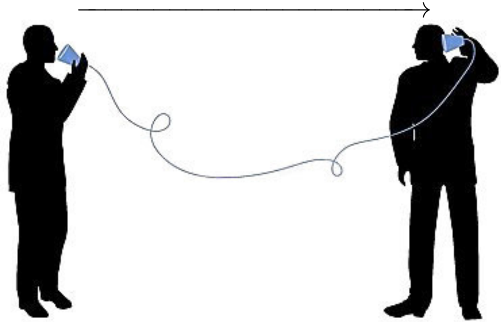
Finite blocklength IT: tutorial overview

Plan:

1. Motivation
2. Data compression (lossless)
3. Channel coding
   3.1 Overview on the example of BSC
   3.2 Converse bounds
   3.3 Achievability bounds
   3.4 Channel dispersion
   3.5 Discussion: ARQ, real codes
4. Anomalous dispersion
   4.1 Channels with feedback
   4.2 Quasi-static channels
   4.3 Energy-per-bit without CSI at receiver

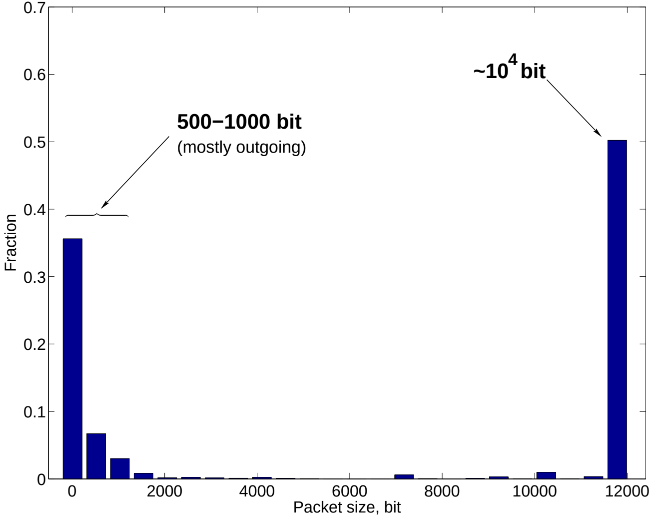
Voice communication

Tolerable delay: 50-200 ms

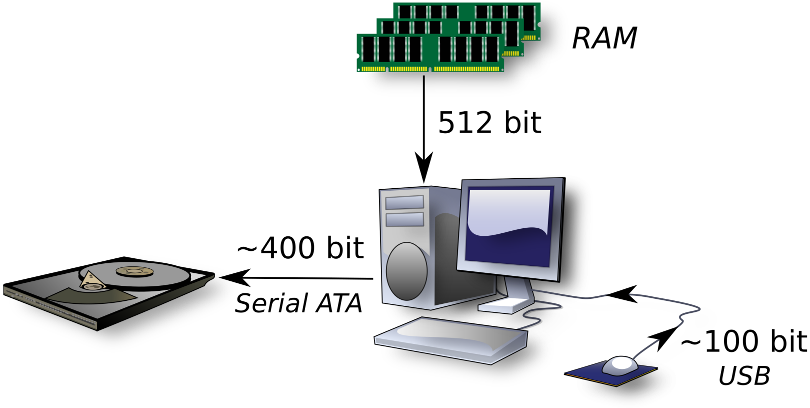


- Packetization: 20-30 ms
  $\Longrightarrow$ payload size $\sim$ 1500 bit (64 kbps, PCM)
- *Skype*: 1300 bit (avg.)
  GSM: 260 bit (uncoded), 456 bit (coded).

Internet traffic: packet sizes



Intra-PC communication



how many bits do we need to compress a 140-character twitter message?

Finite sample size problems lack theoretical simplicity but not practical applications.

T. Cover and M. Hellman

Shannon, 1953

- 173 -

Dr. C.E. Shannon. (In reply).

M. Indjoudjian has raised the question of what might be called a finite delay theory of information. Such a theory would indeed be of great practical importance, but the mathematical difficulties are quite formidable. The class of coding operations with a delay $\leq T$ is not closed in the mathematical sense, for if two such operations or transducers are used in sequence the overall delay may be as much as 2T. Thus we lose the important group theoretic property of closure which is so useful in the "infinite delay" and "time-preserving" theories.

Case study: 1000-bit BSC

- Consider channel $BSC(n = 1000, \delta = 0.11)$
- How many data bits can we transmit with (block) $P_e \leq 10^{-3}$?
- Attempt 1: Repetition

$$k = 47 \text{ bits via [21,1,21]-code}$$

- Attempt 2: Reed-Muller

$$k = 112 \text{ bits via [64,7,32]-code}$$

- Shannon's prediction: $C = 0.5$ bit so

$$k \approx 500 \text{ bit}$$

Case study: 1000-bit BSC

- Consider channel $BSC(n = 1000, \delta = 0.11)$
- How many data bits can we transmit with (block) $P_e \leq 10^{-3}$?
- Attempt 1: Repetition

$$k = 47 \text{ bits via } [21,1,21]\text{-code}$$

- Attempt 2: Reed-Muller

$$k = 112 \text{ bits via } [64,7,32]\text{-code}$$

- Shannon's prediction: $C = 0.5$ bit so

$$k \approx 500 \text{ bit}$$

- Finite blocklength IT:

$$414 \leq k \leq 416$$

Lossless data compression

Compression: Setup

- Given discrete r.v. $X \sim P_X$, find lossless $f : \mathcal{X} \to \{0,1\}^*$ s.t.

$$\ell(f(X)) \to \min$$

  Here: $\ell(\text{binary string}) \triangleq \text{length}$
- Rename values of X: $P_X(1) \geq P_X(2) \geq \cdots$
- The best compressor (for $\mathbb{E}$, median, etc)

$$\begin{aligned} f^*(1) &= \emptyset \\ f^*(2) &= 0 \\ f^*(3) &= 1 \\ f^*(4) &= 01 \\ &\cdots \end{aligned}$$

- Analysis of $\mathbb{E}[\ell(f^*(X))]$: cumbersome!

Compression: Results

- Bounds on $L^* = \ell(f^*(X))$
  - Distribution:

$$\mathbb{P}\left[\log \frac{1}{P_X(X)} \le k\right] \le \mathbb{P}[L^* \le k] \le \mathbb{P}\left[\log \frac{1}{P_X(X)} \le k + \tau\right] + 2^{-\tau}$$

  - Expected length: [Alon-Orlitsky]

$$H(X) - \log_2(e(H(X)+1)) \le \mathbb{E}\left[L^*\right] \le H(X)$$

- WLLN: When $X = (A_1, \ldots, A_n)$ – iid (or ergodic!) vector

$$\frac{1}{n} L^* \approx H(A)$$

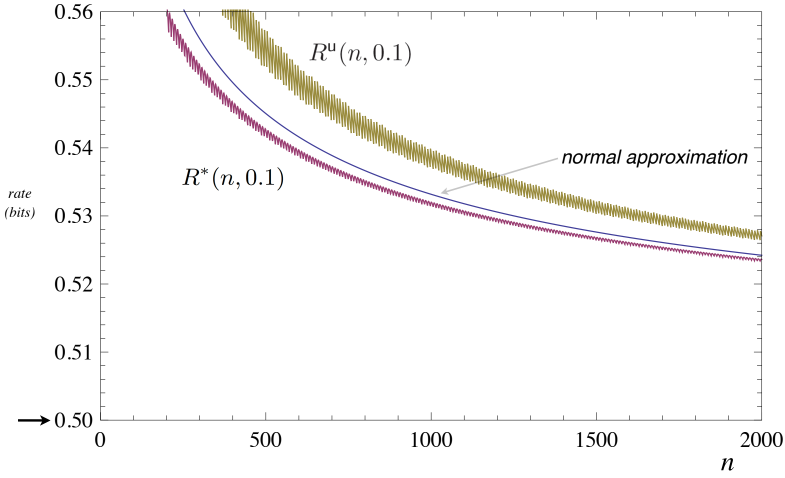
- CLT: When $X = (A_1, \ldots, A_n)$ – iid (or weakly dependent!) vector

$$\frac{1}{n} L^* \approx H(A) + \sqrt{\frac{V(A)}{n}} Z, \qquad Z \sim \mathcal{N}(0,1)$$

where $V(A) = \mathrm{Var}[\log P_X(X)]$ – varentropy.

Behavior of $R^*(n, \epsilon)$ – the ϵ-quantile of compression length

biased coin flips: entropy = 0.5



Takeaway: $\ell(f^*(X)) \approx \log \frac{1}{P_X(X)}$

Takeaway: $\ell(f^*(X)) \approx \log \frac{1}{P_X(X)}$

Channel coding: (next two hours)

- Information density $i(X;Y)$ takes the role of $\log \frac{1}{P_X(X)}$
- Optimal transmission rate

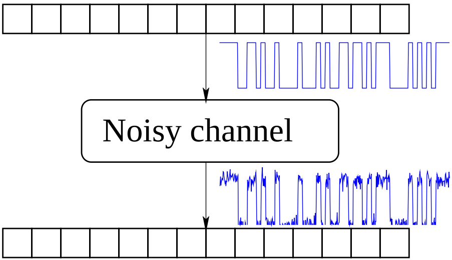
$$R^*(n,\epsilon) \quad \approx \quad \epsilon\text{-quantile of } \frac{1}{n} i(X^n;Y^n)$$

- iid/ergodic asymptotics:

$$R^*(n,\epsilon) \quad \to \quad C = \max_X I(X;Y)$$

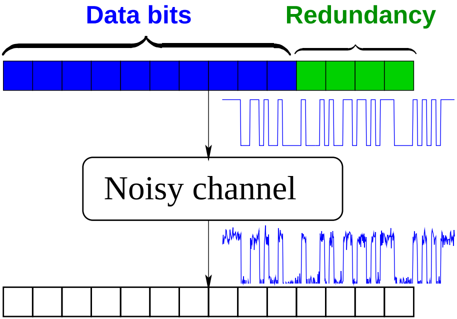
Channel coding: intro

Abstract communication problem



Goal: Decrease corruption of data caused by noise

Channel coding: principles

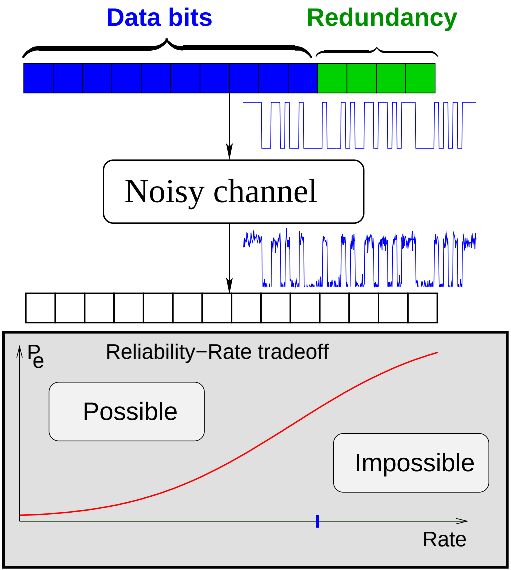


Goal: Decrease corruption of data caused by noise

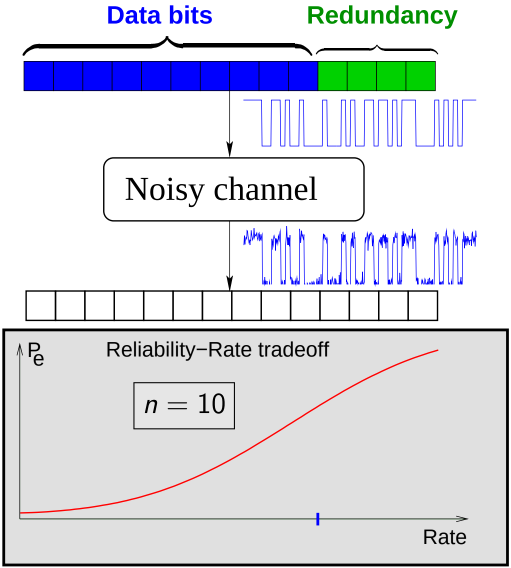
Solution: *Code* to diminish probability of error P_e.

