = \{0,1\}^n$
- Let $c_1, \ldots, c_M \sim Bern(\frac{1}{2})^n$ (random codebook)
- Transmit c_1
- Noise displaces $c_1 \to Y$

$Y = c_1 + Z, \quad Z \sim Bern(\delta)^n$

- Decoder: find closest codeword to Y

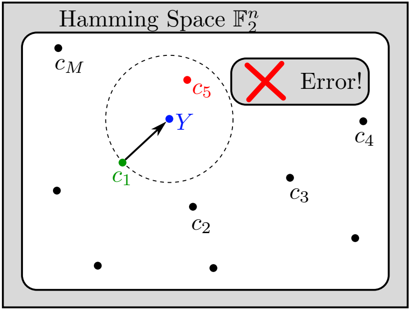


Proof of RCU bound for the BSC

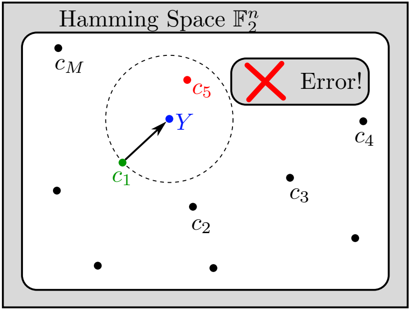
- Input space: $\mathsf{A} = \{0,1\}^n$
- Let $c_1, \ldots, c_M \sim Bern(\frac{1}{2})^n$ (random codebook)
- Transmit c_1
- Noise displaces $c_1 \to Y$

$Y = c_1 + Z, \quad Z \sim Bern(\delta)^n$

- Decoder: find closest codeword to Y



Proof of RCU bound for the BSC



- Input space: $A = \{0,1\}^n$
- Let $c_1, \ldots, c_M \sim Bern(\frac{1}{2})^n$ (random codebook)
- Transmit c_1
- Noise displaces $c_1 \to Y$

$Y = c_1 + Z, \quad Z \sim Bern(\delta)^n$

- Decoder: find closest codeword to Y
- Probability of error analysis:

$$\begin{aligned}\mathbb{P}[\text{error}|Y, \mathrm{wt}(Z) = t] &= \mathbb{P}[\exists j > 1 : c_j \in \mathrm{Ball}(Y,t)] \\ &\le \sum_{j=2}^{M} \mathbb{P}[c_j \in \mathrm{Ball}(Y,t)] \\ &\le 2^{nR} \sum_{k=0}^{t} \binom{n}{k} 2^{-n}\end{aligned}$$

. . . cont'd . . .

- Conditional probability of error:

$$\mathbb{P}[\text{error}|Y, \mathrm{wt}(Z) = t] \le \sum_{k=0}^{t} \binom{n}{k} 2^{-n+nR}$$

- Key observation: For large noise t RHS is > 1. Tighten:

$$\mathbb{P}[\text{error}|Y, \mathrm{wt}(Z) = t] \le \min\left\{1, \sum_{k=0}^{t} \binom{n}{k} 2^{-n+nR}\right\} \qquad (*)$$

- Average $\mathrm{wt}(Z) \sim \mathit{Binomial}(n, \delta) \Longrightarrow$ **Q.E.D.**

Note: Step $(*)$ tightens Gallager's ρ-trick:

$$\mathbb{P}\left[\bigcup_j A_j\right] \le \left(\sum_j \mathbb{P}[A_j]\right)^{\rho}$$

Sphere-packing converse (BSC variation)

Theorem (Elias'55)

For any (n, M, ϵ) code over the BSC(δ):

$$\epsilon \geq f\left(\frac{2^n}{M}\right),$$

where $f(\cdot)$ is a piecewise-linear decreasing convex function:

$$f\left(\sum_{j=0}^{t}\binom{n}{j}\right) = \sum_{j=t+1}^{n}\binom{n}{j}\delta^j(1-\delta)^{n-j} \qquad t = 0, \ldots, n$$

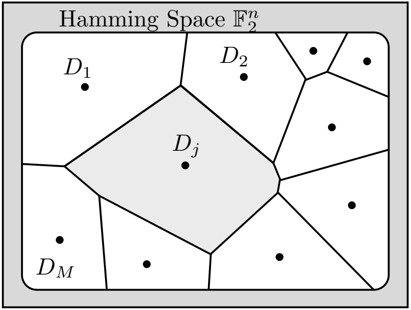
Note: Convexity of f follows from general properties of β_α (below)

Sphere-packing converse (BSC variation)

Proof:

- Denote decoding regions D_j: $\coprod D_j = \{0,1\}^n$
- Probability of error is:

$$\epsilon = \frac{1}{M} \sum_j \mathbb{P}[c_j + Z \notin D_j]$$

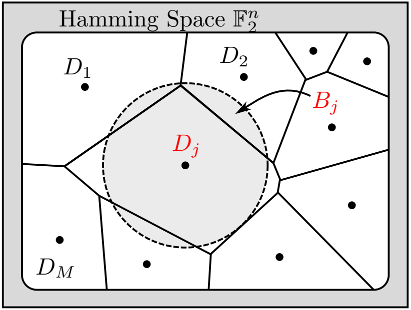


Sphere-packing converse (BSC variation)

Proof:

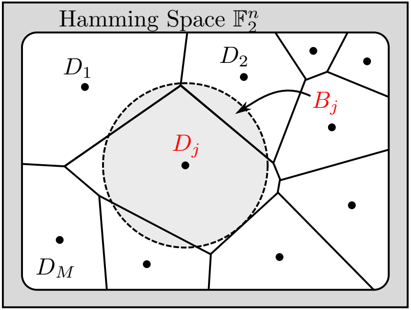
- Denote decoding regions D_j: $\coprod D_j = \{0,1\}^n$
- Probability of error is:

$$\epsilon = \frac{1}{M}\sum_j \mathbb{P}[c_j + Z \notin D_j] \\ \geq \frac{1}{M}\sum_j \mathbb{P}[Z \notin B_j]$$



- $B_j =$ ball centered at 0 s.t. $\mathrm{Vol}(B_j) = \mathrm{Vol}(D_j)$

Sphere-packing converse (BSC variation)



Proof:

- Denote decoding regions D_j: $\coprod D_j = \{0,1\}^n$
- Probability of error is:

$$\epsilon = \frac{1}{M}\sum_j \mathbb{P}[c_j + Z \notin D_j]$$
$$\geq \frac{1}{M}\sum_j \mathbb{P}[Z \notin B_j]$$

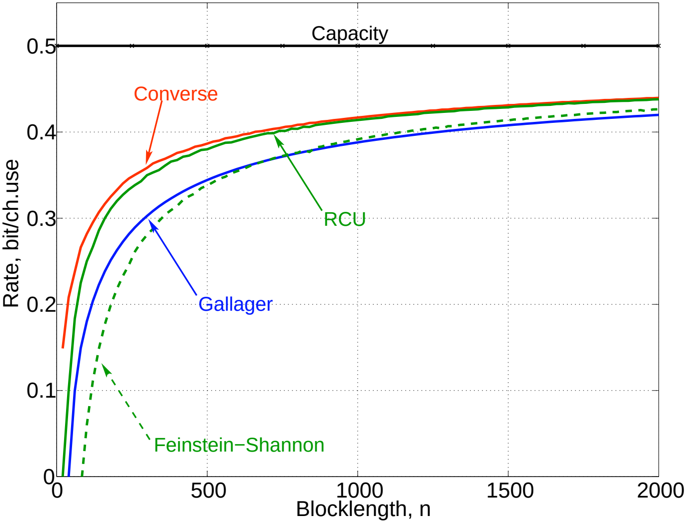
- $B_j =$ ball centered at 0 s.t. $\mathrm{Vol}(B_j) = \mathrm{Vol}(D_j)$
- Simple calculation:

$$\mathbb{P}[Z \notin B_j] = f(\mathrm{Vol}(B_j))$$

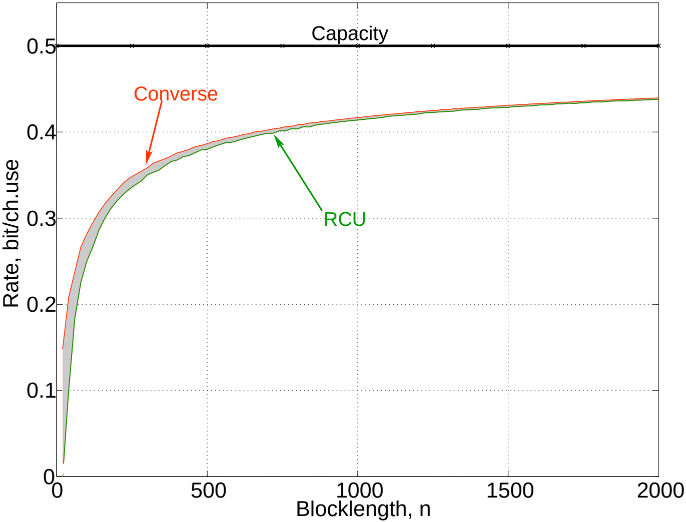
- f – convex, apply Jensen:

$$\epsilon \geq f\left(\frac{1}{M}\sum_{j=1}^{M}\mathrm{Vol}(D_j)\right) = f\left(\frac{2^n}{M}\right)$$

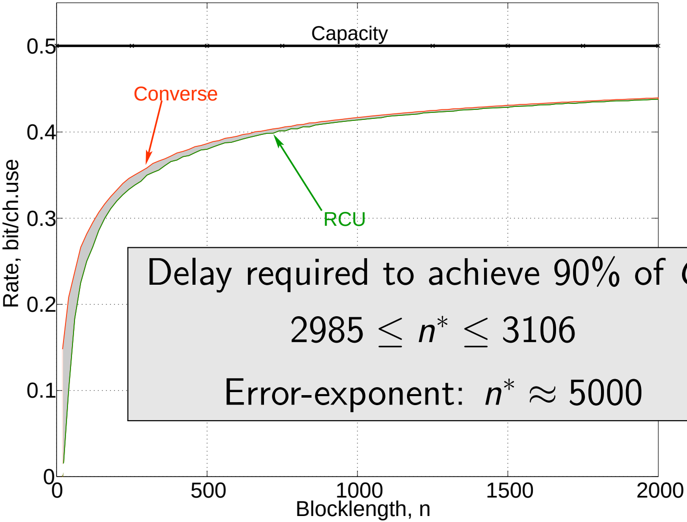
Bounds: example $BSC(0.11)$, $\epsilon = 10^{-3}$



Bounds: example $BSC(0.11)$, $\epsilon = 10^{-3}$



Bounds: example $BSC(0.11)$, $\epsilon = 10^{-3}$



Delay required to achieve 90% of C:

$$2985 \leq n^* \leq 3106$$

Error-exponent: $n^* \approx 5000$

Normal approximation

Theorem

For the BSC(δ) and $0 < \epsilon < 1$,

$$R^*(n,\epsilon) = C - \sqrt{\frac{V}{n}}Q^{-1}(\epsilon) + \frac{1}{2}\frac{\log n}{n} + O\left(\frac{1}{n}\right)$$

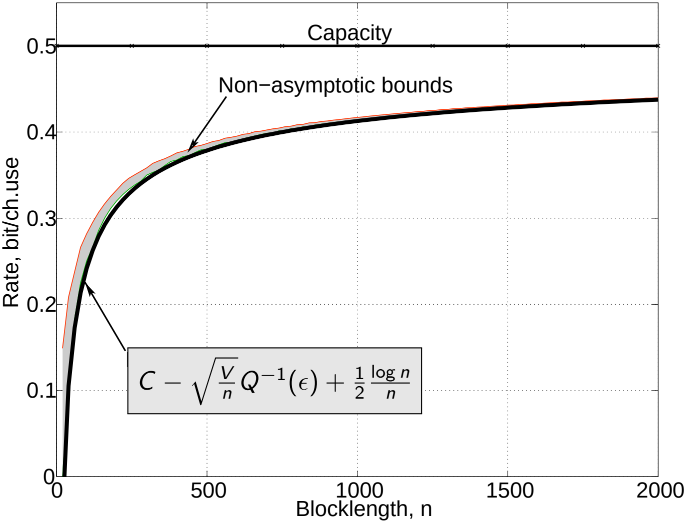
where

$$\begin{aligned} C(\delta) &= \log 2 + \delta \log \delta + (1-\delta)\log(1-\delta) \\ V &= \delta(1-\delta)\log^2 \frac{1-\delta}{\delta} \\ Q(x) &= \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \end{aligned}$$

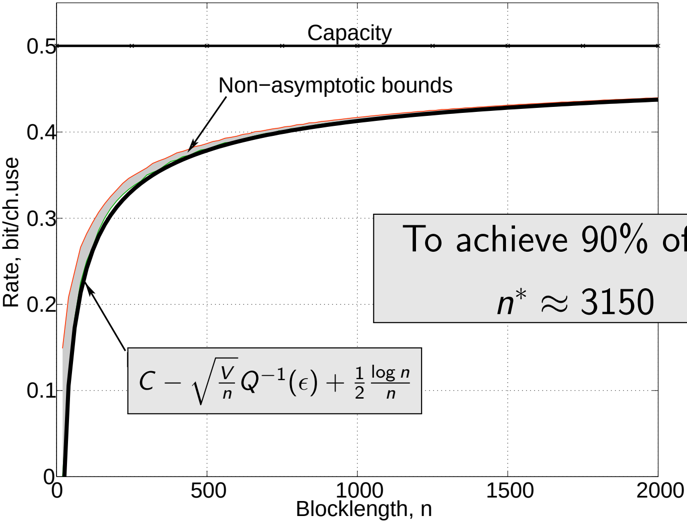
Proof: Bounds + Stirling's formula

Note: Now we see the explicit dependence on ϵ!

Normal approximation: $BSC(0.11)$; $\epsilon = 10^{-3}$



Normal approximation: $BSC(0.11)$; $\epsilon = 10^{-3}$



To achieve 90% of C:

$n^* \approx 3150$

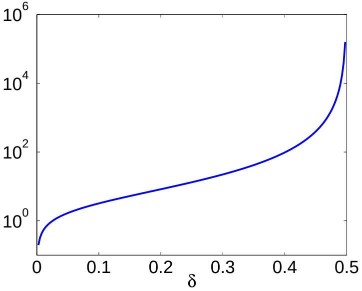
Dispersion and minimal required delay

Delay needed to achieve $R = \eta C$:

$$n^* \gtrsim \left(\frac{Q^{-1}(\epsilon)}{1-\eta} \right)^2 \cdot \frac{V}{C^2}$$

Note: $\frac{V}{C^2}$ is "coding horizon".

Behavior of $\frac{V}{C^2}$ (BSC)



Less noise More noise

BSC summary

Delay required to achieve 90 % of capacity:

- Error-exponents:

$$n^* \approx 5000$$

- True value:

$$2985 \le n^* \le 3106$$

- Channel dispersion:

$$n^* \approx 3150$$

Converse Bounds

Notation

- Take a random transformation $\mathsf{A} \xrightarrow{P_{Y|X}} \mathsf{B}$
  (think $\mathsf{A} = \mathcal{A}^n$, $\mathsf{B} = \mathcal{B}^n$, $P_{Y|X} = P_{Y^n|X^n}$)
- Input distribution P_X induces $P_Y = P_{Y|X} \circ P_X$
  $P_{XY} = P_X P_{Y|X}$
- Fix code:

$$W \xrightarrow{encoder} X \to Y \xrightarrow{decoder} \hat{W}$$

  $W \sim Unif[M]$ and $M = \#$ of codewords
  Input distribution P_X associated to a code:

$$P_X[\cdot] \triangleq \frac{\#\text{ of codewords } \in (\cdot)}{M}.$$

- Goal: Upper bounds on $\log M$ in terms of $\epsilon \triangleq \mathbb{P}[error]$
  As by-product: $R^*(n,\epsilon) \lesssim C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon)$

Fano's inequality

Theorem (Fano)

For any code

$$W \xrightarrow{\text{encoder}} X \xrightarrow{P_{Y|X}} Y \xrightarrow{\text{decoder}} \hat{W}$$

with $W \sim Unif\{1, \ldots, M\}$:

$$\log M \leq \frac{\sup_{P_X} I(X;Y) + h(\epsilon)}{1-\epsilon}, \quad \epsilon = \mathbb{P}[W \neq \hat{W}]$$

Implies *weak converse*:

$$R^*(n, \epsilon) \leq \frac{C}{1-\epsilon} + o(1).$$

Proof: ϵ-small $\implies$ $H(W|\hat{W})$-small $\implies$ $I(X;Y) \approx H(W) = \log M$

A (very long) proof of Fano via *channel substitution*

Consider two distributions on $(W, X, Y, \hat{W})$:

$$\mathbb{P}: \quad P_{WXY\hat{W}} = P_W \times P_{X|W} \times P_{Y|X} \times P_{\hat{W}|Y}$$

DAG: $W \to X \to Y \to \hat{W}$

$$\mathbb{Q}: \quad Q_{WXY\hat{W}} = P_W \times P_{X|W} \times Q_Y \times P_{\hat{W}|Y}$$

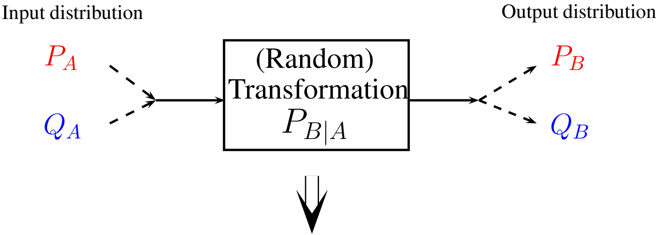
DAG: $W \to X \quad Y \to \hat{W}$

Under $\mathbb{Q}$ the channel is useless:

$$\mathbb{Q}[W = \hat{W}] = \sum_{m=1}^{M} P_W(m) Q_{\hat{W}}(m) = \frac{1}{M} \sum_{m=1}^{M} Q_{\hat{W}}(m) = \frac{1}{M}$$

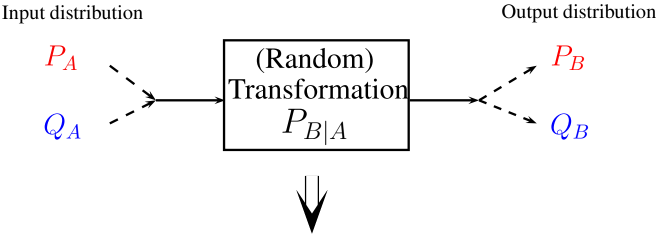
Next step: data-processing for relative entropy $D(\cdot||\cdot)$

Data-processing for $D(\cdot||\cdot)$



$$D(P_A \| Q_A) \qquad \geq \qquad D(P_B \| Q_B)$$

Data-processing for $D(\cdot||\cdot)$



$$D(P_A \| Q_A) \qquad \geq \qquad D(P_B \| Q_B)$$

Apply to transform: $(W, X, Y, \hat{W}) \mapsto 1\{W \neq \hat{W}\}$:

$$\begin{aligned} D(P_{WXY\hat{W}} \| Q_{WXY\hat{W}}) &\geq d(\, \mathbb{P}[W = \hat{W}] \,\|\, \mathbb{Q}[W = \hat{W}]\,) \\ &= d(1 - \epsilon || \tfrac{1}{M}) \end{aligned}$$

where $d(x||y) = x \log \frac{x}{y} + (1 - x) \log \frac{1-x}{1-y}$.

A proof of Fano via *channel substitution*

So far:

$$D(P_{WXY\hat{W}} \| Q_{WXY\hat{W}}) \geq d(1-\epsilon \| \tfrac{1}{M})$$

Lower-bound RHS:

$$d(1-\epsilon \| \tfrac{1}{M}) \geq (1-\epsilon)\log M - h(\epsilon)$$

Analyze LHS:

$$\begin{aligned} D(P_{WXY\hat{W}} \| Q_{WXY\hat{W}}) &= D(P_{XY} \| Q_{XY}) \\ &= D(P_X P_{Y|X} \| P_X Q_Y) \\ &= D(P_{Y|X} \| Q_Y | P_X) \end{aligned}$$

(Recall: $D(P_{Y|X} \| Q_Y | P_X) = \mathbb{E}_{x \sim P_X}[D(P_{Y|X=x} \| Q_Y)]$)

A proof of Fano via *channel substitution*: last step

Putting it all together:

$$\boxed{(1-\epsilon)\log M \le D(P_{Y|X}||Q_Y|P_X) + h(\epsilon) \quad \forall Q_Y \quad \forall \text{code}}$$

Two methods:

1. Compute $\sup_{P_X} \inf_{Q_Y}$ and recall

$$\inf_{Q_Y} D(P_{Y|X}\|Q_Y|P_X) = I(X;Y)$$

2. Take $Q_Y = P_Y^* =$ the caod (capacity achieving output dist.) and recall

$$D(P_{Y|X}\|P_Y^*|P_X) \le \sup_{P_X} I(X;Y) \qquad \forall P_X$$

Conclude:

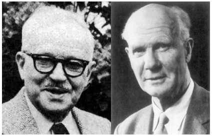
$$(1-\epsilon)\log M \le \sup_{P_X} I(X;Y) + h(\epsilon)$$

Important: Second method is particularly useful for FBL!

Tightening: from $D(\cdot||\cdot)$ to $\beta_\alpha(\cdot,\cdot)$

Question: How about replacing $D(\cdot||\cdot)$ with other divergences?

Answer:

	$D(\cdot\|\|\cdot)$	relative entropy (KL divergence)	weak converse
	$D_\lambda(\cdot\|\|\cdot)$	Rényi divergence	strong converse
	$\beta_\alpha(\cdot,\cdot)$	Neyman-Pearson ROC curve	FBL bounds

Next: What is β_α?

Neyman-Pearson's β_α

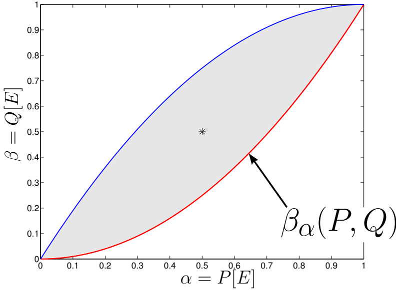
Definition

For every pair of measures P, Q

$$\beta_\alpha(P,Q) \triangleq \inf_{E:P[E]\geq\alpha} Q[E]\,.$$

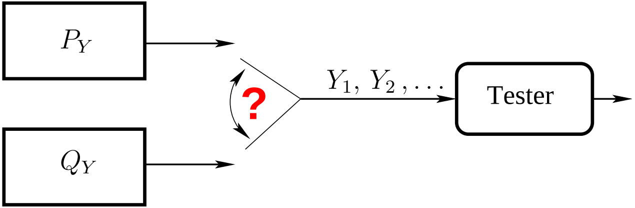
iterate over
all "sets" E
plot pairs $(P[E], Q[E])$

$\Longrightarrow$



Important: Like relative entropy β_α satisfies data-processing.

β_α = binary hypothesis testing



Two types of errors:

$$\begin{aligned} \mathbb{P}[\text{Tester says "}Q_Y\text{"}] &\le \epsilon \\ \mathbb{Q}[\text{Tester says "}P_Y\text{"}] &\to \min \end{aligned}$$

Hence: Solve binary HT $\iff$ compute $\beta_\alpha(P_Y^n, Q_Y^n)$

Stein's Lemma: For many i.i.d. observations

$$\log \beta_{1-\epsilon}(P_Y^n, Q_Y^n) = -nD(P_Y || Q_Y) + o(n)\,.$$

But in fact $\log \beta_\alpha(P_Y^n, Q_Y^n)$ can also be computed exactly!

How to compute β_α?

Theorem (Neyman-Pearson)

β_α is given parametrically by $-\infty \le \gamma \le +\infty$:

$$\mathbb{P}\left[\log \frac{P(X)}{Q(X)} \ge \gamma\right] = \alpha$$

$$\mathbb{Q}\left[\log \frac{P(X)}{Q(X)} \ge \gamma\right] = \beta_\alpha(P, Q)$$

For product measures $\log \frac{P^n(X)}{Q^n(X)} =$ sum of i.i.d. $\Longrightarrow$ from CLT:

$$\log \beta_\alpha(P^n, Q^n) = -nD(P||Q) + \sqrt{nV(P||Q)}Q^{-1}(\alpha) + o(\sqrt{n}),$$

where

$$V(P||Q) = \mathrm{Var}_P\left[\log \frac{P(X)}{Q(X)}\right]$$

Back to proving converse

Recall two measures:

$$\mathbb{P}: \quad P_{WXY\hat{W}} = P_W \times P_{X|W} \times P_{Y|X} \times P_{\hat{W}|Y}$$

DAG: $W \to X \to Y \to \hat{W}$

$$\mathbb{P}[W = \hat{W}] = 1-\epsilon$$

$$\mathbb{Q}: \quad Q_{WXY\hat{W}} = P_W \times P_{X|W} \times Q_Y \times P_{\hat{W}|Y}$$

DAG: $W \to X \quad Y \to \hat{W}$

$$\mathbb{Q}[W = \hat{W}] = \frac{1}{M}$$

Then by definition of β_α:

$$\beta_{1-\epsilon}(P_{WXY\hat{W}}, Q_{WXY\hat{W}}) \le \frac{1}{M}$$

But $\log \frac{P_{WXY\hat{W}}}{Q_{WXY\hat{W}}} = \log \frac{P_X P_{Y|X}}{P_X Q_Y} \quad \Longrightarrow$

$$\log M \le -\log \beta_{1-\epsilon}(P_X P_{Y|X}, P_X Q_Y) \qquad \forall Q_Y \quad \forall \text{code}$$

Meta-converse: minimax version

Theorem

Every (M, ϵ)-code for channel $P_{Y|X}$ satisfies

$$\log M \le -\log\left\{\inf_{P_X}\sup_{Q_Y}\beta_{1-\epsilon}(P_X P_{Y|X}, P_X Q_Y)\right\}.$$

Meta-converse: minimax version

Theorem

Every (M, ϵ)-code for channel $P_{Y|X}$ satisfies

$$\log M \le -\log\left\{\inf_{P_X}\sup_{Q_Y}\beta_{1-\epsilon}(P_X P_{Y|X}, P_X Q_Y)\right\}.$$