Key metrics: Rate and P_e

Channel coding: principles



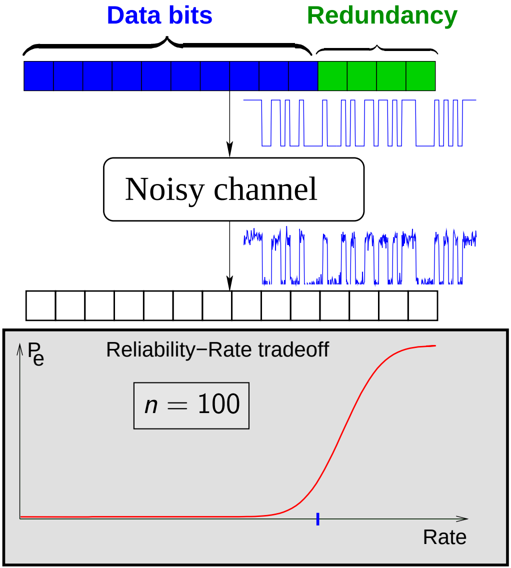
Channel coding: principles



Decreasing P_e further:

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

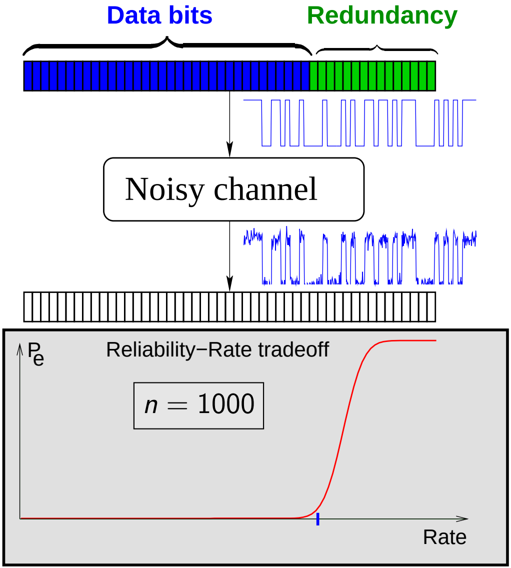
Channel coding: principles



Decreasing P_e further:

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

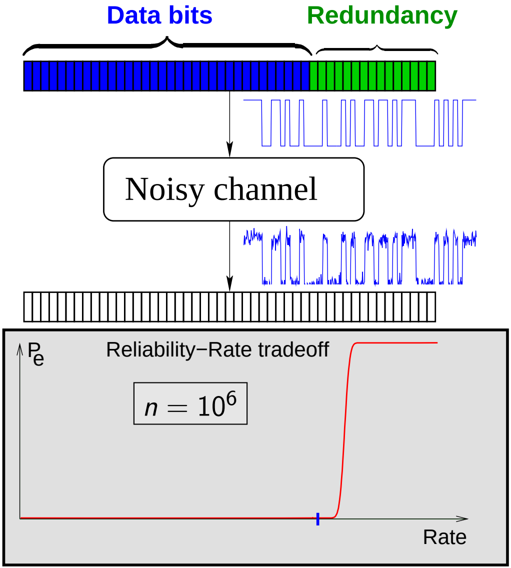
Channel coding: principles



Decreasing P_e further:

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

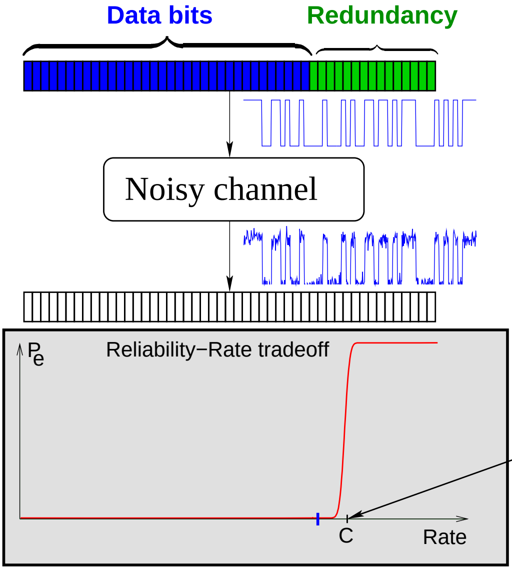
Channel coding: principles



Decreasing P_e further:

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

Channel coding: Shannon capacity

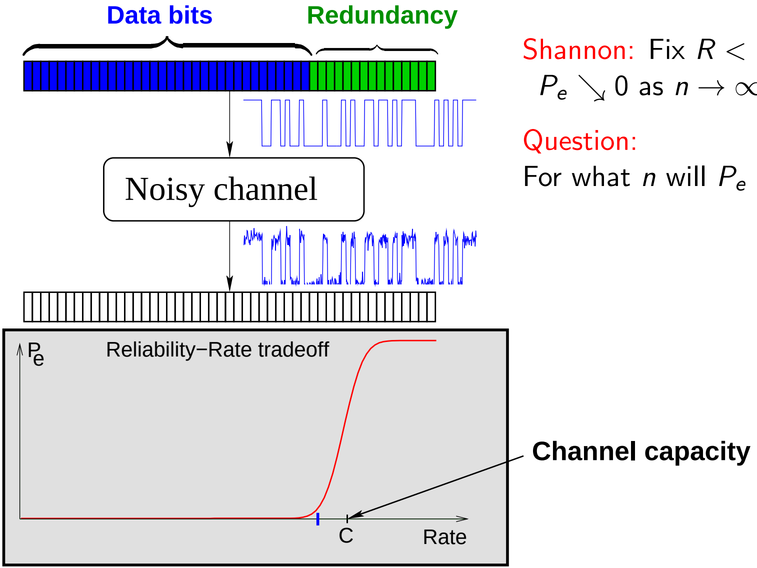


Shannon: Fix $R < C$

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

Channel capacity

Channel coding: Shannon capacity



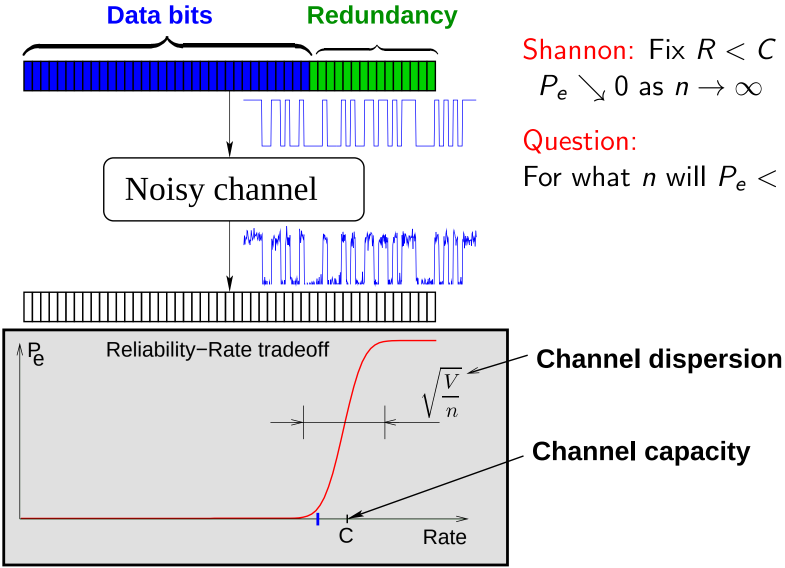
Shannon: Fix $R < C$

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

Question:

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

Channel coding: Gaussian approximation



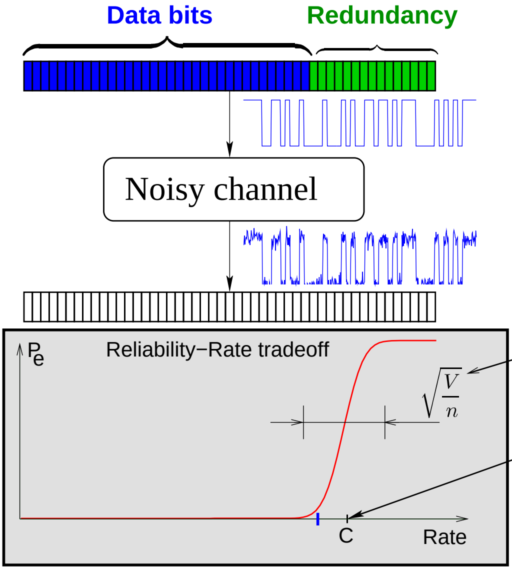
Shannon: Fix $R < C$

$P_e \searrow 0$ 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 $P_e < 10^{-3}$?

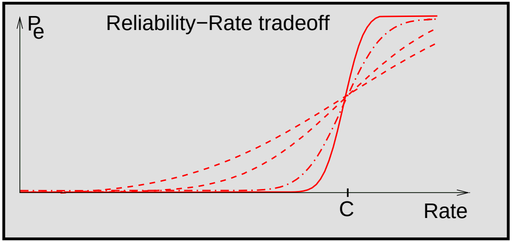
Answer:

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

Channel dispersion

Channel capacity

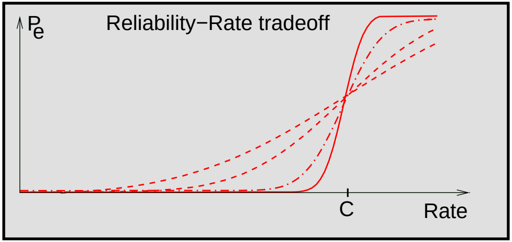
How to describe evolution of the boundary?



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

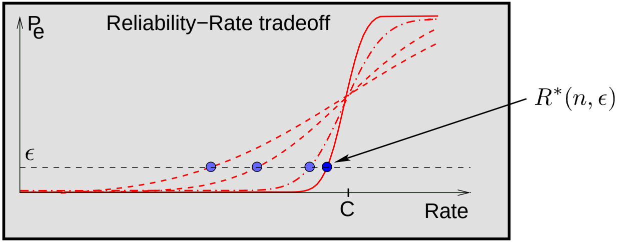
How to describe evolution of the boundary?



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

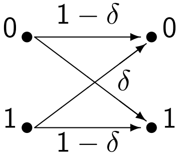
$$\lim_{n\to\infty} R^*(n,\epsilon) = C \quad \Longleftrightarrow \begin{cases} R^*(n,\epsilon) \le C + o(1)\,, \\ R^*(n,\epsilon) \ge C + o(1)\,. \end{cases}$$

  (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:

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

- Solution: Derive bounds from scratch.