- Finding good Q_Y for every P_X is not needed:

$$\inf_{P_X}\sup_{Q_Y}\beta_{1-\epsilon}(P_X P_{Y|X}, P_X Q_Y) = \sup_{Q_Y}\inf_{P_X}\beta_{1-\epsilon}(P_X P_{Y|X}, P_X Q_Y) \quad (*)$$

- Saddle-point property of β_α is similar to $D(\cdot||\cdot)$:

$$\inf_{P_X}\sup_{Q_Y} D(P_X P_{Y|X}\|P_X Q_Y) = \sup_{Q_Y}\inf_{P_X} D(P_X P_{Y|X}\|P_X Q_Y) = C$$

Meta-converse: minimax version

Theorem

Every (M, ϵ)-code for channel $P_{Y|X}$ satisfies

$$\log M \leq -\log \left\{ \inf_{P_X} \sup_{Q_Y} \beta_{1-\epsilon}(P_X P_{Y|X}, P_X Q_Y) \right\} .$$

Bound is tight in two senses:

- There exist *non-signalling assisted* (NSA) codes attaining the upper-bound. [Matthews, Trans. IT'2012]
- **ISIT'2013:** For any (M, ϵ)-code with ML decoder

$$\log M = -\log \left\{ \sup_{Q_Y} \beta_{1-\epsilon}(P_X P_{Y|X}, P_X Q_Y) \right\}$$

Vazquez-Vilar *et al* [WeB4]

Meta-converse: minimax version

Theorem

Every (M, ϵ)*-code for channel* $P_{Y|X}$ *satisfies*

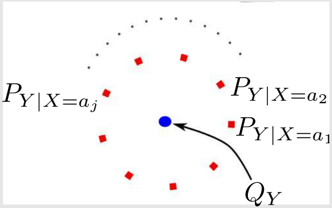
$$\log M \le -\log \left\{ \inf_{P_X} \sup_{Q_Y} \beta_{1-\epsilon}(P_X P_{Y|X}, P_X Q_Y) \right\} .$$

In practice: evaluate with a luckily guessed (suboptimal) Q_Y.

How to guess good Q_Y?

- Try caod P_Y^*
- Analyze channel symmetries
- Use geometric intuition. $\longrightarrow$
  good $Q_Y \approx$ "center" of $P_{Y|X}$
- *Exercise:* Redo BSC.

Space of distributions on $\mathcal{Y}$



Example: Converse for AWGN

The AWGN Channel

$$
\begin{array}{ccccc}
 & & Z \sim \mathcal{N}(0, \sigma^2) & & \\
 & & \downarrow & & \\
X & \longrightarrow & \bigoplus & \longrightarrow & Y
\end{array}
$$

Codewords $X^n \in \mathbb{R}^n$ satisfy power-constraint:

$$\sum_{j=1}^{n} |X_j|^2 \leq nP$$

Goal: Upper-bound # of codewords decodable with $P_e \leq \epsilon$.

Example: Converse for AWGN

- Given $\{c_1, \ldots, c_M\} \in \mathbb{R}^n$ with $\mathbb{P}[W \neq \hat{W}] \leq \epsilon$ on $AWGN(1)$.
- Yaglom-map trick: replacing $n \to n+1$ equalize powers:

$$\|c_j\|^2 = nP \qquad \forall j \in \{1, \ldots, M\}$$

Example: Converse for AWGN

- Given $\{c_1, \ldots, c_M\} \in \mathbb{R}^n$ with $\mathbb{P}[W \neq \hat{W}] \leq \epsilon$ on $AWGN(1)$.
- Yaglom-map trick: replacing $n \to n+1$ equalize powers:

$$\|c_j\|^2 = nP \qquad \forall j \in \{1, \ldots, M\}$$

- Take $Q_{Y^n} = \mathcal{N}(0, 1+P)^n$ (the caod!)
- Optimal test "$P_{X^n Y^n}$ vs. $P_{X^n} Q_{Y^n}$" (Neyman-Pearson):

$$\log \frac{P_{Y^n|X^n}}{Q_{Y^n}} = nC + \frac{\log e}{2} \cdot \left(\frac{\|Y^n\|^2}{1+P} - \|Y^n - X^n\|^2 \right)$$

where $C = \frac{1}{2}\log(1+P)$.

- Under $\mathbb{P}$: $Y^n = X^n + \mathcal{N}(0, \mathbf{I}_n)$
  $\Longrightarrow$ distribution of LLR (CLT approx.)

$$\approx nC + \sqrt{nV} Z \,, \qquad Z \sim \mathcal{N}(0,1)$$

Simple algebra: $V = \frac{\log^2 e}{2} \left(1 - \frac{1}{(1+P)^2} \right)$

...cont'd ...

- Under $\mathbb{P}$: distribution of LLR (CLT approx.)

$$\approx nC + \sqrt{nV}Z\,, \qquad Z \sim \mathcal{N}(0,1)$$

- Take $\gamma = nC - \sqrt{nV}Q^{-1}(\epsilon) \implies$

$$\mathbb{P}\left[\log \frac{d\mathbb{P}}{d\mathbb{Q}} \geq \gamma\right] \approx 1 - \epsilon\,.$$

- Under $\mathbb{Q}$: standard change-of-measure shows

$$\mathbb{Q}\left[\log \frac{d\mathbb{P}}{d\mathbb{Q}} \geq \gamma\right] \approx \exp\{-\gamma\}\,.$$

...cont'd ...

- Under $\mathbb{P}$: distribution of LLR (CLT approx.)

$$\approx nC + \sqrt{nV}Z\,, \qquad Z \sim \mathcal{N}(0,1)$$

- Take $\gamma = nC - \sqrt{nV}Q^{-1}(\epsilon) \implies$

$$\mathbb{P}\left[\log\frac{d\mathbb{P}}{d\mathbb{Q}} \geq \gamma\right] \approx 1 - \epsilon\,.$$

- Under $\mathbb{Q}$: standard change-of-measure shows

$$\mathbb{Q}\left[\log\frac{d\mathbb{P}}{d\mathbb{Q}} \geq \gamma\right] \approx \exp\{-\gamma\}\,.$$

- By Neyman-Pearson

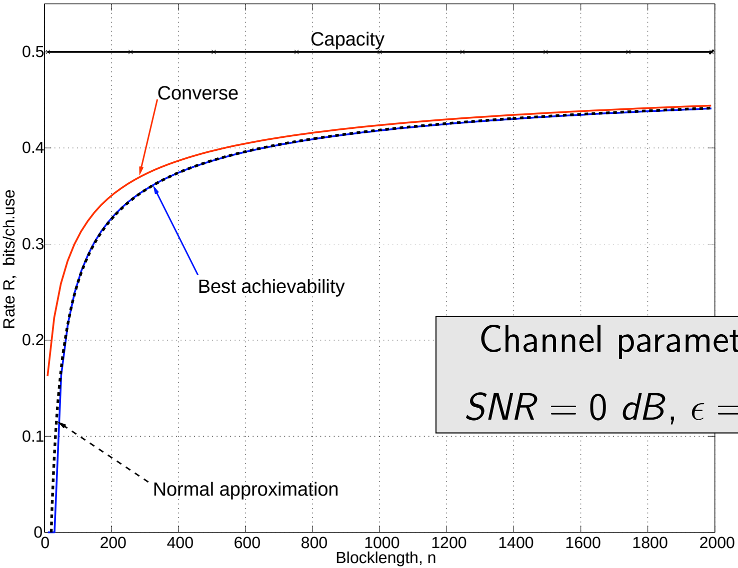
$$\log\beta_{1-\epsilon}(P_{Y^n|X^n=c}, Q_{Y^n}) \approx -nC + \sqrt{nV}Q^{-1}(\epsilon)$$

- Punchline: $\forall(n, M, \epsilon)$-code

$$\log M \lesssim nC - \sqrt{nV}Q^{-1}(\epsilon)$$

N.B.! RHS can be exactly expressed via *non-central* χ^2 *dist.*
... and computed in MATLAB (w/o any CLT approx).

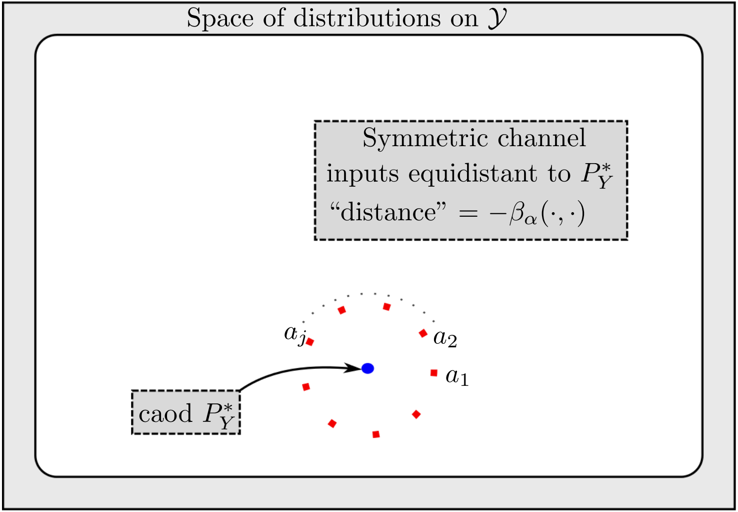
AWGN: Converse from $\beta_\alpha(P, Q)$ with $Q_Y = \mathcal{N}(0,1)^n$



Channel parameters:

$SNR = 0\ dB$, $\epsilon = 10^{-3}$

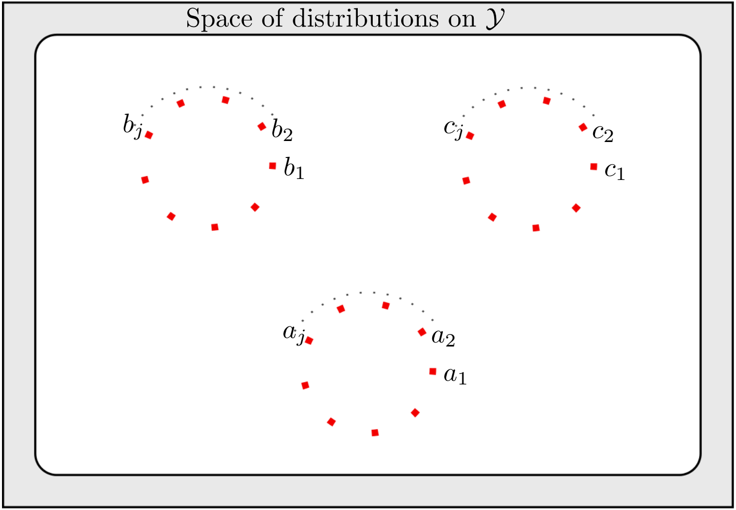
From one Q_Y to many



points
$\Updownarrow$
ch. inputs
(a_j, b_j, c_j)

Symmetric channel: choice of Q_Y is clear

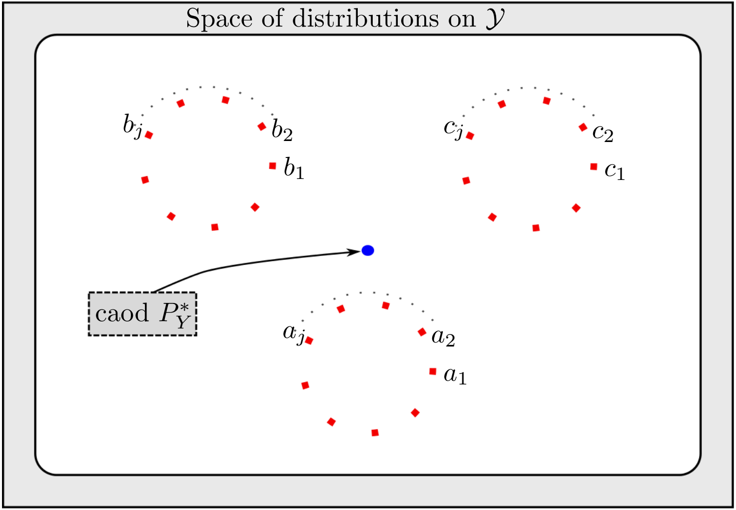
From one Q_Y to many



points
$\Updownarrow$
ch. inputs
(a_j, b_j, c_j)

General channels: Inputs cluster (by composition, power-allocation, ...)
(Clusters $\iff$ orbits of channel symmetry gp.)

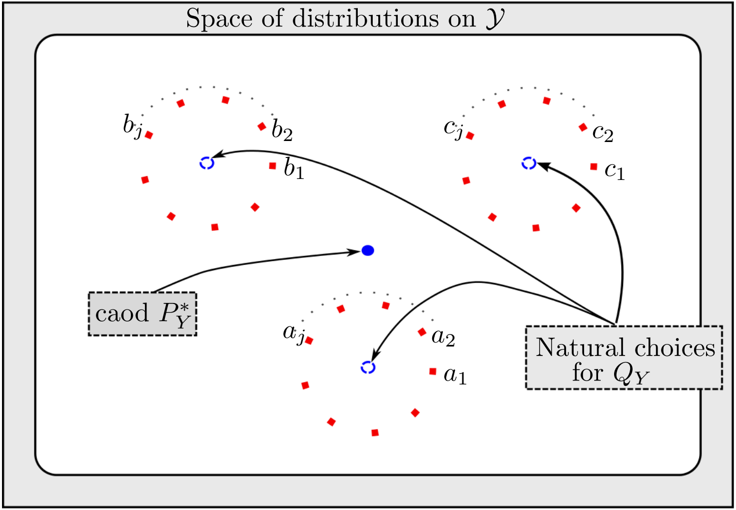
From one Q_Y to many



points
$\Updownarrow$
ch. inputs
(a_j, b_j, c_j)

General channels: Caod is no longer equidistant to all inputs (read: analysis horrible!)

From one Q_Y to many



points $\Updownarrow$ ch. inputs (a_j, b_j, c_j)

Solution: Take Q_Y different for each cluster!

I.e. think of $Q_{Y|X}$

General meta-converse principle

Steps:

- Select auxiliary channel $Q_{Y|X}$ (art)

  E.g.: $Q_{Y|X=x}$ = center of a cluster of x
- Prove converse bound for channel $Q_{Y|X}$

  E.g.: $\mathbb{Q}[W = \hat{W}] \lesssim \frac{\text{\# of clusters}}{M}$
- Find $\beta_\alpha(\mathbb{P}, \mathbb{Q})$, i.e. compare:

$$\mathbb{P} : \ P_{WXY\hat{W}} = P_W \times P_{X|W} \times P_{Y|X} \times P_{\hat{W}|Y}$$

vs.

$$\mathbb{Q} : \ P_{WXY\hat{W}} = P_W \times P_{X|W} \times Q_{Y|X} \times P_{\hat{W}|Y}$$

- Amplify converse for $Q_{Y|X}$ to a converse for $P_{Y|X}$:

$$\beta_{1-P_e(P_{Y|X})} \le 1 - P_e(Q_{Y|X}) \qquad \forall \text{code}$$

Meta-converse theorem: point-to-point channels

Theorem

For any code $\epsilon \triangleq \mathbb{P}[\text{error}]$ *and* $\epsilon' \triangleq \mathbb{Q}[\text{error}]$ *satisfy*

$$\beta_{1-\epsilon}(P_X P_{Y|X}, P_X Q_{Y|X}) \le 1 - \epsilon'$$

Advanced examples of $Q_{Y|X}$:

- General DMC: $Q_{Y|X=x} = P_{Y|X} \circ \hat{P}_x$
  Why? To reduce DMC to symmetric DMC
- Parallel AWGN: $Q_{Y|X=x} = f(\text{power-allocation})$
  Why? Since water-filling is not FBL-optimal
- Feedback: $\mathbb{Q}[Y \in \cdot | W = w] = \mathbb{P}[Y \in \cdot | W \neq w]$
  Why? To get bounds in terms of Burnashev's C_1
- PAPR of codes: $Q_{Y^n|X^n=x^n} = f(\text{peak power of } x)$
  Why? To show peaky codewords waste power

Meta-converse generalizes many classical methods

Theorem

For any code $\epsilon \triangleq \mathbb{P}[\mathit{error}]$ *and* $\epsilon' \triangleq \mathbb{Q}[\mathit{error}]$ *satisfy*

$$\beta_{1-\epsilon}(P_X P_{Y|X}, P_X Q_{Y|X}) \le 1 - \epsilon'$$

Corollaries:

- Fano's inequality
- Wolfowitz strong converse
- Shannon-Gallager-Berlekamp's sphere-packing
  + improvements: [Valembois-Fossorier'04], [Wiechman-Sason'08]
- Haroutounian's sphere-packing
- list-decoding converses
- Berlekamp's low-rate converse
- Verdú-Han and Poor-Verdú information spectrum converses
- Arimoto's converse (+ extension to feedback)

Meta-converse generalizes many classical methods

Theorem

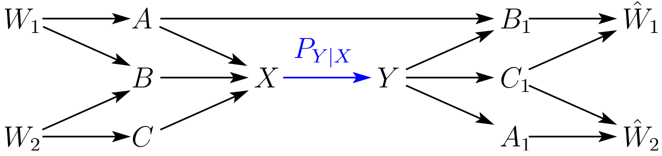
For any code $\epsilon \triangleq \mathbb{P}[\mathit{error}]$ *and* $\epsilon' \triangleq \mathbb{Q}[\mathit{error}]$ *satisfy*

$$\beta_{1-\epsilon}(P_X P_{Y|X}, P_X Q_{Y|X}) \le 1 - \epsilon'$$

Corollaries:

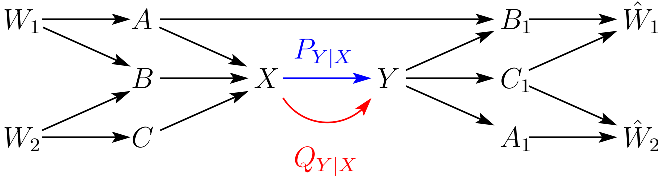
- Fano's inequality
- Wolfowitz strong converse
- Shannon-Gallager-Berlekamp's sphere-packing
  $+$ improvements: [Valembois-Fossorier'04], [Wiechman-Sason'08]
- Haroutounian's sphere-packing
- list-decoding converses E.g.: $\mathbb{Q}[W \in \{\text{list}\}] = \frac{|\{\text{list}\}|}{M}$
- Berlekamp's low-rate converse
- Verdú-Han and Poor-Verdú information spectrum converses
- Arimoto's converse ($+$ extension to feedback)

Meta-converse in networks



$$\{\text{error}\} = \{W_1 \neq \hat{W}_1\} \cup \{W_2 \neq \hat{W}_2\}$$

Meta-converse in networks



$$\{\text{error}\} = \{W_1 \neq \hat{W}_1\} \cup \{W_2 \neq \hat{W}_2\}$$

- Probability of error depends on channel:

$$\begin{aligned} \mathbb{P}[\text{error}] &= \epsilon\,, \\ \mathbb{Q}[\text{error}] &= \epsilon'\,. \end{aligned}$$

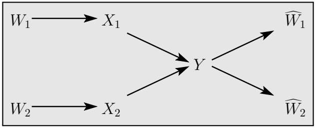
- Same idea: use code as a suboptimal binary HT: $P_{Y|X}$ vs. $Q_{Y|X}$
- ... and compare to the best possible test:

$$D(P_{XY} \| Q_{XY}) \geq d(1-\epsilon \| 1-\epsilon')$$

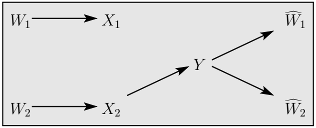
$$\beta_{1-\epsilon}(P_X P_{Y|X}, P_X Q_{Y|X}) \leq 1-\epsilon'$$

Example: MAC (weak-converse)

$\mathbb{P}$:



$\mathbb{Q}$:



$$\mathbb{P}[\hat{W}_{1,2} = W_{1,2}] = 1 - \epsilon \qquad \mathbb{Q}[\hat{W}_{1,2} = W_{1,2}] = \frac{1}{M_1}$$

... apply data processing of $D(\cdot||\cdot)$ ...

$$\Downarrow$$

$$d(1-\epsilon \| \tfrac{1}{M_1}) \le D(P_{Y|X_1X_2} \| Q_{Y|X_1} | P_{X_1} P_{X_2})$$

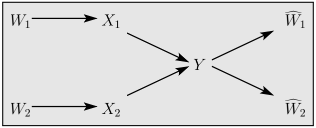
Optimizing $Q_{Y|X_1}$:

$$\log M_1 \le \frac{I(X_1; Y|X_2) + h(\epsilon)}{1-\epsilon}$$

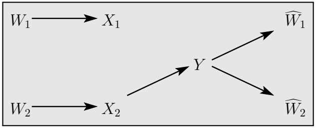
Also with $X_1 \leftrightarrow X_2 \implies$ weak converse (usual pentagon)

Example: MAC (FBL?)

$\mathbb{P}$:



$\mathbb{Q}$:



$$\mathbb{P}[\hat{W}_{1,2} = W_{1,2}] = 1 - \epsilon \qquad \mathbb{Q}[\hat{W}_{1,2} = W_{1,2}] = \frac{1}{M_1}$$

$$\ldots \text{ use } \beta_\alpha(\cdot,\cdot) \ldots$$
$$\Downarrow$$
$$\beta_{1-\epsilon}(P_{X_1 X_2 Y}, P_{X_1} Q_{X_2 Y}) \le \frac{1}{M_1}$$

On-going work: This β_α is highly non-trivial to compute.
[Huang-Moulin, MolavianJazi-Laneman, Yagi-Oohama]

Achievability Bounds

Notation

- A random transformation $\mathsf{A} \xrightarrow{P_{Y|X}} \mathsf{B}$
- (M, ϵ) codes:

$$W \to \mathsf{A} \to \mathsf{B} \to \hat{W} \qquad W \sim Unif\{1, \ldots, M\}$$

$$\mathbb{P}[W \neq \hat{W}] \leq \epsilon$$

- For every $P_{XY} = P_X P_{Y|X}$ define **information density**:

$$\imath_{X;Y}(x;y) \triangleq \log \frac{dP_{Y|X=x}}{dP_Y}(y)$$

  - $\mathbb{E}\left[\imath_{X;Y}(X;Y)\right] = I(X;Y)$
  - $\mathrm{Var}[\imath_{X;Y}(X;Y)|X] = V$
  - Memoryless channels: $\imath_{A^n;B^n}(A^n;B^n) =$ sum of iid.

$$\imath_{A^n;B^n}(A^n;B^n) \stackrel{d}{\approx} nI(A;B) + \sqrt{nV}Z, \qquad Z \sim \mathcal{N}(0,1)$$

- Goal: Prove FBL bounds.
  As by-product: $R^*(n,\epsilon) \gtrsim C - \sqrt{\frac{V}{n}}Q^{-1}(\epsilon)$

Achievability bounds: classical ideas

Goal: select codewords $C_1, \ldots, C_M$ in the input space A.

Two principal approaches:

- Random coding: generate $C_1, \ldots, C_M$ – iid with P_X and compute average probability of error [Shannon'48, Erdös'47].
- Maximal coding: choose C_j one by one until the output space is exhausted [Gilbert'52, Feinstein'54, Varshamov'57].

Complication: Many inequivalent ways to apply these ideas! Which ones are the best for FBL?

Classical bounds

- Feinstein'55 bound: $\exists(M,\epsilon)$-code:

$$M \geq \sup_{\gamma \geq 0} \left\{ \gamma(\epsilon - \mathbb{P}[\imath_{X;Y}(X;Y) \leq \log \gamma]) \right\}$$

- Shannon'57 bound: $\exists(M,\epsilon)$-code:

$$\epsilon \leq \inf_{\gamma \geq 0} \left\{ \mathbb{P}[\imath_{X;Y}(X;Y) \leq \log \gamma] + \frac{M-1}{\gamma} \right\} .$$

- Gallager'65 bound: $\exists(n,M,\epsilon)$-code over memoryless channel:

$$\epsilon \leq \exp\left\{ -nE_r\left(\frac{\log M}{n}\right) \right\} .$$

- Up to $M \leftrightarrow (M-1)$ Feinstein and Shannon are equivalent.

New bounds: RCU

Theorem (Random Coding Union Bound)

For any P_X there exists a code with M codewords and

$$\epsilon \le \mathbb{E}\left[\min\left\{1, (M-1)\pi(X,Y)\right\}\right]$$

$$\pi(a,b) = \mathbb{P}[\imath_{X;Y}(\bar{X};Y) \ge \imath_{X;Y}(X;Y) \,|\, X=a, Y=b]$$

where $P_{XY\bar{X}}(a,b,c) = P_X(a)P_{Y|X}(b|a)P_X(c)$

Proof:

- Reason as in RCU for BSC with $d_{Ham}(\cdot,\cdot) \leftrightarrow -\imath_{X;Y}(\cdot,\cdot)$
- For example ML decoder: $\hat{W} = \mathrm{argmax}_j\, \imath_{X;Y}(C_j;Y)$
- Conditional prob. of error:

$$\mathbb{P}[\text{error} \,|\, X,Y] \le (M-1)\mathbb{P}[\imath_{X;Y}(\bar{X};Y) \ge \imath_{X;Y}(X;Y) \,|\, X,Y]$$

- Same idea: take $\min\{\cdot, 1\}$ before averaging over (X,Y).