New achievability bound



BSC

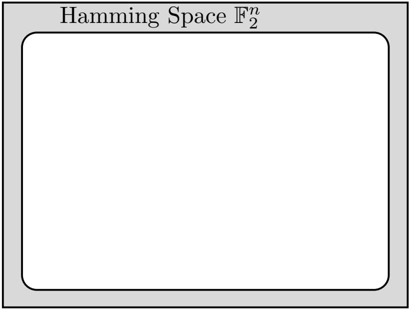
Theorem (RCU)

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

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

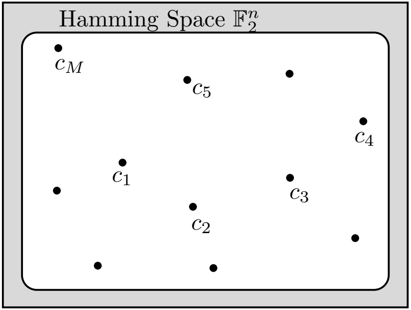
Proof of RCU bound for the BSC

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



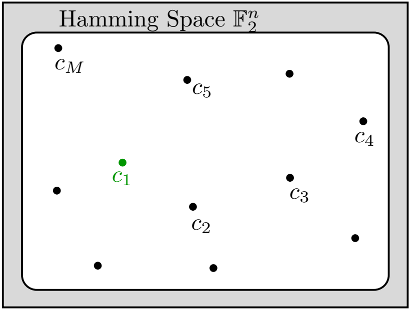
Proof of RCU bound for the BSC

- 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

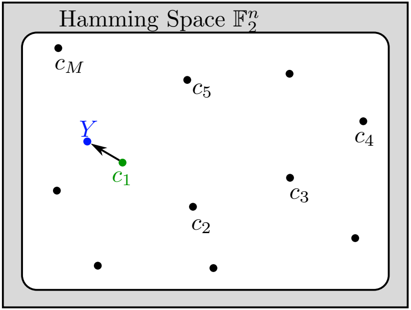
- Input space: $A = \{0,1\}^n$
- Let $c_1, \ldots, c_M \sim Bern(\frac{1}{2})^n$ (random codebook)
- Transmit c_1



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 $c_1 \to Y$

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

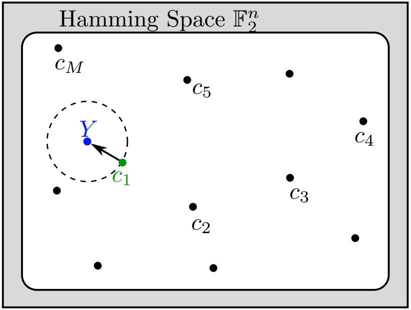


Proof of RCU bound for the BSC

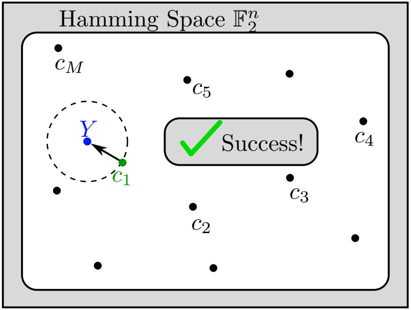
- Input space: $A = \{0,1\}^n$
- Let $c_1, \dots, c_M \sim Bern(\frac{1}{2})^n$ (random codebook)
- Transmit c_1
- Noise displaces $c_1 \to Y$

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

- Decoder: find closest codeword to Y



Proof of RCU bound for the BSC



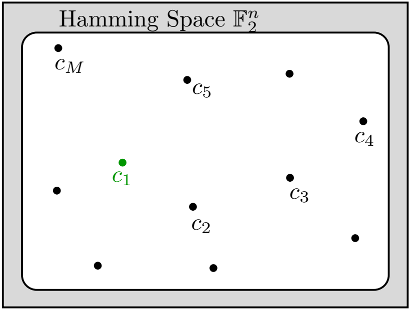
- Input space: $A = \{0,1\}^n$
- Let $c_1, \ldots, c_M \sim Bern(\frac{1}{2})^n$ (random codebook)
- Transmit c_1
- Noise displaces $c_1 \to Y$

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

- Decoder: find closest codeword to Y

Proof of RCU bound for the BSC

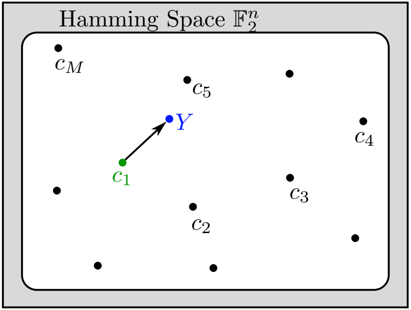
- Input space: $A = \{0, 1\}^n$
- Let $c_1, \ldots, c_M \sim Bern(\frac{1}{2})^n$ (random codebook)
- Transmit c_1



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 $c_1 \to Y$

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

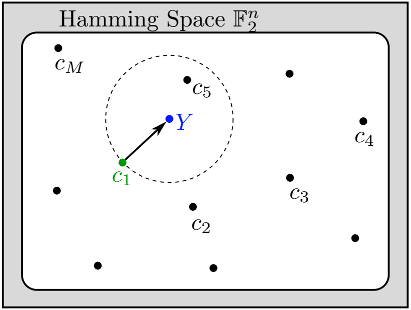


Proof of RCU bound for the BSC

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

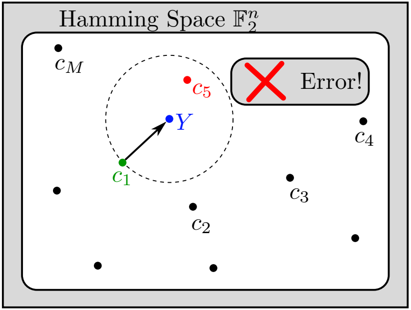


Proof of RCU bound for the BSC

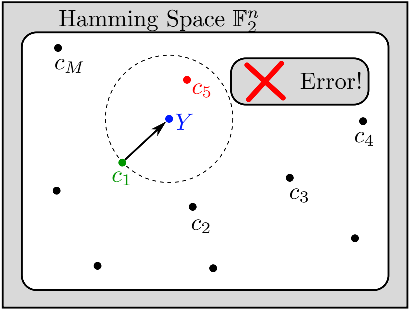
- Input space: $A = \{0,1\}^n$
- Let $c_1, \ldots, c_M \sim Bern(\frac{1}{2})^n$ (random codebook)
- Transmit c_1
- Noise displaces $c_1 \to Y$

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

- Decoder: find closest codeword to Y



Proof of RCU bound for the BSC



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

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

...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 $\mathrm{wt}(Z) \sim \mathit{Binomial}(n, \delta) \implies$ **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:

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

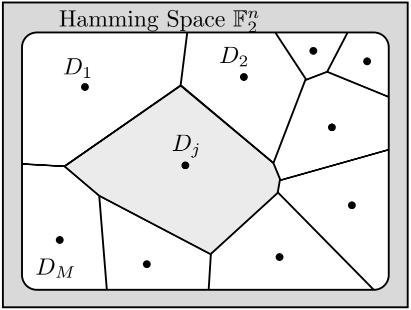
Note: Convexity of f follows from general properties of β_α (below)

Sphere-packing converse (BSC variation)

Proof:

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

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

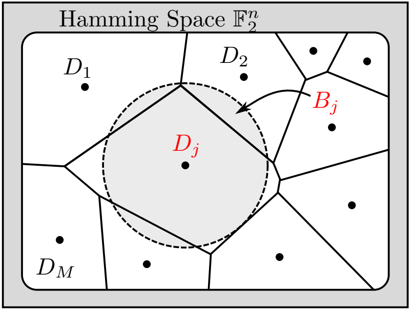


Sphere-packing converse (BSC variation)

Proof:

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

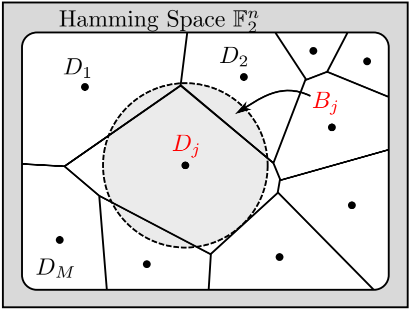
$$\epsilon = \frac{1}{M}\sum_j \mathbb{P}[c_j + Z \notin D_j]$$
$$\geq \frac{1}{M}\sum_j \mathbb{P}[Z \notin B_j]$$



- 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:

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

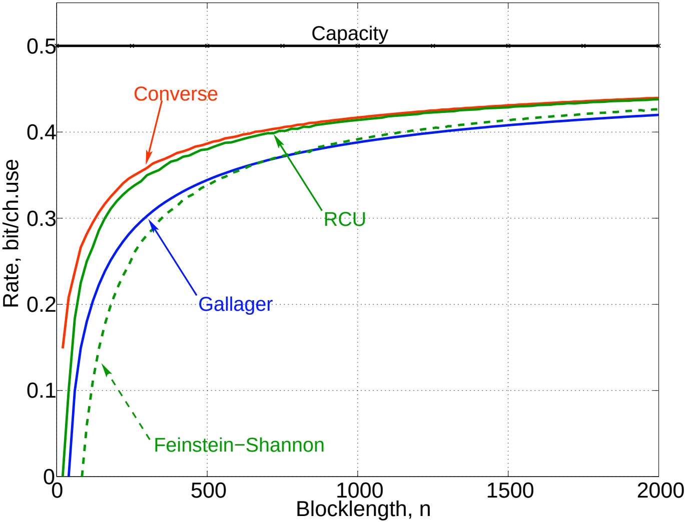
- B_j = ball centered at 0 s.t. $\mathrm{Vol}(B_j) = \mathrm{Vol}(D_j)$
- Simple calculation:

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

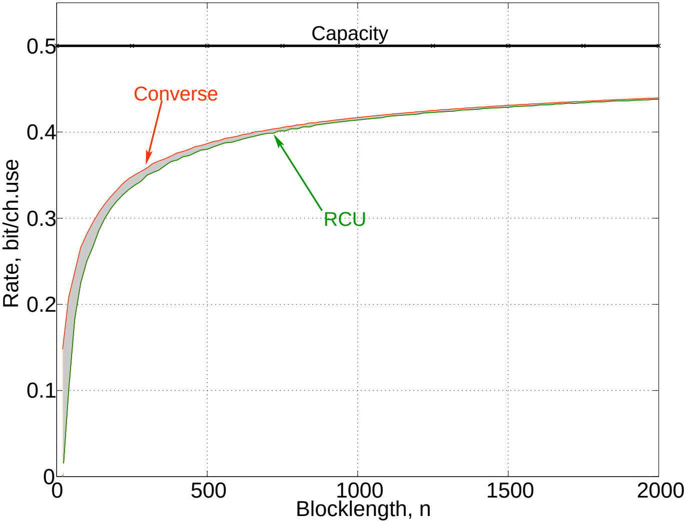
- f – convex, apply Jensen:

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

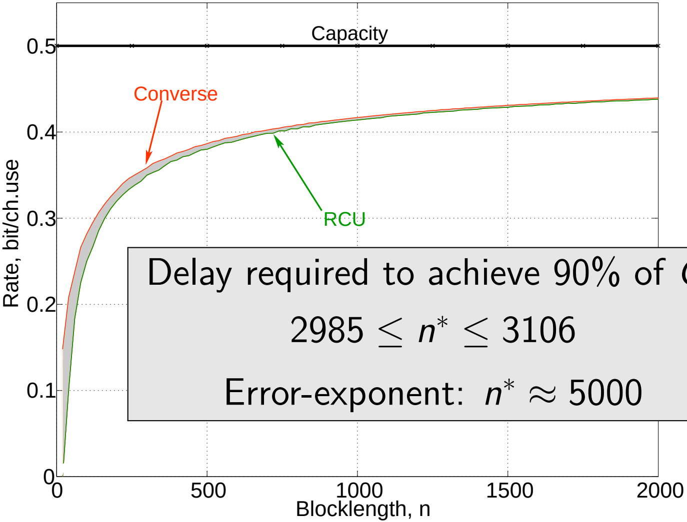
Bounds: example $BSC(0.11)$, $\epsilon = 10^{-3}$



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* $0 < \epsilon < 1$,

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

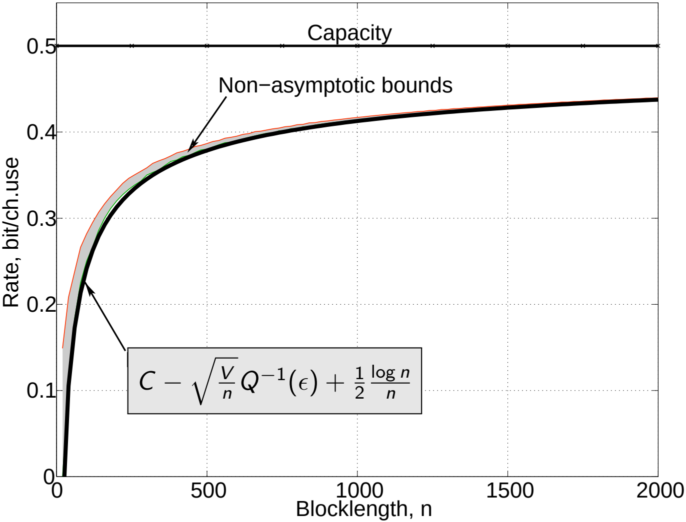
where

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

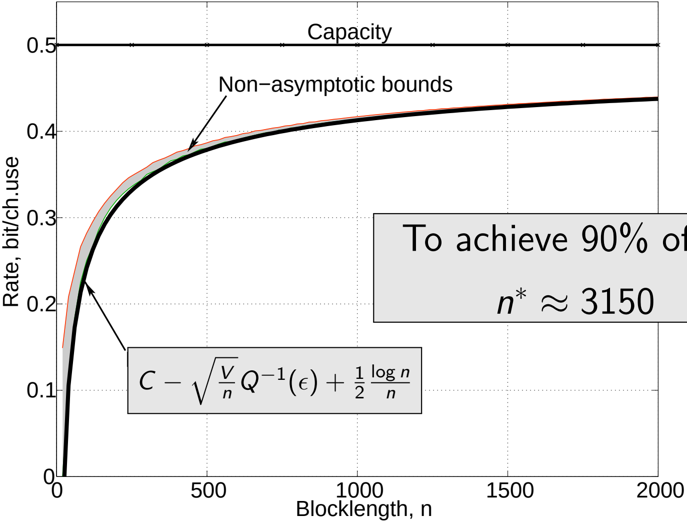
Proof: Bounds + Stirling's formula

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:

$$n^* \approx 3150$$

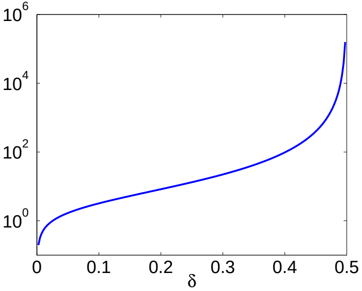
Dispersion and minimal required delay

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:

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

- Channel dispersion:

$$n^* \approx 3150$$

Channel coding: converse bounds

Notation

- Take a random transformation $\mathrm{A} \xrightarrow{P_{Y|X}} \mathrm{B}$
  (think $\mathrm{A} = \mathcal{A}^n$, $\mathrm{B} = \mathcal{B}^n$, $P_{Y|X} = P_{Y^n|X^n}$)
- 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 $W \sim \mathit{Unif}\{1, \ldots, M\}$*:*

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

Implies *weak converse*:

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

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 $(W, X, Y, \hat{W})$:

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

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

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

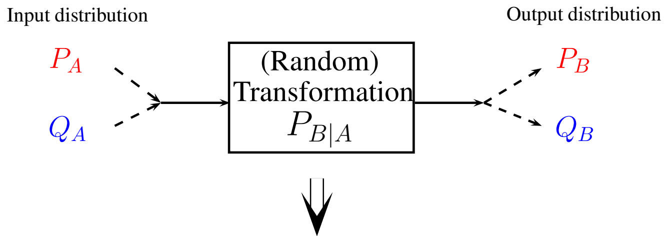
$$\text{DAG:} \quad W \to X \;\underline{\quad\quad}\; Y \to \hat{W}$$

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

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

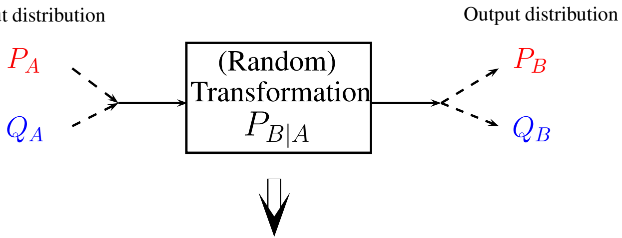
Next step: data-processing for relative entropy $D(\cdot||\cdot)$

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



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

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:

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

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:

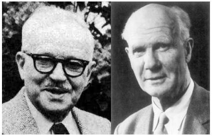
$$(1-\epsilon)\log M \le \sup_{P_X} I(X;Y) + h(\epsilon)$$

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?

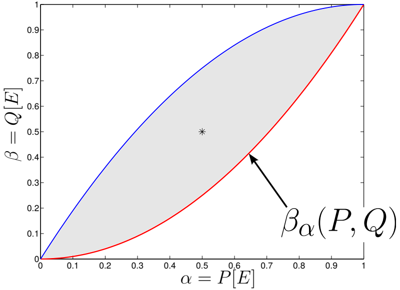
Neyman-Pearson's β_α

Definition

For every pair of measures P, Q

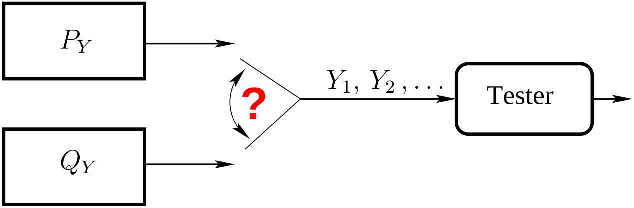
$$\beta_\alpha(P, Q) \triangleq \inf_{E: P[E] \geq \alpha} Q[E] .$$

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:

$$\mathbb{P}[\text{Tester says "}Q_Y\text{"}] \leq \epsilon$$
$$\mathbb{Q}[\text{Tester says "}P_Y\text{"}] \to \min$$

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

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

Meta-converse: minimax version

Theorem

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

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

- 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*

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

Bound is tight in several senses:

- There exist *non-signalling assisted* (NSA) codes attaining the upper-bound. [Matthews, T-IT'2012]
- For any (M, ϵ)-code with ML decoder [Vazquez-Vilar *et al*'13]

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

- **ISIT'2016:** almost tight for regular codes [Barman-Fawzi]

Meta-converse: minimax version

Theorem

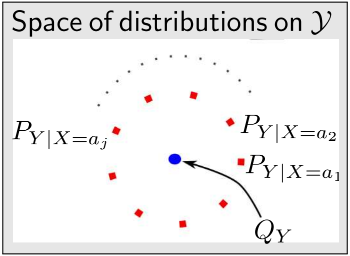
Every (M, ϵ)*-code for channel* $P_{Y|X}$ *satisfies*

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

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. ⟶
  good $Q_Y \approx$ "center" of $P_{Y|X}$
- *Exercise:* Redo BSC.



Example: Converse for AWGN

The AWGN Channel

$$Z \sim \mathcal{N}(0, \sigma^2)$$

$$\downarrow$$

$$X \longrightarrow \bigoplus \longrightarrow Y$$

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 $\mathbb{P}$: $Y^n = X^n + \mathcal{N}(0, \mathbf{I}_n)$
  By CLT distribution of LLR (even $|X^n = x^n$!)

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

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

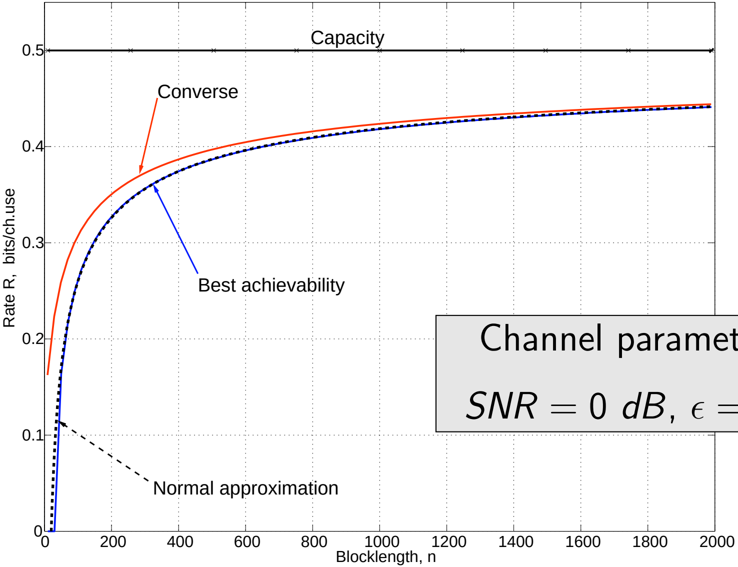
$$\log \beta_{1-\epsilon}(P_{Y^n|X^n=c}, Q_{Y^n}) \approx -nC + \sqrt{nV}Q^{-1}(\epsilon)$$

- 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 $Q_Y = \mathcal{N}(0,1)^n$

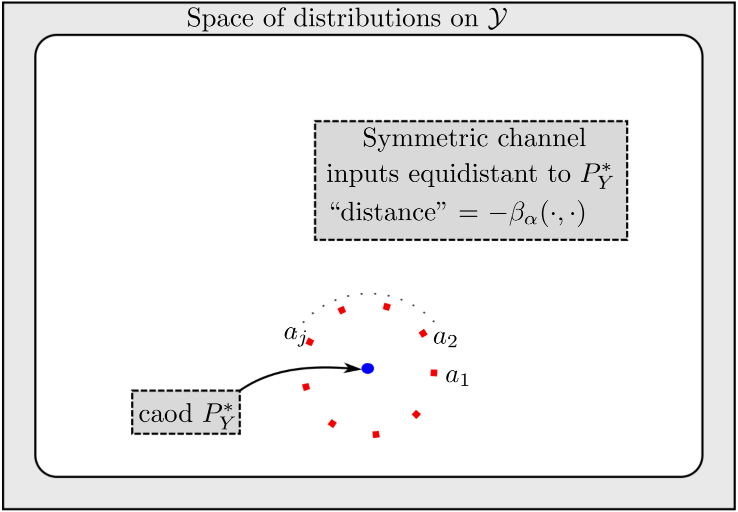


Channel parameters:

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

spectre

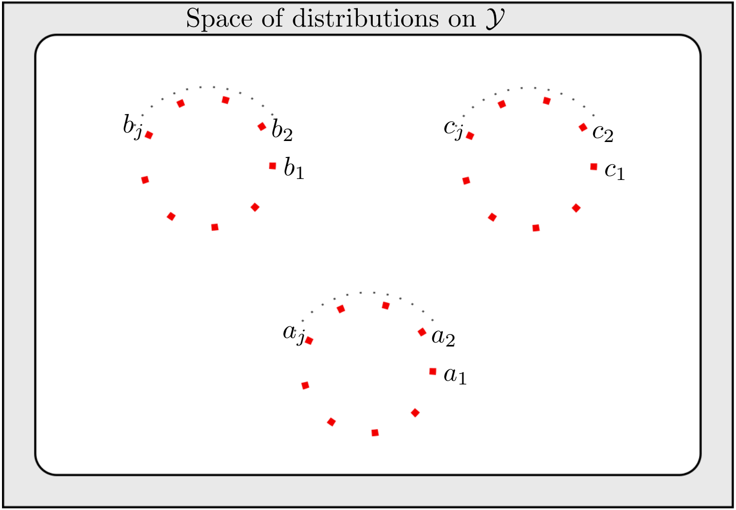
From one Q_Y to many



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

Symmetric channel: choice of Q_Y is clear

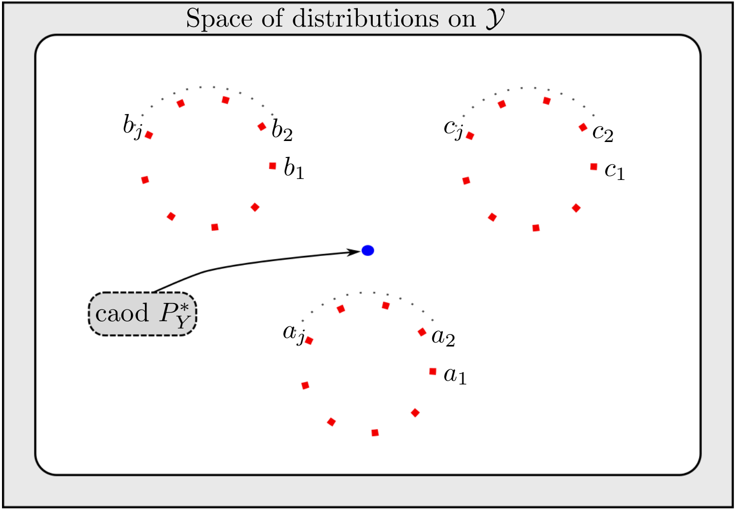
From one Q_Y to many



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

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

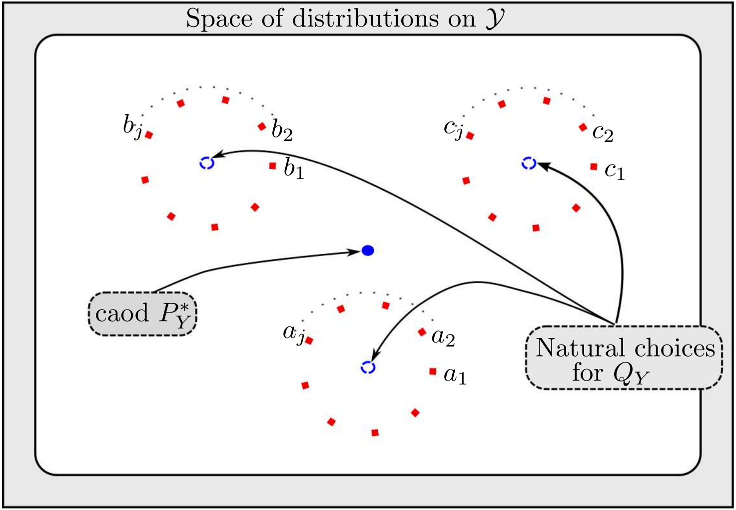
From one Q_Y to many



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 $\epsilon \triangleq \mathbb{P}[error]$ *and* $\epsilon' \triangleq \mathbb{Q}[error]$ *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
- Block fading: $Q_{Y^n|X^n=x^n} = f(\|x^n\|_4)$
  Why? To curb "flat" codewords with $\frac{\|x\|_4}{\|x\|_2} < 3$