New bounds: RCU

Theorem (Random Coding Union Bound)

For any P_X there exists a code with M codewords and

$$\epsilon \le \mathbb{E}\left[\min\left\{1, (M-1)\pi(X,Y)\right\}\right]$$
$$\pi(a,b) = \mathbb{P}[\imath_{X;Y}(\bar{X};Y) \ge \imath_{X;Y}(X;Y)\,|\,X=a, Y=b]$$

where $P_{XY\bar{X}}(a,b,c) = P_X(a)P_{Y|X}(b|a)P_X(c)$

Highlights:

- Strictly stronger than Feinstein-Shannon and Gallager
- Not easy to analyze asymptotics
- Computational complexity $O(n^{2(|X|-1)|Y|})$

New bounds: DT

Theorem (Dependence Testing Bound)

For any P_X there exists a code with M codewords and

$$\epsilon \le \mathbb{E}\left[\exp\left\{-\left|\imath_{X;Y}(X;Y) - \log\frac{M-1}{2}\right|^+\right\}\right].$$

Highlights:

- Strictly stronger than Feinstein-Shannon
- … and no optimization over γ!
- Easier to compute than RCU
- Easier asymptotics: $\epsilon \le \mathbb{E}\left[e^{-n\left|\frac{1}{n}\imath(X^n;Y^n)-R\right|^+}\right]$
  $\approx Q\left(\sqrt{\frac{n}{V}}\{I(X;Y)-R\}\right)$
- Has a form of f-divergence: $1-\epsilon \ge D_f(P_{XY}\|P_X P_Y)$

DT bound: Proof

- Codebook: random $C_1, \dots C_M \sim P_X$ iid
- Feinstein decoder:

$$\hat{W} = \text{smallest } j \text{ s.t. } \imath_{X;Y}(C_j; Y) > \gamma$$

- j-th codeword's probability of error:

$$\mathbb{P}[\text{error} \mid W = j] \le \underbrace{\mathbb{P}[\imath_{X;Y}(X;Y) \le \gamma]}_{a} + (j-1)\underbrace{\mathbb{P}[\imath_{X;Y}(\bar{X};Y) > \gamma]}_{b}$$

In a: C_j too far from Y
In b: C_k with $k < j$ is too close to Y

- Average over W:

$$\mathbb{P}[\text{error}] \le \mathbb{P}\left[\imath_{X;Y}(X;Y) \le \gamma\right] + \frac{M-1}{2}\mathbb{P}\left[\imath_{X;Y}(\bar{X};Y) > \gamma\right]$$

DT bound: Proof

- Recap: for every γ there exists a code with

$$\epsilon \le \mathbb{P}\left[\imath_{X;Y}(X;Y) \le \gamma\right] + \frac{M-1}{2}\mathbb{P}\left[\imath_{X;Y}(\bar{X};Y) > \gamma\right] .$$

- Key step: closed-form optimization of γ.
  Note: $\imath_{X;Y} = \log \frac{dP_{XY}}{dP_{\bar{X}Y}}$

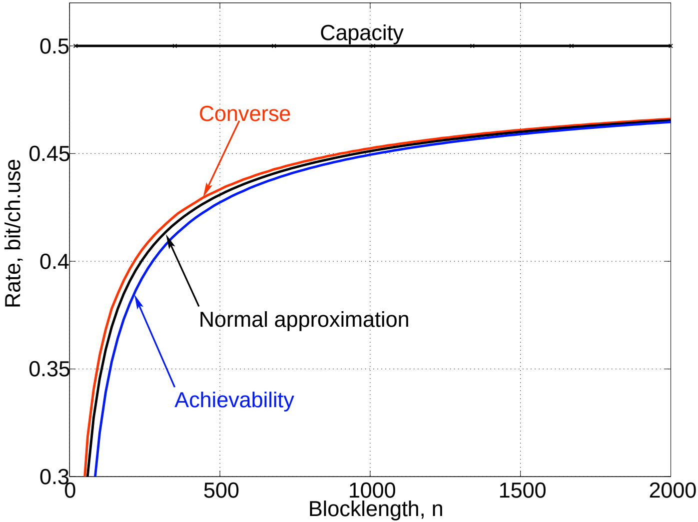
$$\frac{M+1}{2}\left(\frac{2}{M+1}P_{XY}\left[\frac{dP_{XY}}{dP_{\bar{X}Y}} \le e^{\gamma}\right] + \frac{M-1}{M+1}P_{\bar{X}Y}\left[\frac{dP_{XY}}{dP_{\bar{X}Y}} > e^{\gamma}\right]\right)$$

  Bayesian dependence testing!
  Optimum threshold: Ratio of priors $\Longrightarrow$ $\boxed{\gamma^* = \log \frac{M-1}{2}}$

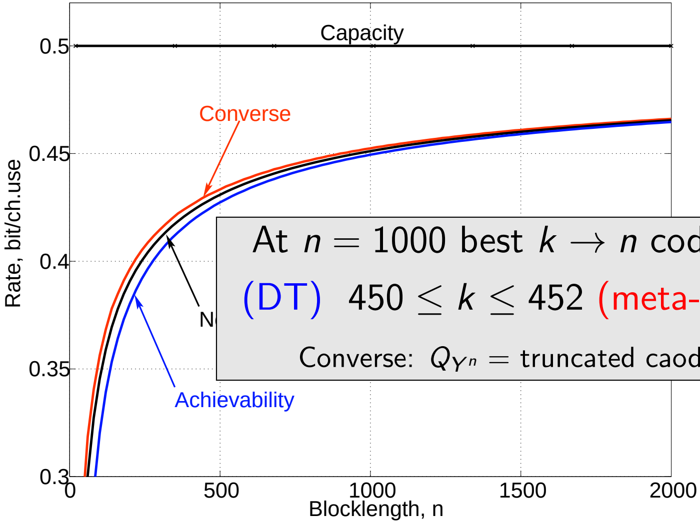
- Change of measure argument:

$$P\left[\frac{dP}{dQ} \le \tau\right] + \tau Q\left[\frac{dP}{dQ} > \tau\right] = \mathbb{E}_P\left[\exp\left\{-\left|\log\frac{dP}{dQ} - \log\tau\right|^{+}\right\}\right] .$$

Example: Binary Erasure Channel $BEC(0.5)$, $\epsilon = 10^{-3}$



Example: Binary Erasure Channel $BEC(0.5)$, $\epsilon = 10^{-3}$



At $n = 1000$ best $k \to n$ code:

(DT) $450 \leq k \leq 452$ (meta-c.)

Converse: $Q_{Y^n} =$ truncated caod

Input constraints: $\kappa\beta$ bound

Theorem

For all Q_Y and τ there exists an (M, ϵ)-code inside $\mathsf{F} \subset \mathsf{A}$

$$M \geq \frac{\kappa_\tau}{\sup_x \beta_{1-\epsilon+\tau}(P_{Y|X=x}, Q_Y)}$$

where

$$\kappa_\tau = \inf_{\{E\colon P_{Y|X}[E|x] \geq \tau \ \forall x \in \mathsf{F}\}} Q_Y[E]$$

Highlights:

- Key for channels with cost constraints (e.g. AWGN).
- Bound parameterized by the output distribution.
- Reduces coding to binary HT.

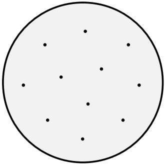
$\kappa\beta$ bound: idea

Decoder:

- Take received $\mathbf{y}$.
- Test $\mathbf{y}$ against *each* codeword $\mathbf{c}_i$, $i = 1, \ldots M$:

  Run optimal binary HT for :

$$\begin{array}{ccc} \mathcal{H}_0 & : & P_{Y|X=c_i} \\ \mathcal{H}_1 & : & Q_Y \end{array}$$



$$\begin{array}{rcl} \mathbb{P}[\text{detect } \mathcal{H}_0] & = & 1 - \epsilon + \tau \\ \mathbb{Q}[\text{detect } \mathcal{H}_0] & = & \beta_{1-\epsilon+\tau}(P_{Y|X=x}, Q_Y) \end{array}$$

- First test that returns $\mathcal{H}_0$ becomes **the decoded codeword**.
- If all $\mathcal{H}_1$ – **declare error**.

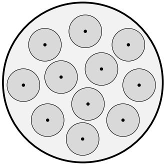
$\kappa\beta$ bound: idea

Decoder:

- Take received $\mathbf{y}$.
- Test $\mathbf{y}$ against *each* codeword $\mathbf{c}_i$, $i = 1, \ldots M$:

  Run optimal binary HT for :

$$\begin{array}{ccc} \mathcal{H}_0 & : & P_{Y|X=c_i} \\ \mathcal{H}_1 & : & Q_Y \end{array}$$



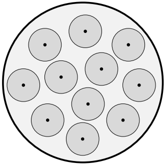
$$\begin{array}{rcl} \mathbb{P}[\text{detect } \mathcal{H}_0] & = & 1 - \epsilon + \tau \\ \mathbb{Q}[\text{detect } \mathcal{H}_0] & = & \beta_{1-\epsilon+\tau}(P_{Y|X=x}, Q_Y) \end{array}$$

- First test that returns $\mathcal{H}_0$ becomes **the decoded codeword**.
- If all $\mathcal{H}_1$ – **declare error**.

$\kappa\beta$ bound: idea

Codebook:

- Pick codewords s.t. "balls" are τ-disjoint: $\mathbb{P}[Y \in B_x \cap \text{others} | x] \le \tau$
- Key step: Cannot pick more codewords $\Longrightarrow$ $\bigcup_{j=1}^{M} \{j\text{-th decoding region}\}$ is a composite HT:



$$\begin{array}{lcll} \mathcal{H}_0 & : & P_{Y|X=x} & x \in \mathsf{F} \\ \mathcal{H}_1 & : & Q_Y & \end{array}$$

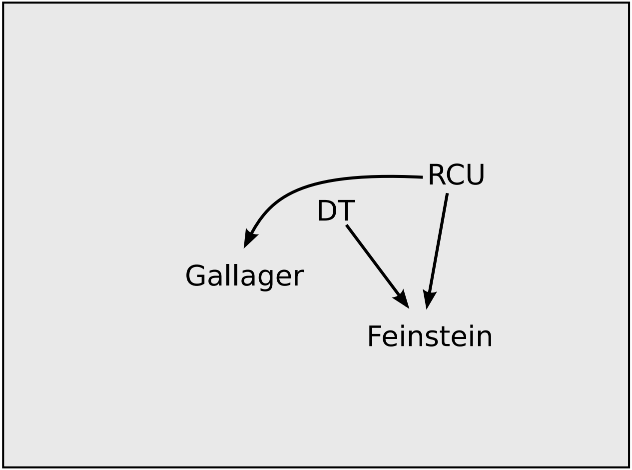
- Performance of the best test:

$$\kappa_\tau = \inf_{\{E:\, P_{Y|X}[E|x] \ge \tau \ \forall x \in \mathsf{F}\}} Q_Y[E]\,.$$

- Thus:

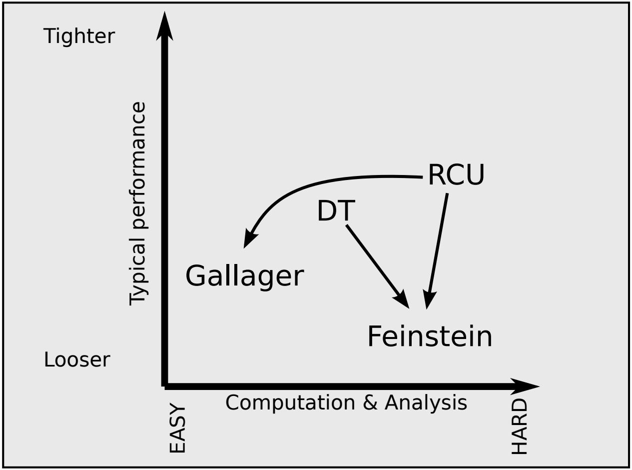
$$\begin{aligned} \kappa_\tau &\le \mathbb{Q}[\text{all } M \text{ "balls"}] \\ &\le M \sup_x \beta_{1-\epsilon+\tau}(P_{Y|X=x}, Q_Y) \end{aligned}$$

Hierarchy of achievability bounds (no cost constr.)



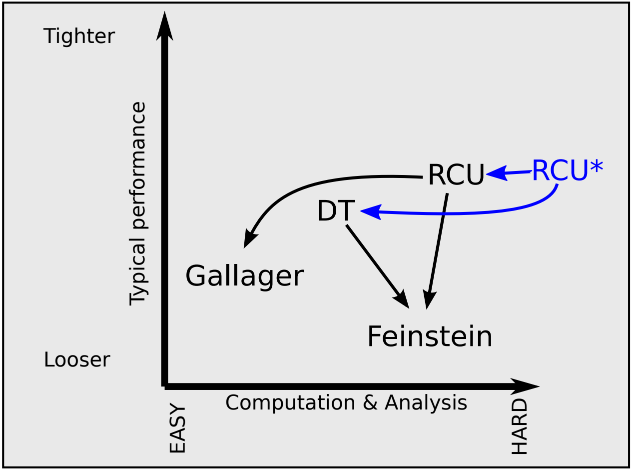
- Arrows show logical implication

Hierarchy of achievability bounds (no cost constr.)



- Arrows show logical implication
- Performance $\leftrightarrow$ computation **rule of thumb**.

Hierarchy of achievability bounds (no cost constr.)



- Arrows show logical implication
- Performance $\leftrightarrow$ computation **rule of thumb**.
- **ISIT'2013**: Haim-Kochman-Erez [WeB4]

Channel Dispersion

Connection to CLT

Recap:

- Let $P_{Y^n|X^n} = P^n_{Y|X}$ be memoryless. FBL fundamental limit:

$$R^*(n,\epsilon) = \max\left\{\frac{1}{n}\log M : \exists (n, M, \epsilon)\text{-code}\right\}$$

- Converse bounds (roughly):

$$R^*(n,\epsilon) \lesssim \epsilon\text{-th quantile of } \frac{1}{n}\log\frac{dP_{Y^n|X^n}}{dQ_{Y^n}}$$

- Achievability bounds (roughly):

$$R^*(n,\epsilon) \gtrsim \epsilon\text{-th quantile of } \frac{1}{n}\imath_{X^n;Y^n}(X^n;Y^n)$$

- Both random variables have form: $\frac{1}{n}\cdot$(sum of iid) $\Longrightarrow$ by CLT

$$R^*(n,\epsilon) = C + \theta\left(\frac{1}{\sqrt{n}}\right)$$

This section: Study $\sqrt{n}$-term.