Meta-converse generalizes many classical methods

Theorem

For any code $\epsilon \stackrel{\triangle}{=} \mathbb{P}[error]$ *and* $\epsilon' \stackrel{\triangle}{=} \mathbb{Q}[error]$ *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

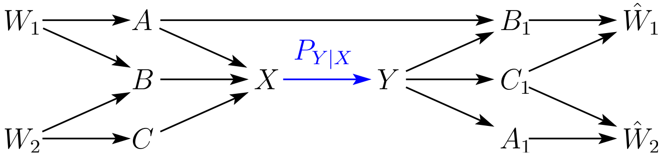
For any code $\epsilon \triangleq \mathbb{P}[error]$ *and* $\epsilon' \triangleq \mathbb{Q}[error]$ *satisfy*

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

Corollaries:

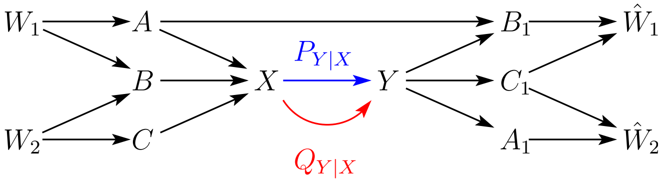
- Fano's inequality
- Wolfowitz strong converse
- Shannon-Gallager-Berlekamp's sphere-packing
  + improvements: [Valembois-Fossorier'04], [Wiechman-Sason'08]
- Haroutounian's sphere-packing
- list-decoding converses E.g.: $\mathbb{Q}[W \in \{\text{list}\}] = \frac{|\{\text{list}\}|}{M}$
- Berlekamp's 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:

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

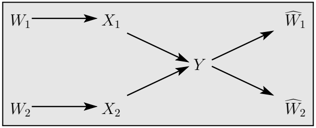
- Same idea: use code as a suboptimal binary HT: $P_{Y|X}$ vs. $Q_{Y|X}$
- ... and compare to the best possible test:

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

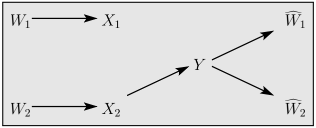
$$\beta_{1-\epsilon}(P_X P_{Y|X}, P_X Q_{Y|X}) \leq 1-\epsilon'$$

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 $D(\cdot||\cdot)$ ...

$$\Downarrow$$

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

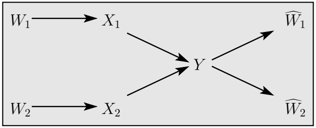
Optimizing $Q_{Y|X_1}$:

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

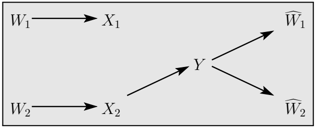
Also with $X_1 \leftrightarrow X_2 \implies$ weak converse (usual pentagon)

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]

Channel coding: achievability bounds

Notation

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

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

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

- 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: $\hat{W} = \operatorname{argmax}_j \imath_{X;Y}(C_j;Y)$
- 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:

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

In ⓐ: C_j too far from Y
In ⓑ: 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

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

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

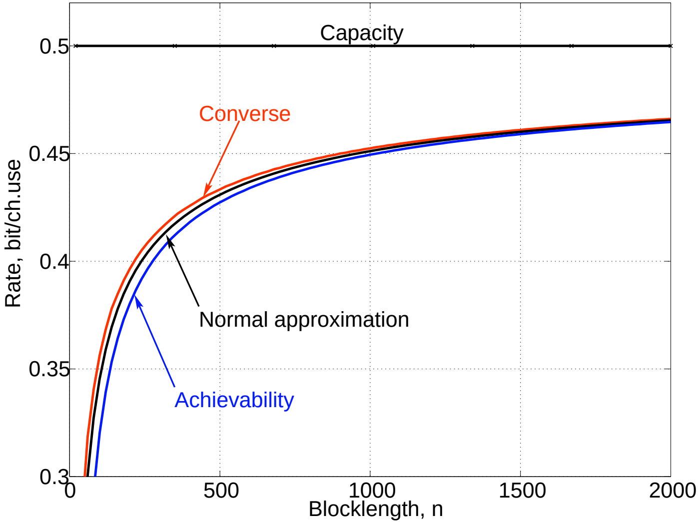
$$\frac{M+1}{2}\left(\frac{2}{M+1}P_{XY}\left[\frac{dP_{XY}}{dP_{\bar{X}Y}} \le e^{\gamma}\right] + \frac{M-1}{M+1}P_{\bar{X}Y}\left[\frac{dP_{XY}}{dP_{\bar{X}Y}} > e^{\gamma}\right]\right)$$

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

- Change of measure argument:

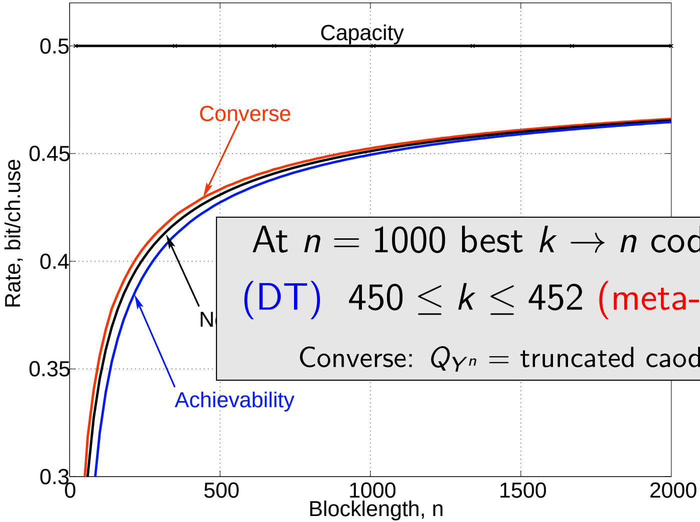
$$P\left[\frac{dP}{dQ} \le \tau\right] + \tau Q\left[\frac{dP}{dQ} > \tau\right] = \mathbb{E}_P\left[\exp\left\{-\left|\log \frac{dP}{dQ} - \log \tau\right|^+\right\}\right] .$$

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



spectre

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



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

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

Converse: Q_{Y^n} = truncated caod

spectre

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.

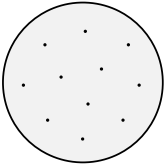
$\kappa\beta$ bound: idea

Decoder:

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

  Run optimal binary HT for :

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



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

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

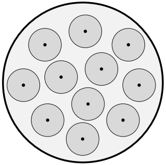
$\kappa\beta$ bound: idea

Decoder:

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

  Run optimal binary HT for :

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



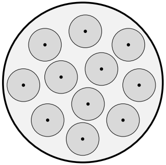
$$\begin{aligned} \mathbb{P}[\text{detect } \mathcal{H}_0] &= 1 - \epsilon + \tau \\ \mathbb{Q}[\text{detect } \mathcal{H}_0] &= \beta_{1-\epsilon+\tau}(P_{Y|X=x}, Q_Y) \end{aligned}$$

- 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 $\implies$ $\bigcup_{j=1}^{M} \{j\text{-th decoding region}\}$ is a composite HT:



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

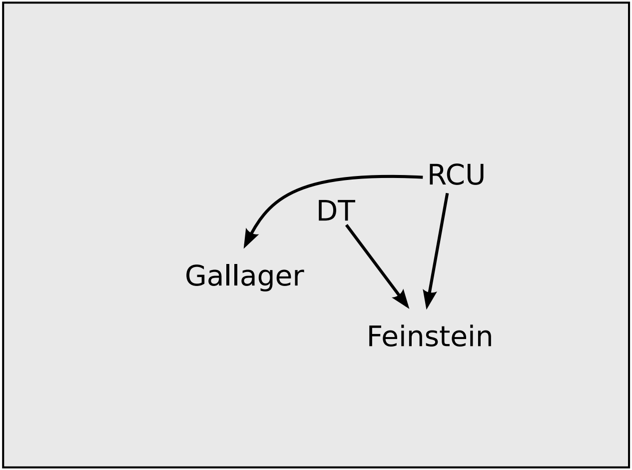
- Performance of the best test:

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

- Thus:

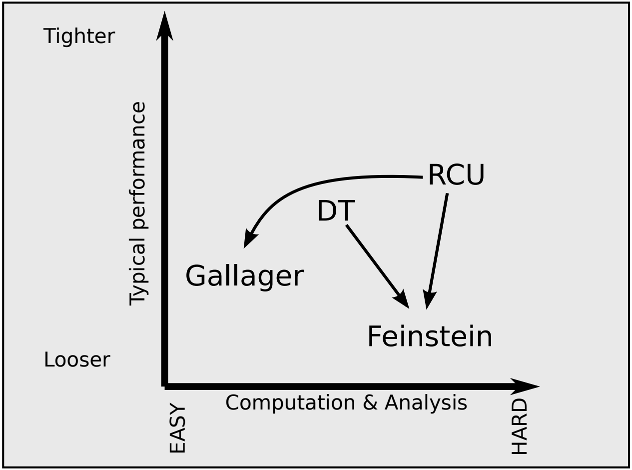
$$\begin{aligned} \kappa_\tau &\le \mathbb{Q}[\text{all } M \text{ "balls"}] \\ &\le M \sup_x \beta_{1-\epsilon+\tau}(P_{Y|X=x}, Q_Y) \end{aligned}$$

Hierarchy of achievability bounds (no cost constr.)



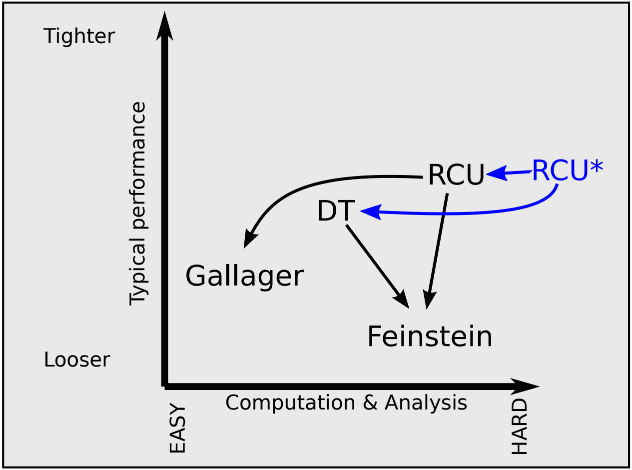
- Arrows show logical implication

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**.
- RCU* in [Haim-Kochman-Erez, ISIT'13]

Channel coding: 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) $\implies$ 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

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

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

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

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

where $LVar\{X_i\} \triangleq \lim \frac{1}{n} \mathrm{Var}[\sum_{i=1}^n X_i]$.

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

$$V = \mathrm{LVar}\{C(S_i)\} + \mathbb{E}[V(S)]$$

- Quasi-static fading channels [Yang-Durisi-Koch-P.'14]:

$$V = 0 \text{ (!)}$$

Dispersion: product vs generic channels

- Relation to alphabet size:

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

- Dispersion is additive:

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

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

- $\Longrightarrow$ 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):

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

- For **all DMC** [Moulin'12, Tomamichel-Tan'13]

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

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:

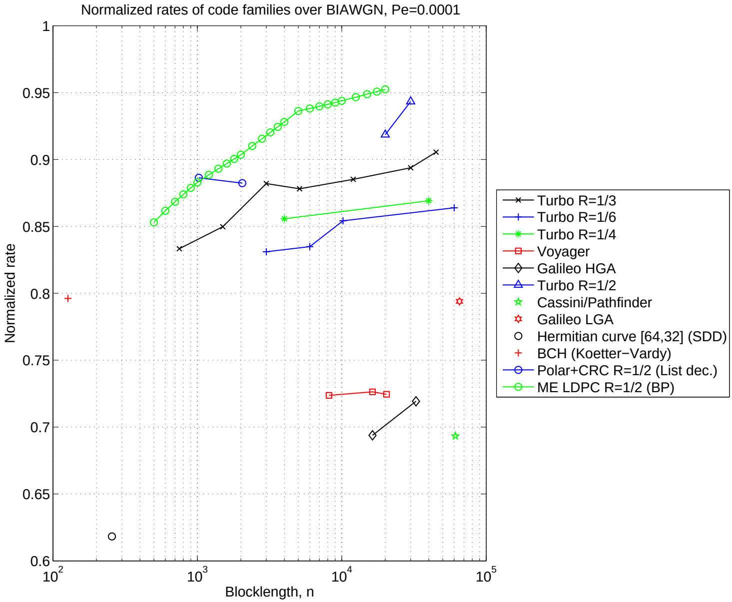
Definition (Normalized rate)

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

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

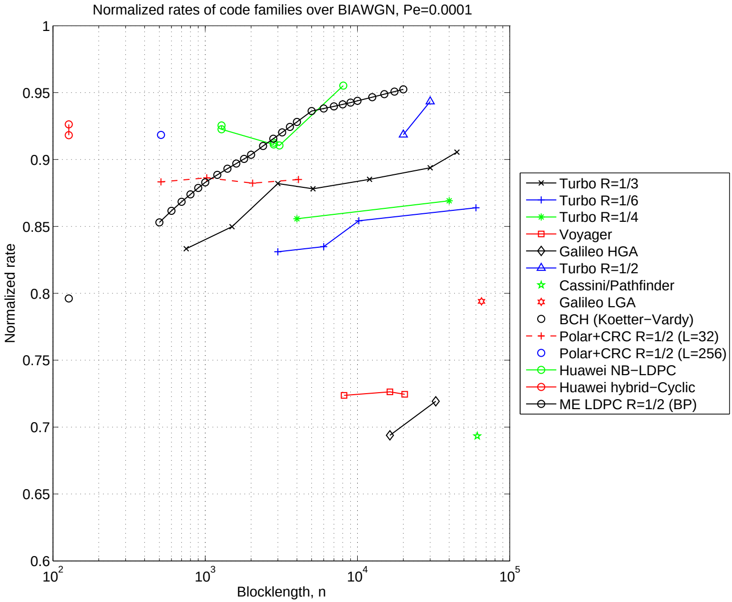
- Agreement: $\epsilon = 10^{-3}$ or 10^{-4}
- Take $R^*(n, \epsilon, SNR) \approx C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon)$
- A family of channels needed (e.g. AWGN or BSC)

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



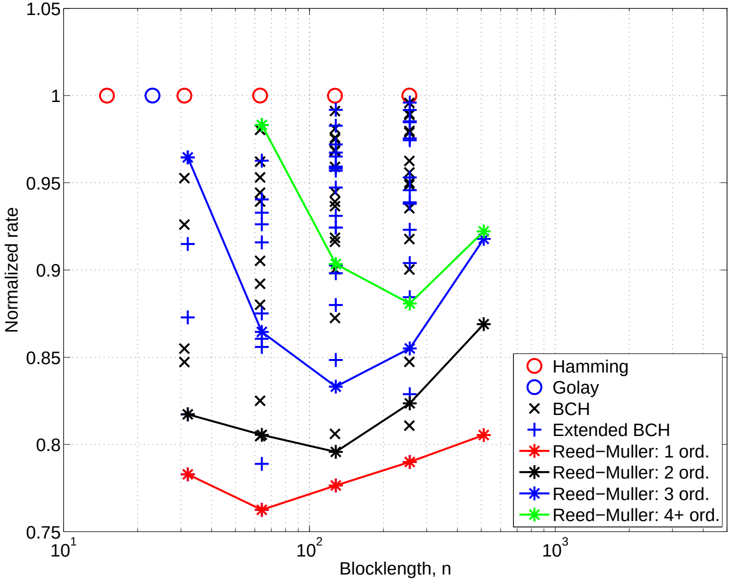
spectre

Codes vs. fundamental limits (from 1970's to 2016)



spectre

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?

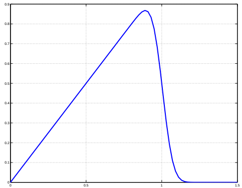
Optimizing ARQ systems

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

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

(almost) universal adaptation rule!

Anomalous dispersion

Recap: From $I(X;Y)$ to β_α

- Asymptotics (Shannon):

$$\lim_{n\to\infty} R^*(n,\epsilon) = C \stackrel{\triangle}{=} \max_{P_X} I(X;Y)\,.$$

- Non-asymptotic analog of C [PPV'10]:

$$R^*(n,\epsilon) = \max_{P_X} \qquad -\frac{1}{n}\log \beta_{1-\epsilon}(P_X P_{Y|X}, P_X Q_Y) + O\left(\frac{\log n}{n}\right)$$

Recap: From $I(X;Y)$ to β_α

- Asymptotics (Shannon):

$$\lim_{n\to\infty} R^*(n,\epsilon) = C \stackrel{\triangle}{=} \max_{P_X} I(X;Y)\,.$$

- Non-asymptotic analog of C [PPV'10]:

$$R^*(n,\epsilon) = \max_{P_X} \min_{Q_Y} -\frac{1}{n}\log \beta_{1-\epsilon}(P_X P_{Y|X}, P_X Q_Y) + O\left(\frac{\log n}{n}\right)$$

Recap: From $I(X;Y)$ to β_α

- Asymptotics (Shannon):

$$\lim_{n\to\infty} R^*(n,\epsilon) = C \stackrel{\triangle}{=} \max_{P_X} I(X;Y)\,.$$

- Non-asymptotic analog of C [PPV'10]:

$$R^*(n,\epsilon) = \max_{P_X} \min_{Q_Y} -\frac{1}{n}\log \beta_{1-\epsilon}(P_X P_{Y|X}, P_X Q_Y) + O\left(\frac{\log n}{n}\right)$$

- By Neyman-Pearson:

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

- What is P_{X^n}, Q_{Y^n}?
  A: saddle-point of β_α (or approx.)

Channel dispersion

- Memoryless channels: saddle-point P_X, Q_Y are both memoryless (approx.)
- $\Longrightarrow$ from CLT:

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

where

$$C = \mathbb{E}\left[\log\frac{P_{Y|X}(Y|X)}{Q_Y(Y)}\right] \qquad V = \mathrm{Var}\left[\log\frac{P_{Y|X}(Y|X)}{Q_Y(Y)}\right]$$

- Punchline:

  Channel dispersion = Variance of info density

Anomalous dispersion terms

- Feedback coding + variable-length [PPV'11]:

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

- Var-length data-compression [Kostina-P.-Verdú'14]:

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

- Long-term ("average-over-the-codebook") power-constraint [Yang-Caire-Durisi-P.'15]:

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

Anomalous dispersion terms (cont'd)

- Quasi-static fading channels [Yang-Durisi-Koch-P.'14]:

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

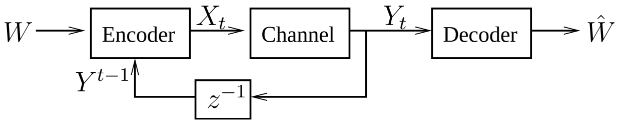
- Energy-per-bit under multiplicative fading [Yang-Durisi-P.'15]:

$$E_b(k,\epsilon) \approx (-1.59\ dB) + c_\epsilon \sqrt[3]{\frac{\log k}{k}}$$

Contrast with CLT-like AWGN: $E_b(k,\epsilon) \approx (-1.59\ dB) + \sqrt{\frac{V}{k}} Q^{-1}(\epsilon)$

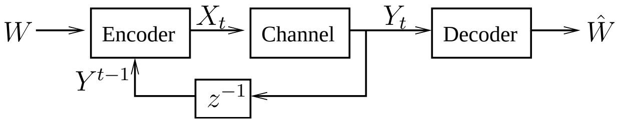
Channels with feedback

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)

$$\epsilon \leq \mathbb{E}[f(\tau)],$$

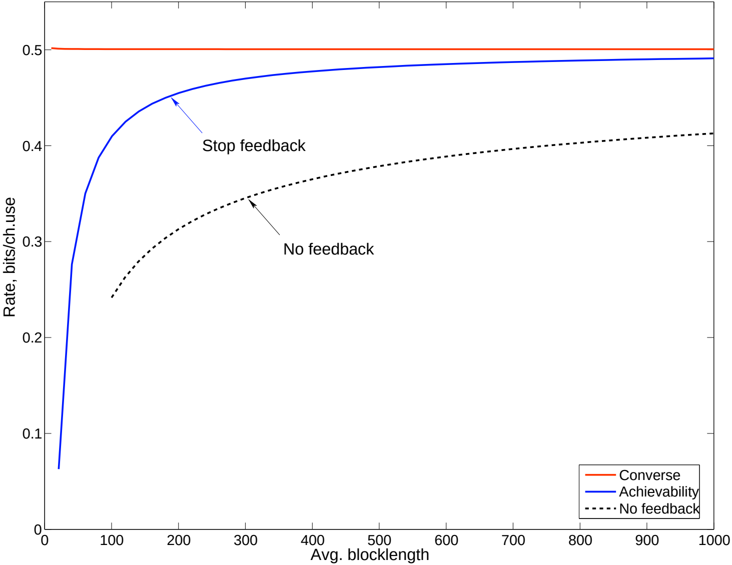
where

$$f(n) \triangleq \mathbb{E}\left[1\{\tau \leq n\} 2^{\ell R - S_\tau}\right]$$

$$\tau \triangleq \inf\{n \geq 0 : i(X^n; Y^n) \geq \gamma\}$$

$$i(X^n; Y^n) \triangleq \text{sum of iid Bernoulli}(\delta).$$

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



Why dispersion is zero?

- Channel coding goal: have LLR $\geq \log M$ w.h.p.
- LLR $= S_n =$ sum-of-iid.
- Fixed blocklength: set n so large that

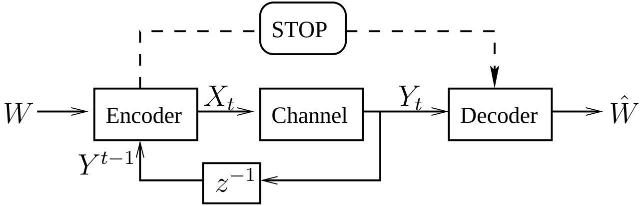
$$P_e \approx \mathbb{P}\left[S_n < \log M\right] .$$

  Thus get $\sqrt{n}$ fluctuations.
- Expected-length:

$$S_\tau = \log M \pm \mathrm{const}$$

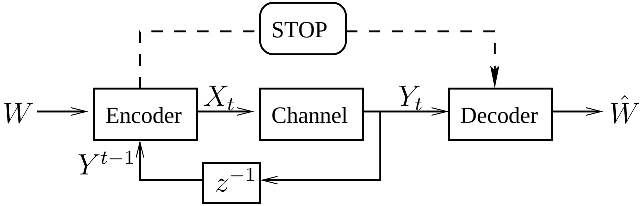
  Thus $\mathbb{P}[S_\tau < \log M] \approx 0$
- Key: $\ell = \mathbb{E}\left[\tau\right]$ averages-out normal fluctuations!