General definition of channel dispersion

Definition

For any channel we define channel dispersion as

$$V = \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{n\left(C - R^*(n,\epsilon)\right)^2}{2 \ln \frac{1}{\epsilon}}$$

Rationale is the expansion (see below)

$$R^*(n,\epsilon) = C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon) + o\left(\frac{1}{\sqrt{n}}\right) \quad (*)$$

and the fact $Q^{-1}(\epsilon) \sim 2 \ln \frac{1}{\epsilon}$ for $\epsilon \to 0$

Recall: Approximation via $(*)$ is remarkably tight

General definition of channel dispersion

Definition

For any channel we define channel dispersion as

$$V = \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{n\left(C - R^*(n,\epsilon)\right)^2}{2 \ln \frac{1}{\epsilon}}$$

Heuristic connection to error exponents $E(R)$:

$$E(R) = \frac{(R-C)^2}{2} \cdot \frac{\partial^2 E(R)}{\partial R^2} + o((R-C)^2)$$

and thus

$$V = \left(\frac{\partial^2 E(R)}{\partial R^2}\right)^{-1}$$

Dispersion of memoryless channels

- DMC [Dobrushin'61,Strassen'62]:

$$V = \mathrm{Var}[\imath_{X;Y}(X;Y)] \qquad X \sim \text{capacity-achieving}$$

- AWGN channel [PPV'08]:

$$V = \frac{\log^2 e}{2}\left[1 - \left(\frac{1}{1+\mathrm{SNR}}\right)^2\right]$$

- Parallel AWGN [PPV'09]:

$$V = \sum_{j=1}^{L} V_{AWGN}\left(\frac{W_j}{\sigma_j^2}\right) \qquad \{W_j\}\text{–waterfilling powers}$$

- DMC with input constraints [Hayashi'09,P'10]:

$$V = \mathrm{Var}[\imath_{X;Y}(X;Y)|X] \quad X \sim \text{capacity-achieving}$$

Dispersion of channels with memory

From [PPV'09, PPV'10, PV'11]:

- Non-white Gaussian noise with PSD $N(f)$:

$$V = \frac{\log^2 e}{2} \int_{-1/2}^{1/2} \left[1 - \frac{|N(f)|^4}{P^2 \xi^2}\right]^+ df, \qquad \int_{-1/2}^{1/2} \left[\xi - \frac{|N(f)|^2}{P}\right]^+ df = 1$$

- AWGN subject to stationary fading process H_i (CSI at receiver):

$$V = \mathrm{PSD}_{\frac{1}{2}\log(1+PH_i^2)}(0) + \frac{\log^2 e}{2}\left(1 - \mathbb{E}^2\left[\frac{1}{1+PH_0^2}\right]\right)$$

- State-dependent discrete additive noise (CSI at receiver):

$$V = \mathrm{PSD}_{C(S_i)}(0) + \mathbb{E}\left[V(S)\right]$$

- **ISIT'13**: Quasi-static fading channels: $V = 0$ (!)
  Yang-Durisi-Koch-P. in [WeA4]

Dispersion: product vs generic channels

- Relation to alphabet size:

$$V \le 2\log^2 \min(|\mathcal{A}|, |\mathcal{B}|) - C^2 \,.$$

- Dispersion is additive:

$$\left\{ \begin{array}{ccccc} \mathcal{A}_1 & \to & \boxed{DMC_1} & \to & \mathcal{B}_1 \\ \mathcal{A}_2 & \to & \boxed{DMC_2} & \to & \mathcal{B}_2 \end{array} \right\} = \mathcal{A}_1 \times \mathcal{A}_2 \to \boxed{DMC} \to \mathcal{B}_1 \times \mathcal{B}_2$$

$$C = C_1 + C_2\,, \qquad V_\epsilon = V_{1,\epsilon} + V_{2,\epsilon}$$

- $\Longrightarrow$ product DMCs have atypically low dispersion.

Dispersion and normal approximation

Let $P_{Y|X}$ be DMC and

$$V_\epsilon \triangleq \begin{cases} \max_{P_X} \mathrm{Var}[i(X,Y)|X]\,, & \epsilon < 1/2\,, \\ \min_{P_X} \mathrm{Var}[i(X,Y)|X]\,, & \epsilon > 1/2 \end{cases}$$

where optimization is over all P_X s.t. $I(X;Y) = C$.

Theorem (Strassen'62)

$$R^*(n,\epsilon) = C - \sqrt{\frac{V_\epsilon}{n}} Q^{-1}(\epsilon) + O\left(\frac{\log n}{n}\right)$$

But [PPV'10]: a counter-example with

$$R^*(n,\epsilon) = C + \Theta\left(n^{-\frac{2}{3}}\right)$$

Dispersion and normal approximation

Let $P_{Y|X}$ be DMC and

$$V_\epsilon \triangleq \begin{cases} \max_{P_X} \mathrm{Var}[i(X,Y)|X]\,, & \epsilon < 1/2\,, \\ \min_{P_X} \mathrm{Var}[i(X,Y)|X]\,, & \epsilon > 1/2 \end{cases}$$

where optimization is over all P_X s.t. $I(X;Y) = C$.

Theorem (Strassen'62, PPV'10)

$$R^*(n,\epsilon) = C - \sqrt{\frac{V_\epsilon}{n}} Q^{-1}(\epsilon) + O\left(\frac{\log n}{n}\right)$$

unless DMC is exotic in which case $O\left(\frac{\log n}{n}\right)$ becomes $O(n^{-\frac{2}{3}})$.

Further results on $O\left(\frac{\log n}{n}\right)$

- For **BEC** we have:

$$R^*(n,\epsilon) = C - \sqrt{\frac{V}{n}}Q^{-1}(\epsilon) + 0 \cdot \frac{\log n}{n} + O\left(\frac{1}{n}\right)$$

- For most other symmetric channels (incl. **BSC** and **AWGN***):

$$R^*(n,\epsilon) = C - \sqrt{\frac{V}{n}}Q^{-1}(\epsilon) + \frac{1}{2}\frac{\log n}{n} + O\left(\frac{1}{n}\right)$$

- For **most DMC** (under mild conditions):

$$R^*(n,\epsilon) \geq C - \sqrt{\frac{V_\epsilon}{n}}Q^{-1}(\epsilon) + \frac{1}{2}\frac{\log n}{n} + O\left(\frac{1}{n}\right)$$

- ISIT'13: For **all DMC**

$$R^*(n,\epsilon) \leq C - \sqrt{\frac{V_\epsilon}{n}}Q^{-1}(\epsilon) + \frac{1}{2}\frac{\log n}{n} + O\left(\frac{1}{n}\right)$$

Tomamichel-Tan, Moulin in [WeA4]

Applications

Evaluating performance of real-world codes

- Comparing codes: usual method – waterfall plots P_e vs. SNR
- **Problem:** Not fair for different rates.

  $\Longrightarrow$ define rate-invariant metric:

Evaluating performance of real-world codes

- Comparing codes: usual method – waterfall plots P_e vs. SNR
- **Problem:** Not fair for different rates.

  $\Longrightarrow$ define rate-invariant metric:

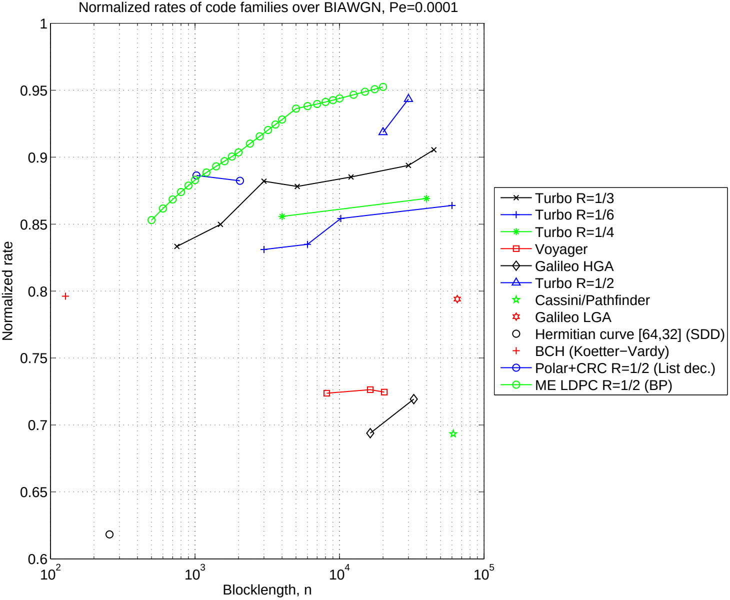
Definition (Normalized rate)

Given rate R code find SNR at which $P_e = \epsilon$.

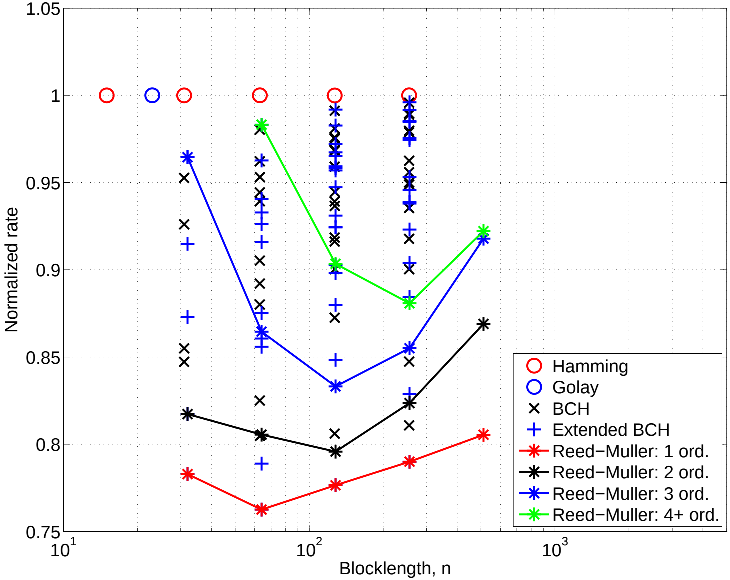
$$R_{norm} = \frac{R}{R^*(n, \epsilon, SNR)}$$

- Agreement: $\epsilon = 10^{-3}$ or 10^{-4}
- Take $R^*(n, \epsilon, SNR) \approx C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon)$
- A family of channels needed (e.g. AWGN or BSC)

Codes vs. fundamental limits (from 1970's to 2012)



Performance of short algebraic codes (BSC, $\epsilon = 10^{-3}$)



Optimizing ARQ systems

- End-user wants $P_e = 0$
- Usual method: automatic repeat request (ARQ)

$$\text{average throughput} = \text{Rate} \times (1 - \mathbb{P}[\text{error}])$$

- **Question:** Given k bits what rate (equiv. ϵ) maximizes throughput?

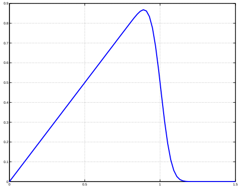
Optimizing ARQ systems

- End-user wants $P_e = 0$
- Usual method: automatic repeat request (ARQ)

$$\text{average throughput} = \text{Rate} \times (1 - \mathbb{P}[\text{error}])$$

- **Question:** Given k bits what rate (equiv. ϵ) maximizes throughput?
- Assume (C, V) is known. Then approximately

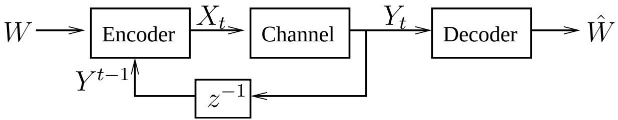
$$T^*(k) \approx \max_R R \cdot \left(1 - Q\left(\sqrt{\frac{kR}{V}}\left\{\frac{C}{R} - 1\right\}\right)\right)$$



- Solution: $\epsilon^*(k) \sim \frac{1}{\sqrt{kt \log kt}}, \quad t = \frac{C}{V}$
- Punchline: For $k \sim 1000$ bit and reasonable channels

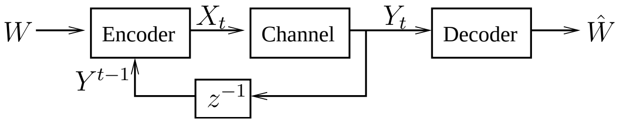
$$\boxed{\epsilon \approx 10^{-3}..10^{-2}}$$

Benefits of feedback: From ARQ to Hybrid ARQ



- Memoryless channels: feedback does not improve C [Shannon'56]
- Question: What about higher order terms?

Benefits of feedback: From ARQ to Hybrid ARQ



- Memoryless channels: feedback does not improve C [Shannon'56]
- Question: What about higher order terms?