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

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

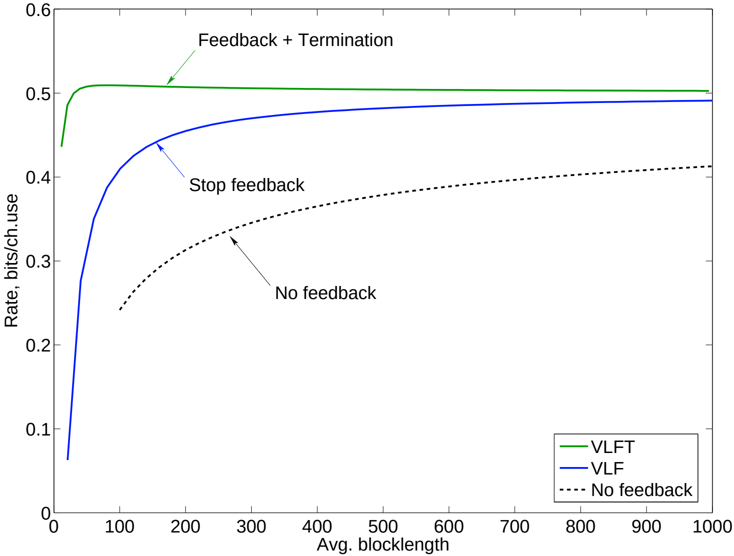
Codes employing feedback + termination (BSC version)

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:

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

- Stop feedback + variable-length:

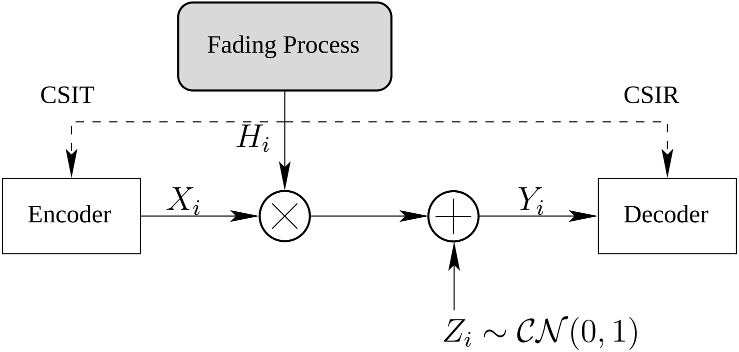
  $n \lesssim 200$ penalty term: $O\left(\frac{\log n}{n}\right)$

- Feedback + variable-length + termination:

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

Quasi-static wireless channels

Wireless channels: fading and AWGN



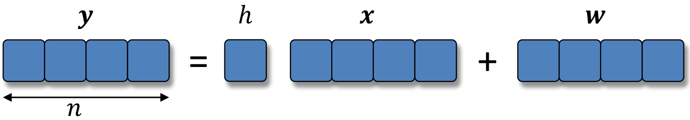
- AWGN channel with fluctuating SNR:

$$Y_i = H_i X_i + Z_i \,.$$

- Channel gain: known at Rx (CSIR) and/or Tx (CSIT)
- Channel dynamics: quasi-static $H_1 = H_2 = \cdots$

Quasi-static fading channel: SISO

I/O relation



Notes:

- H–random but fixed for all channel uses (nonergodic)
- Power-constraint: $\|\boldsymbol{x}\|^2 \le nP$
- Outage capacity:

$$C_\epsilon \stackrel{\triangle}{=} \lim_{n\to\infty} R^*(n,\epsilon) = \sup\{R : P_{\text{out}}(R) \le \epsilon\}$$

$$P_{\text{out}}(R) \stackrel{\triangle}{=} \mathbb{P}\left[\log(1 + P|H|^2) \le R\right]$$

- C_ϵ is same in all 4 cases: noCSI, CSIR, CSIT, CSIRT

Reason: Can easily learn the channel.

Main result: Dispersion is zero (SISO)

Theorem

If pdf of H is smooth and $P'_{\text{out}}(C_\epsilon) > 0$, *then*

$$R^*(n,\epsilon) = C_\epsilon + 0 \cdot \cancel{\frac{1}{\sqrt{n}}} + O\left(\frac{\log n}{n}\right)$$

In all 4 cases: noCSI, CSIR, CSIT, CSIRT

Why zero dispersion?

- AWGN channel

$$R^*(n,\epsilon) \approx C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon)$$

$$\Updownarrow$$

$$\epsilon \approx \mathbb{P}\left[C + \sqrt{\frac{V}{n}} Z \le R^*(n,\epsilon) \right], \quad Z \sim \mathcal{N}(0,1)$$

Why zero dispersion?

- AWGN channel

$$R^*(n,\epsilon) \approx C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon)$$

$$\Updownarrow$$

$$\epsilon \approx \mathbb{P}\left[C + \sqrt{\frac{V}{n}} Z \le R^*(n,\epsilon)\right], \quad Z \sim \mathcal{N}(0,1)$$

- Quasi-static case

$$\epsilon \approx \mathbb{P}\left[C(H) + \sqrt{\frac{V(H)}{n}} Z \le R^*(n,\epsilon)\right]$$

Why zero dispersion?

- AWGN channel

$$R^*(n,\epsilon) \approx C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon)$$

$$\Updownarrow$$

$$\epsilon \approx \mathbb{P}\left[C + \sqrt{\frac{V}{n}} Z \leq R^*(n,\epsilon) \right], \quad Z \sim \mathcal{N}(0,1)$$

- Quasi-static case

$$\epsilon \approx \mathbb{P}\left[C(H) + \sqrt{\frac{V(H)}{n}} Z \leq R^*(n,\epsilon) \right]$$
$$= \mathbb{P}\left[\frac{Z}{\sqrt{n}} \leq \frac{R^*(n,\epsilon) - C(H)}{\sqrt{V(H)}} \right]$$

Why zero dispersion?

- AWGN channel

$$R^*(n,\epsilon) \approx C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon)$$

$$\Updownarrow$$

$$\epsilon \approx \mathbb{P}\left[C + \sqrt{\frac{V}{n}} Z \le R^*(n,\epsilon) \right], \quad Z \sim \mathcal{N}(0,1)$$

- Quasi-static case

$$\begin{aligned} \epsilon &\approx \mathbb{P}\left[C(H) + \sqrt{\frac{V(H)}{n}} Z \le R^*(n,\epsilon) \right] \\ &= \mathbb{P}\left[\frac{Z}{\sqrt{n}} \le \frac{R^*(n,\epsilon) - C(H)}{\sqrt{V(H)}} \right] \\ &\to \mathbb{P}\left[C(H) \le R^*(n,\epsilon) \right], \quad n \to \infty \end{aligned}$$

Why zero dispersion?

Lemma

Let $A \perp\!\!\!\perp B$ *s.t.* B *has smooth pdf,* $\mathbb{E}[A] = 0, \mathbb{E}[A^2] = 1$*:*

$$\mathbb{P}\left[B \geq \frac{A}{\sqrt{n}}\right] = \mathbb{P}[B \geq 0] + O\left(\frac{1}{n}\right).$$

Why zero dispersion?

Lemma

Let $A \perp\!\!\!\perp B$ *s.t.* B *has smooth pdf,* $\mathbb{E}[A] = 0, \mathbb{E}[A^2] = 1$*:*

$$\mathbb{P}\left[B \geq \frac{A}{\sqrt{n}}\right] = \mathbb{P}[B \geq 0] + O\left(\frac{1}{n}\right).$$

$$\epsilon \approx \mathbb{P}\left[\frac{Z}{\sqrt{n}} \leq \frac{R^*(n,\epsilon) - C(H)}{\sqrt{V(H)}}\right]$$

Why zero dispersion?

Lemma

Let $A \perp\!\!\!\perp B$ *s.t.* B *has smooth pdf,* $\mathbb{E}[A] = 0, \mathbb{E}[A^2] = 1$*:*

$$\mathbb{P}\left[B \geq \frac{A}{\sqrt{n}}\right] = \mathbb{P}[B \geq 0] + O\left(\frac{1}{n}\right).$$

$$\epsilon \approx \mathbb{P}\left[\frac{Z}{\sqrt{n}} \leq \frac{R^*(n,\epsilon) - C(H)}{\sqrt{V(H)}}\right]$$

$$= \mathbb{P}\big[C(H) \leq R^*(n,\epsilon)\big] + O(1/n)$$

Why zero dispersion?

Lemma

Let $A \perp\!\!\!\perp B$ *s.t.* B *has smooth pdf,* $\mathbb{E}[A] = 0, \mathbb{E}[A^2] = 1$*:*

$$\mathbb{P}\left[B \geq \frac{A}{\sqrt{n}}\right] = \mathbb{P}[B \geq 0] + O\left(\frac{1}{n}\right).$$

$$\begin{aligned} P_{\mathrm{out}}(C_\epsilon) = \epsilon \approx \mathbb{P}\left[\frac{Z}{\sqrt{n}} \leq \frac{R^*(n,\epsilon) - C(H)}{\sqrt{V(H)}}\right] \\ = \mathbb{P}\big[C(H) \leq R^*(n,\epsilon)\big] + O(1/n) \\ = P_{\mathrm{out}}\big(R^*(n,\epsilon)\big) + O(1/n) \end{aligned}$$

Why zero dispersion?

Lemma

Let $A \perp\!\!\!\perp B$ *s.t.* B *has smooth pdf,* $\mathbb{E}[A] = 0, \mathbb{E}[A^2] = 1$*:*

$$\mathbb{P}\left[B \geq \frac{A}{\sqrt{n}}\right] = \mathbb{P}[B \geq 0] + O\left(\frac{1}{n}\right).$$

$$P_{\mathrm{out}}(C_\epsilon) = \epsilon \approx \mathbb{P}\left[\frac{Z}{\sqrt{n}} \leq \frac{R^*(n,\epsilon) - C(H)}{\sqrt{V(H)}}\right]$$

$$= \mathbb{P}\big[C(H) \leq R^*(n,\epsilon)\big] + O(1/n)$$

$$= P_{\mathrm{out}}\left(R^*(n,\epsilon)\right) + O(1/n)$$

$$\boxed{R^*(n,\epsilon) \approx C_\epsilon + O\left(\frac{1}{n}\right)}$$

Why zero dispersion?