Theorem

For any DMC with capacity C and $0 < \epsilon < 1$ we have for codes with feedback and variable length:

$$R_f^*(n, \epsilon) = \frac{C}{1-\epsilon} + O\left(\frac{\log n}{n}\right) .$$

Note: dispersion is zero!

Stop feedback bound (BSC version)

Theorem

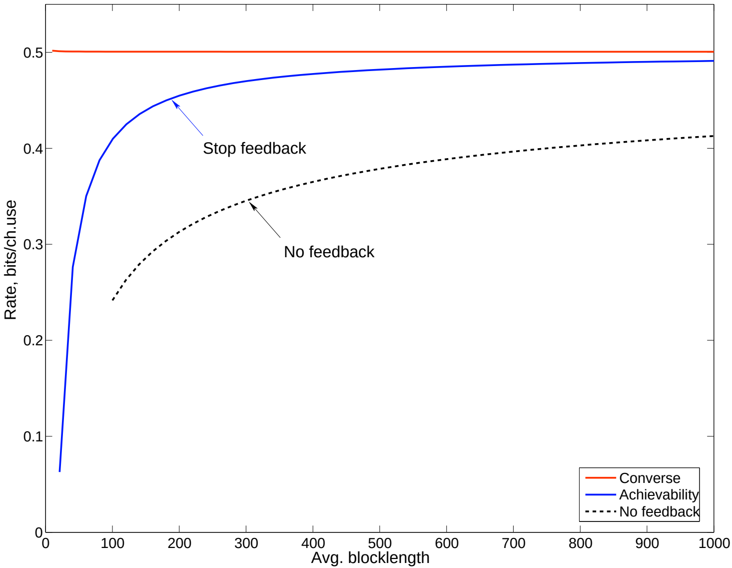
For any $\gamma > 0$ there exists a stop feedback code of rate R, average length $\ell = \mathbb{E}[\tau]$ and probability of error over BSC(δ)

$$\epsilon \le \mathbb{E}\left[f(\tau)\right],$$

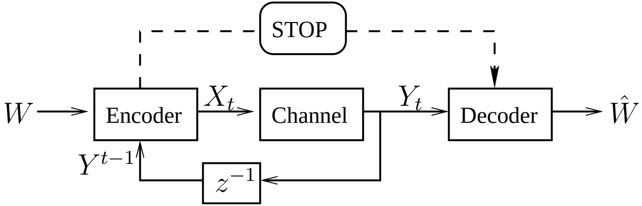
where

$$\begin{aligned} f(n) &\triangleq \mathbb{E}\left[1\{\tau \le n\}2^{\ell R - S_\tau}\right] \\ \tau &\triangleq \inf\{n \ge 0 : S_n \ge \gamma\} \\ S_n &\triangleq n\log(2-2\delta) + \log\frac{\delta}{1-\delta}\cdot\sum_{k=1}^{n} Z_k \\ Z_k &\sim \textit{i.i.d. Bernoulli}(\delta)\,. \end{aligned}$$

Feedback codes for BSC(0.11), $\epsilon = 10^{-3}$

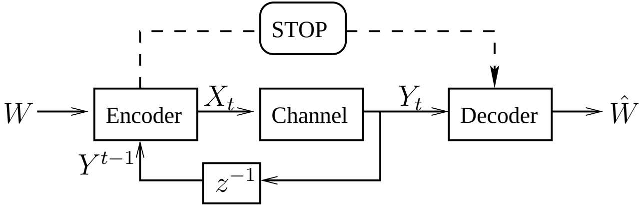


Effects of flow control



- Modeling of packet termination
- Often: reliability of start/end $\gg$ reliability of payload

Effects of flow control



- Modeling of packet termination
- Often: reliability of start/end $\gg$ reliability of payload

Theorem

If reliable termination is available, then there exist codes with variable length and feedback achieving

$$R_t^*(n, 0) \geq C + O\left(\frac{1}{n}\right) .$$

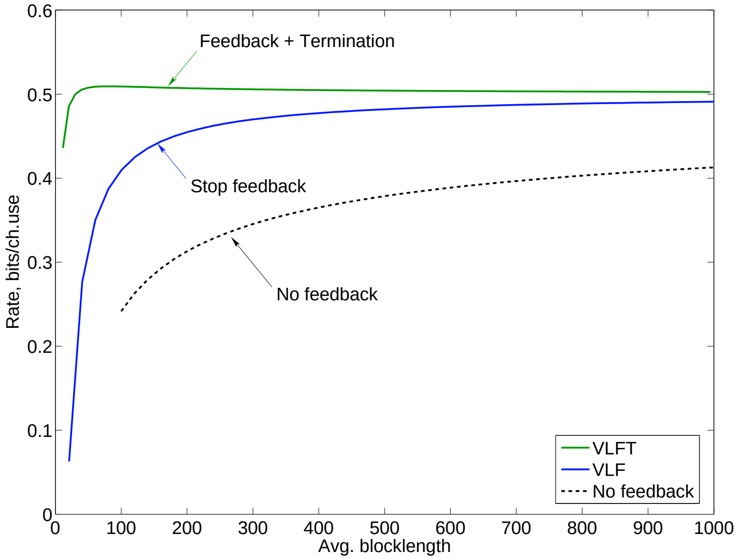
Codes employing feedback + termination (BSC version)

Theorem

Consider a BSC(δ) with feedback and reliable termination. There exists a code sending k bits with zero error and average length

$$\ell \leq \sum_{n=0}^{\infty} \sum_{t=0}^{n} \binom{n}{t} \delta^t (1-\delta)^{n-t} \min\left\{1, \sum_{k=0}^{t} \binom{n}{k} 2^{k-n}\right\} .$$

Feedback + termination for the BSC(0.11)



Benefit of feedback

Delay to achieve 90% of the capacity of the BSC(0.11):

- No feedback:

$$n \approx 3100$$

- Stop feedback + variable-length:

$$n \lesssim 200$$

- Feedback + variable-length + termination:

$$n \lesssim 20$$

Benefit of feedback

Delay to achieve 90% of the capacity of the BSC(0.11):

- No feedback:

$$n \approx 3100 \qquad \text{penalty term:} \quad O\left(\frac{1}{\sqrt{n}}\right)$$

- Stop feedback + variable-length:

$$n \lesssim 200 \qquad \text{penalty term:} \quad O\left(\frac{\log n}{n}\right)$$

- Feedback + variable-length + termination:

$$n \lesssim 20 \qquad \text{penalty term:} \quad O\left(\frac{1}{n}\right)$$

Gaussian channel: Energy per bit

$$
\begin{array}{ccccc}
 & & Z_i \sim \mathcal{N}\left(0, \frac{N_0}{2}\right) & & \\
 & & \downarrow & & \\
X_i & \longrightarrow & \bigoplus & \longrightarrow & Y_i
\end{array}
$$

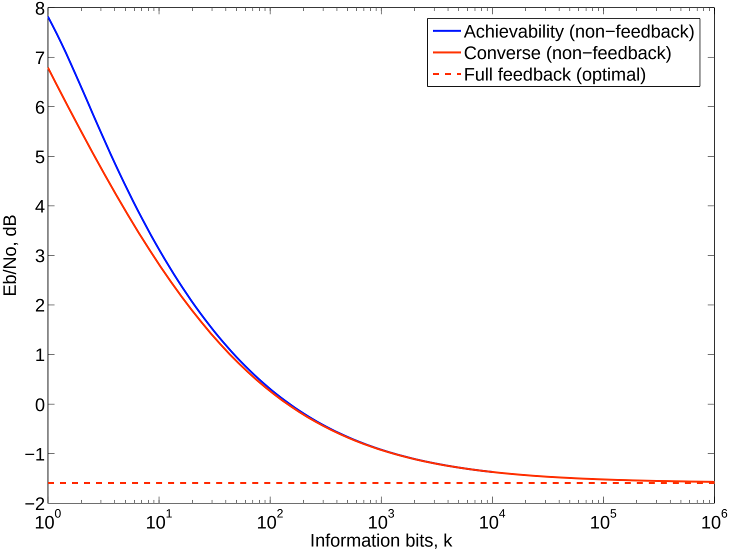
Problem: minimal energy-per-bit E_b *vs.* payload size k:

$$\mathbb{E}\left[\sum_{i=1}^{n} |X_i|^2\right] \le kE_b \,.$$

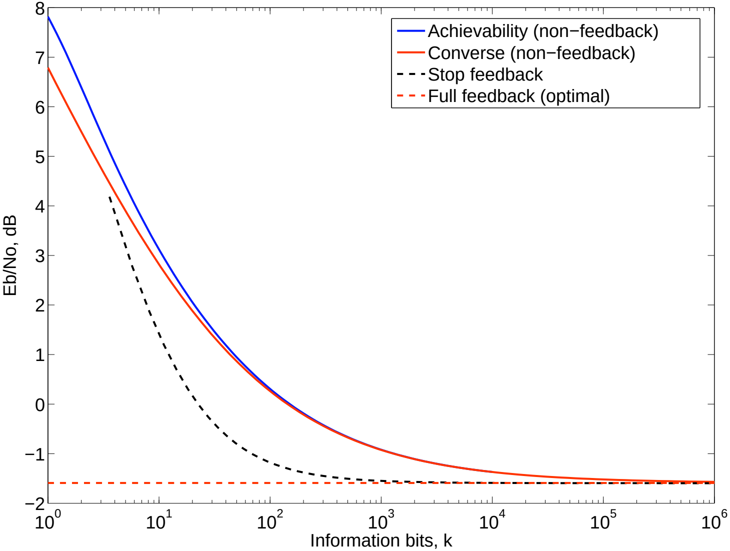
Asymptotically: [Shannon'49]

$$\min\left(\frac{E_b}{N_0}\right) \to \log 2 = -1.6 \text{ dB} \qquad , k \to \infty \,.$$

Energy per bit vs. # of information bits ($\epsilon = 10^{-3}$)

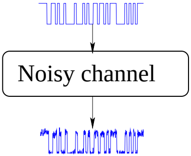


Energy per bit vs. # of information bits ($\epsilon = 10^{-3}$)

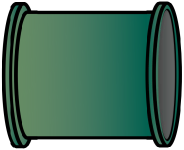


Summary

Classical: $(n \to \infty)$

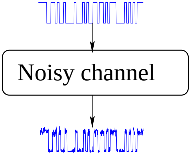


Optimal coding $\Longrightarrow$

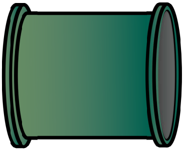


C

Finite blocklength: $(n - \text{finite})$



Optimal coding $\Longrightarrow$



$\mathcal{N}\left(C, \frac{V}{n}\right)$

What we had to skip

- **Hypothesis testing methods in Quantum IT:** [Wang-Colbeck-Renner'09], [Matthews-Wehner'12], [Tomamichel'12],[Kumagai-Hayashi'13]
- **Channels with state:** [Ingber-Feder'10], [Tomamichel-Tan'13], [Yang-Durisi-Koch-P.'12]
- **FBL theory of lattice codes:** [Ingber-Zamir-Feder'12]
- **Feedback codes:** [Naghshvar-Javidi'12], [Williamson-Chen-Wesel'12]
- **Random coding bounds and approximations:** [Martinez-Guillen i Fabregas'11], [Kosut-Tan'12]
- **Other FBL questions:** [Riedl-Coleman-Singer'11], [Varshney-Mitter-Goyal'12], [Asoodeh-Lapidoth-Wang'12], [P.-Wu'13]

...and many more (apologies!) ... (cf: References)

New results at ISIT'2013: two terminals

- Universal lossless compression: Kosut-Sankar [MoD5]
- Random number generation: Kumagai-Hayashi [WeA3]
- Quasi-static SIMO: Yang-Durisi-Koch-P. [WeA4]
- $O(\log n) = \frac{1}{2} \log n$: Tomamichel-Tan, Moulin [WeA4]
- Meta-converse is tight: Vazquez-Vilar et al [WeB4]
- Meta-converse for unequal error protection: Shkel-Tan-Draper [WeB4]
- RCU* bound: Haim-Kochman-Erez [WeB4]
- Cost constraints: Kostina-Verdú [WeB4]
- Lossless compression: Kontoyiannis-Verdú [WeB5]
- Feedback: Chen-Williamson-Wesel [ThD6]

New results at ISIT'2013: multi-terminal

- achievability bounds: Yassae-Aref-Gohari [TuD1]
- random binning: Yassae-Aref-Gohari [ThA1]
- interference channel: Le-Tan-Motani [ThA1]
- Gaussian line network: Subramanian-Vellambi-Land [ThA1]
- Slepian-Wolf for mixed sources: Nomura-Han [ThA7]
- one-help-one and Wyner-Ziv: Watanabe-Kuzuoka-Tan [FrC5]

References: Lossless compression

A. A. Yushkevich, "On limit theorems connected with the concept of entropy of Markov chains", Uspekhi Matematicheskikh Nauk, 8:5(57), pp. 177-180, 1953

V. Strassen, "Asymptotische Abschätzungen in Shannon's Informationstheorie," *Trans. Third Prague Conf. Information Theory*, 1962, Czechoslovak Academy of Sciences, Prague, pp. 689-723. Translation: `http://www.math.cornell.edu/~pmlut/strassen.pdf`

I. Kontoyiannis, "Second-order noiseless source coding theorems," IEEE Trans. Inform. Theory, vol. 43, no. 3, pp. 1339-1341, July 1997

M. Hayashi, "Second-order asymptotics in fixed-length source coding and intrinsic randomness," IEEE Trans. Inform. Theory, vol. 54, no. 10, pp. 4619-4637, October 2008.

W. Szpankowski and S. Verdú, "Minimum expected length of fixed-to-variable lossless compression without prefix constraints," IEEE Trans. on Information Theory, vol. 57, no. 7, pp. 4017-4025, July 2011.

S. Verdú and I. Kontoyiannis, "Lossless Data Compression Rate: Asymptotics and Non-Asymptotics," 46th Annual Conference on Information Sciences and Systems, Princeton, New Jersey, March 21-23, 2012

R. Nomura and T. S. Han, "Second-Order Resolvability, Intrinsic Randomness, and Fixed-Length Source Coding for Mixed Sources: Information Spectrum Approach," IEEE Trans. Inform. Theory, vol.59, no.1, pp.1,16, Jan. 2013

I. Kontoyiannis and S. Verdú, "Optimal lossless compression: Source varentropy and dispersion," Proc. 2013 IEEE Int. Symposium on Information Theory, Istanbul, Turkey, July 7-12, 2013

References: multi-terminal compression

S. Sarvotham, D. Baron, and R. G. Baraniuk, "Non-asymptotic performance of symmetric Slepian-Wolf coding," in Conference on Information Sciences and Systems, 2005.

D.-K. He, L. A. Lastras-Montanõ, E.-H. Yang, A. Jagmohan, and J. Chen, "On the redundancy of Slepian-Wolf coding," IEEE Trans. Inform. Theory, vol. 55, no. 12, pp. 5607-27, Dec 2009.

V. Y. F. Tan and O. Kosut, "The dispersion of Slepian-Wolf coding," in Proc. 2012 IEEE Intl. Symp. Inform. Theory (ISIT), Jul. 2012, pp. 915-919.

S. Kuzuoka, "On the redundancy of variable-rate Slepian-Wolf coding," Inform. Theory and its Applications (ISITA), 2012 International Symposium on (pp. 155-159), Oct. 2012.

S. Verdú, "Non-asymptotic achievability bounds in multiuser information theory," in Allerton Conference, 2012.

References: Channel coding (1948-2008)

A. Feinstein, "A new basic theorem of information theory," *IRE Trans. Inform. Theory*, vol. 4, no. 4, pp. 2-22, 1954.

P. Elias, "Coding for two noisy channels," *Proc. Third London Symposium on Information Theory,* pp. 61-76, Buttersworths, Washington, DC, Sept. 1955

C. E. Shannon, "Certain results in coding theory for noisy channels," *Inform. Control*, vol. 1, pp. 6-25, 1957.

J. Wolfowitz, "The coding of messages subject to chance errors," *Illinois J. Math.*, vol. 1, pp. 591-606, 1957.

C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," Bell System Tech. Journal, vol. 38, pp. 611-656, 1959.

R. L. Dobrushin, "Mathematical problems in the Shannon theory of optimal coding of information," *Proc. 4th Berkeley Symp. Mathematics, Statistics, and Probability,* 1961, vol. 1, pp. 211-252.

V. Strassen, "Asymptotische Abschätzungen in Shannon's Informationstheorie," *Trans. Third Prague Conf. Information Theory*, 1962, Czechoslovak Academy of Sciences, Prague, pp. 689-723. Translation: `http://www.math.cornell.edu/~pmlut/strassen.pdf`

R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inform. Theory*, vol. 11, no. 1, pp. 3-18, 1965.

C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels I", *Inform. Contr.*, vol. 10, pp. 65-103, 1967

J. Wolfowitz, "Notes on a general strong converse," *Inform. Contr.*, vol. 12, pp. 1-4, 1968.

H. V. Poor and S. Verdú, "A lower bound on the error probability in multihypothesis testing," *IEEE Trans. Inform. Theory*, vol. 41, no. 6, pp. 1992-1993, 1995.

A. Valembois and M. P. C. Fossorier, "Sphere-packing bounds revisited for moderate block lengths," *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 2998-3014, 2004.

G. Wiechman and I. Sason, "An improved sphere-packing bound for finite-length codes over symmetric memoryless channels," *IEEE Trans. Inform. Theory*, vol. 54, no. 5, pp. 1962-1990, 2008.

References: Channel coding (2008-)

Y. Polyanskiy, H. V. Poor and S. Verdú, "New channel coding achievability bounds," *Proc. 2008 IEEE Int. Symp. Information Theory (ISIT),* Toronto, Canada, 2008.

L. Wang, R. Colbeck and R. Renner, "Simple channel coding bounds," *Proc. 2009 IEEE Int. Symp. Information Theory (ISIT),* Seoul, Korea, July 2009.

M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Trans. Inform. Theory*, vol. 55, no. 11, pp. 4947-4966, Nov 2009.

Y. Polyanskiy, H. V. Poor and S. Verdú, "Dispersion of Gaussian channels," 2009 IEEE Int. Symp. Inf. Theory (ISIT), Seoul, Korea, Jul. 2009.

L. Varshney, S. K. Mitter and V. Goyal, "Channels that die," Communication, Control, and Computing, Allerton, 2009.

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

Y. Polyanskiy, H. V. Poor and S. Verdú, "Dispersion of the Gilbert-Elliot channel," IEEE Trans. Inf. Theory, vol. 57, no. 4, pp. 1829-1848, Apr. 2011.

Y. Polyanskiy and S. Verdú, "Channel dispersion and moderate deviations limits for memoryless channels," 48th Allerton Conference 2010, Allerton Retreat Center, Monticello, IL, USA, Sep. 2010.

Y. Polyanskiy and S. Verdú, "Arimoto channel coding converse and Rényi divergence," 48th Allerton Conference 2010, Allerton Retreat Center, Monticello, IL, USA, Sep. 2010.

A. Ingber and M. Feder, "Finite blocklength coding for channels with side information at the receiver." Electrical and Electronics Engineers in Israel (IEEEI), 2010 IEEE 26th Convention of. IEEE, 2010.

S. Asoodeh, A. Lapidoth and L. Wang, "It takes half the energy of a photon to send one bit reliably on the Poisson channel with feedback." arXiv:1010.5382, 2010

T. J. Riedl, T. Coleman and A. Singer, "Finite block-length achievable rates for queuing timing channels." Information Theory Workshop (ITW), 2011.

A. Martinez and A. Guillén i Fàbregas, "Random-coding bounds for threshold decoders: Error exponent and saddlepoint approximation," in Proc. Int. Symp. Info. Theory, Aug. 2011.

References: Channel coding (2008-)

Y. Polyanskiy, H. V. Poor and S. Verdú, "Minimum energy to send k bits through the Gaussian channel with and without feedback," IEEE Trans. Inf. Theory, vol. 57, no. 8, pp. 4880 - 4902, Aug. 2011.

P.-N. Chen, H.-Y. Lin, and S. Moser, "Ultra-small block-codes for binary discrete memoryless channels," in Proc. IEEE Inf. Theory Works., Paraty, Brazil, Oct. 2011.

Y. Polyanskiy and S. Verdú, "Scalar coherent fading channel: dispersion analysis," 2011 IEEE Int. Symp. Inf. Theory (ISIT), St. Petersburg, Russia, Aug. 2011.

Y. Polyanskiy, H. V. Poor and S. Verdú, "Feedback in the non-asymptotic regime," IEEE Trans. Inf. Theory, vol. 57, no. 8, pp. 4903 - 4925, Aug. 2011.

W. Yang, G. Durisi, T. Koch, Y. Polyanskiy, "Diversity versus channel knowledge at finite block-length," 2012 Inf. Theory Workshop (ITW), Lausanne, Switzerland, Sep 2012.

Y. Polyanskiy and S. Verdú, "Empirical distribution of good channel codes with non-vanishing error probability," IEEE Trans. Inf. Theory, submitted Dec. 2012.
http://people.lids.mit.edu/yp/homepage/data/optcodes_journal.pdf

Y. Altug and A.Wagner., "Moderate deviations in channel coding." arXiv:1208.1924, 2012

J. Hoydis, R. Couillet, P. Piantanida and M. Debbah, "A random matrix approach to the finite blocklength regime of MIMO fading channels." Proc 2012 IEEE ISIT, 2012.

P. Moulin, "The log-volume of optimal constant-composition codes for memoryless channels, within $O(1)$ bits." Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on. IEEE, 2012.

J. Scarlett and A. Martinez, "Mismatched Decoding: Finite-Length Bounds, Error Exponents and Approximations." arXiv preprint arXiv:1303.6166 (2013).

W. Matthews and S. Wehner, "Finite blocklength converse bounds for quantum channels" arXiv:1210.4722, 2012

M. Tomamichel, "A framework for non-asymptotic quantum information theory." arXiv:1203.2142, 2012

References: Channel coding (2008-)

A. Ingber, R. Zamir, M. Feder, "Finite-Dimensional Infinite Constellations," IEEE Trans. Inform. Theory, vol.59, no.3, pp.1630,1656, March 2013

M. Tomamichel and V. Tan, "ϵ-Capacities and Second-Order Coding Rates for Channels with General State." arXiv:1305.6789, 2013

M. Naghshvar and T. Javidi, "Active sequential hypothesis testing." arXiv:1203.4626, 2012

A. Williamson, T.-Y. Chen and R. D. Wesel, "A rate-compatible sphere-packing analysis of feedback coding with limited retransmissions." Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on, 2012.

Y. Polyanskiy, "Saddle point in the minimax converse for channel coding," IEEE Trans. Inf. Theory, vol. 59, no. 5, pp. 2576-2595, May 2013.

Y. Polyanskiy and Y. Wu, "Peak-to-average power ratio of good codes for Gaussian channel", arXiv:1302.0084, Jan 2013.

References: Channel coding (multi-terminal)

T. S. Han, Information-Spectrum Methods in Information Theory. Springer Berlin Heidelberg, Feb 2003.

Y. Huang and P. Moulin, “Finite blocklength coding for multiple access channels,” in Proc. IEEE Intl. Symp. Inform. Theory (ISIT), Jul 2012, pp. 831-835.

E. MolavianJazi and J. N. Laneman, “Simpler achievable rate regions for multiaccess with finite blocklength,” Proc. 2012 IEEE Int. Symp. Information Theory, July 2012.

V. Tan and O. Kosut, “On the dispersions of three network information theory problems.” Information Sciences and Systems (CISS), 2012 46th Annual Conference on. IEEE, 2012.

P. Moulin, “A new metaconverse and outer region for finite-blocklength MACs.” Information Theory and Applications Workshop (ITA), 2013. IEEE, 2013.

S. Verdú, “Non-asymptotic achievability bounds in multiuser information theory,” in Allerton Conference, 2012.

V. Tan, “On Dispersions of Discrete Memoryless Channels with Noncausal State Information at the Encoder.” arXiv:1204.0431(2012).

V. Tan, “The Capacity of the General Gelfand-Pinsker Channel and Achievable Second-Order Coding Rates.” arXiv:1210.1091(2012).

M. H. Yassaee, M. R. Aref, and A. Gohari, “A Technique for Deriving One-Shot Achievability Results in Network Information Theory.” arXiv preprint arXiv:1303.0696 (2013).

References: Lossy compression, joint source-channel coding

I. Kontoyiannis, "Pointwise redundancy in lossy data compression and universal lossy data compression," IEEE Trans. Inform. Theory, vol. 46, no. 1, pp. 136-152, Jan. 2000

A. Ingber and Y. Kochman, "The dispersion of lossy source coding," in Data Compression Conference (DCC), Snowbird, UT, Mar. 2011, pp. 53-62.

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

A. Ingber, D Wang and Y. Kochman, "Dispersion theorems via second order analysis of functions of distributions." Information Sciences and Systems (CISS), 2012 46th Annual Conference on. IEEE, 2012.

V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," IEEE Trans. Inform. Theory, vol. 58, no. 6, Jun. 2012.

A. Tauste Campo, G. Vazquez-Vilar, A. Guillén i Fàbregas and A. Martinez, "Converse bounds for finite-length joint source-channel coding," in Proc. 50th Allerton Conf. on Comm., Cont. and Comp., Monticello, IL, USA, Oct. 2012.

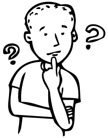
D. Wang, A. Ingber and Y. Kochman, "A strong converse for joint source-channel coding." Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on. IEEE, 2012.

V. Kostina and S. Verdú, "To code or not to code: Revisited." Information Theory Workshop (ITW), 2012 IEEE. IEEE, 2012.

V. Kostina and S. Verdú, "Lossy Joint Source-Channel Coding in the Finite Blocklength Regime", IEEE Trans. on Infor. Theory, vol. 59, no. 5, pp. 2545-2575, May 2013

N. Datta, J. M. Renes, R. Renner and M. M. Wilde, "One-shot lossy quantum data compression." arXiv:1304.2336, 2013

Thank you!



Do not hesitate to ask questions!

Yury Polyanskiy <yp@mit.edu>

Sergio Verdú <verdu@princeton.edu>