Lemma

Let $A \perp\!\!\!\perp B$ s.t. B has smooth pdf, $\mathbb{E}[A] = 0, \mathbb{E}[A^2] = 1$:

$$\mathbb{P}\left[B \geq \frac{A}{\sqrt{n}}\right] = \mathbb{P}[B \geq 0] + O\left(\frac{1}{n}\right).$$

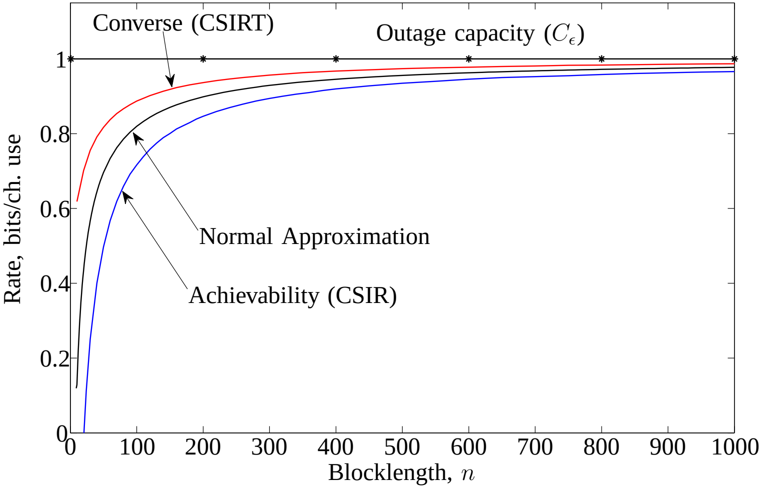
$$P_{\text{out}}(C_\epsilon) = \epsilon \approx \mathbb{P}\left[\frac{Z}{\sqrt{n}} \leq \frac{R^*(n,\epsilon) - C(H)}{\sqrt{V(H)}}\right]$$

$$= \mathbb{P}\big[C(H) \leq R^*(n,\epsilon)\big] + O(1/n)$$

$$= P_{\text{out}}\big(R^*(n,\epsilon)\big) + O(1/n)$$

$$\boxed{R^*(n,\epsilon) = C_\epsilon + O\left(\frac{\log n}{n}\right)}$$

Normal Approximation $R_{\mathrm{approx}}(n, \epsilon)$



spectre

Upper Bound on $R^*(n, \epsilon)$ (CSIRT)

- From meta-converse:

$$R^*(n, \epsilon) \le \sup_{P_X} \frac{1}{n} \log \frac{1}{\beta_{1-\epsilon}(P_{XYH}, P_X \times Q_{YH})}$$

Upper Bound on $R^*(n, \epsilon)$ (CSIRT)

- From meta-converse:

$$R^*(n, \epsilon) \leq \sup_{P_X} \frac{1}{n} \log \frac{1}{\beta_{1-\epsilon}(P_{XYH}, P_X \times Q_{YH})}$$

- SIMO: choose Q_{YH} to be *caod*
- MIMO: choose $Q_{YH|X}$ weakly dependent on X

Lower bound on $R^*(n, \epsilon)$ (no CSI)

- Use $\kappa\beta$ bound [PPV '10]

$$R^*(n,\epsilon) \geq \frac{1}{n} \log \frac{\kappa_\tau(F, Q_Y)}{\sup_{x \in F} \beta_{1-\epsilon+\tau}(P_{Y\,|\,X=x}, Q_Y)}$$

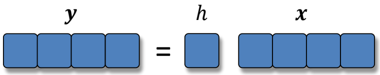
- A stochastically degraded channel

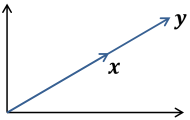
$$X \to Y \to \mathrm{span}(Y)$$

- $\beta_{1-\epsilon+\tau}$: angle thresholding
- κ_τ: computed in closed form

Lower bound on $R^*(n, \epsilon)$ (no CSI)

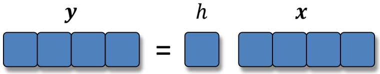
- Angle thresholding

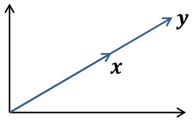


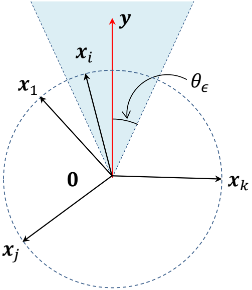


Lower bound on $R^*(n, \epsilon)$ (no CSI)

- Angle thresholding

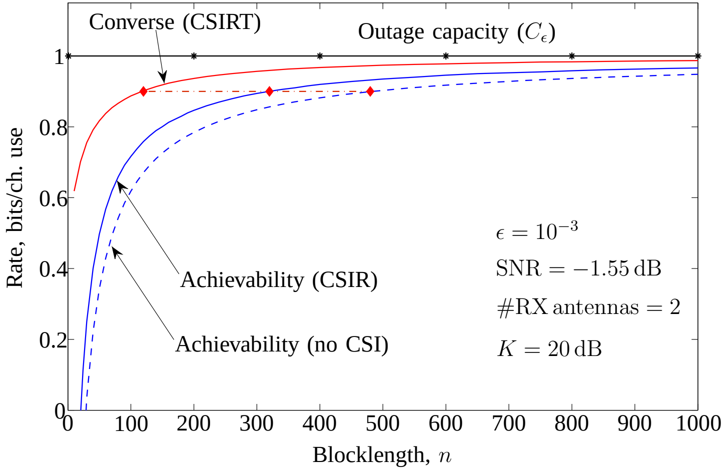




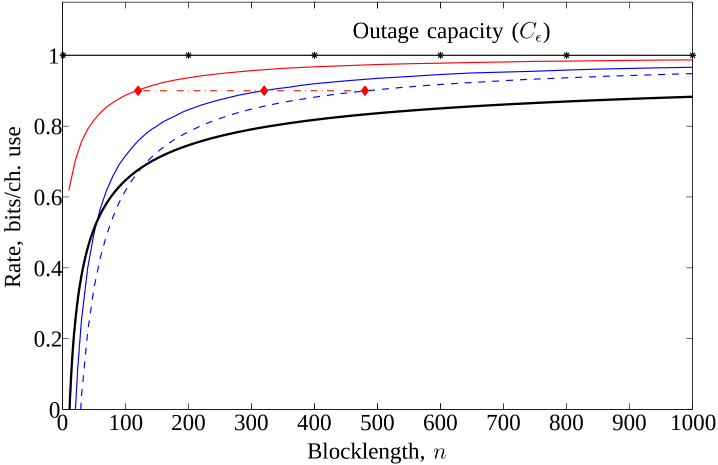


- Declare x_i iff $\theta(x_i, Y) \leq \theta_\epsilon$
- Declare error otherwise
- Knowledge of h or statistics of h not required

Bounds on $R^*(n, \epsilon)$



Bounds on $R^*(n, \epsilon)$



spectre

Does zero dispersion imply fast convergence to C_ϵ?

$$R^*(n,\epsilon) = C_\epsilon + O\left(\frac{\log n}{n}\right)$$

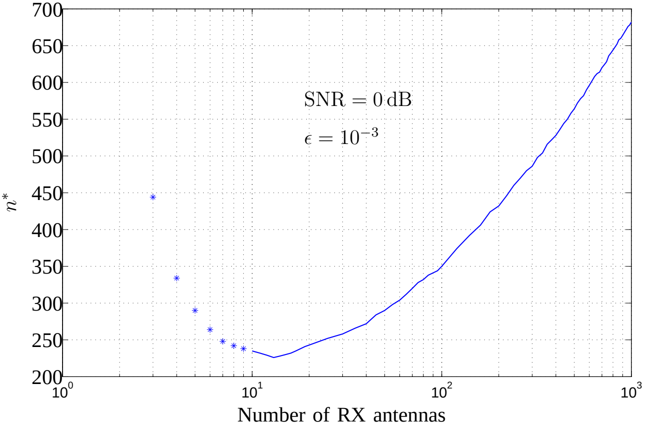
Does zero dispersion imply fast convergence to C_ϵ?

$$R^*(n,\epsilon) = C_\epsilon + O\left(\frac{\log n}{n}\right)$$

- n^*: blocklength required to achieve 90% of C_ϵ
- $n^*_{\mathrm{AWGN}} \approx 1400$

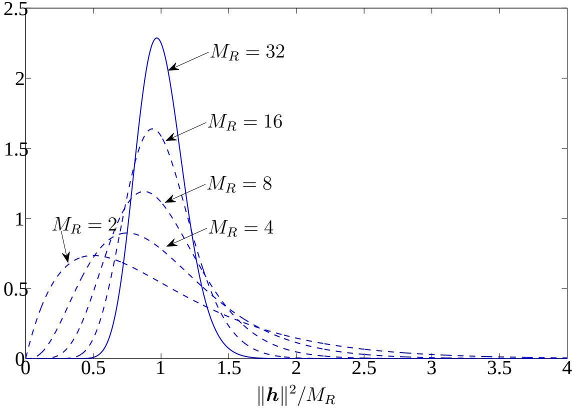
Does zero dispersion imply fast convergence to C_ϵ?

$$R^*(n, \epsilon) = C_\epsilon + O\left(\frac{\log n}{n}\right)$$

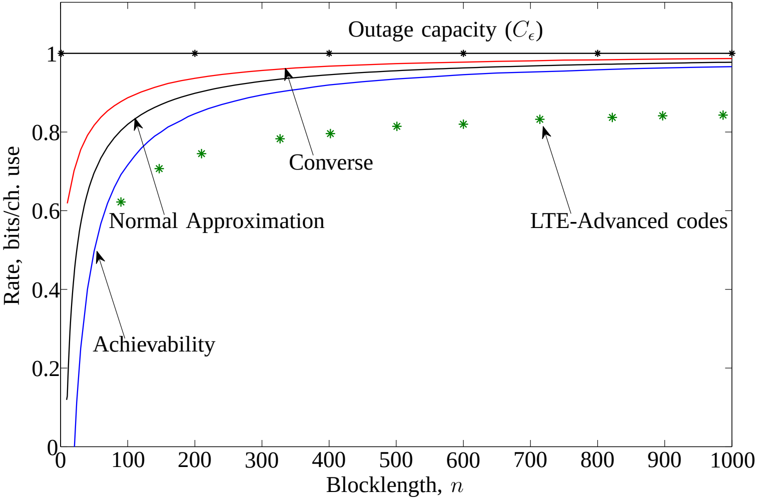


Does zero dispersion imply fast convergence to C_ϵ?

$$O\left(\frac{\log n}{n}\right) \to \sqrt{\frac{V_{\mathrm{AWGN}}}{n}}Q^{-1}(\epsilon) + O\left(\frac{\log n}{n}\right), \qquad M_R \to \infty$$

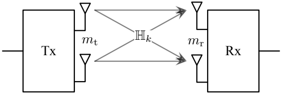


Performance of LTE-Advanced coding schemes



Quasi-static fading channels: MIMO

$$Y_i = \mathbb{H}X_i + Z_i, \qquad \mathbb{H} \in \mathbb{C}^{m_r \times m_t}$$



Theorem

If $\mathbb{H}$ has smooth density with polynomial tails:

$$R^*(n, \epsilon) = C_\epsilon + 0 \cdot \cancel{\frac{1}{\sqrt{n}}} + O\left(\frac{\log n}{n}\right)$$

Under all 4 cases: noCSI, CSIT, CSIR, CSIRT.

- Technical heavy-lifting
- Input achieving C_ϵ unknown (Telatar conjecture)

Energy-efficiency in wireless channels

Energy per bit: Gaussian channel

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

- Problem: Encode k bits into finite-energy waveforms:

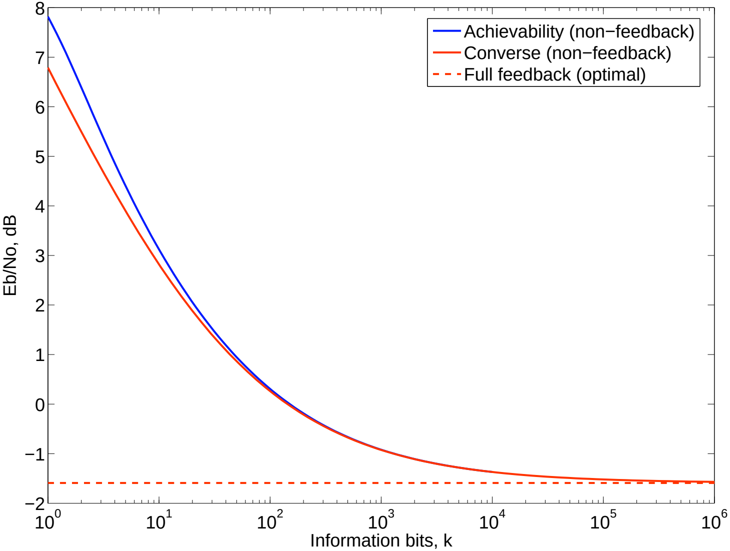
$$\mathbb{E}\left[\sum_{i=1}^{\infty} |X_i|^2\right] \le kE_b\,.$$

- **No rate-constraint** (aka wideband limit)
- Asymptotically: [Shannon'49]

$$\min\left(\frac{E_b}{N_0}\right) \to \log_e 2 = -1.59 \text{ dB} \qquad , k \to \infty\,.$$

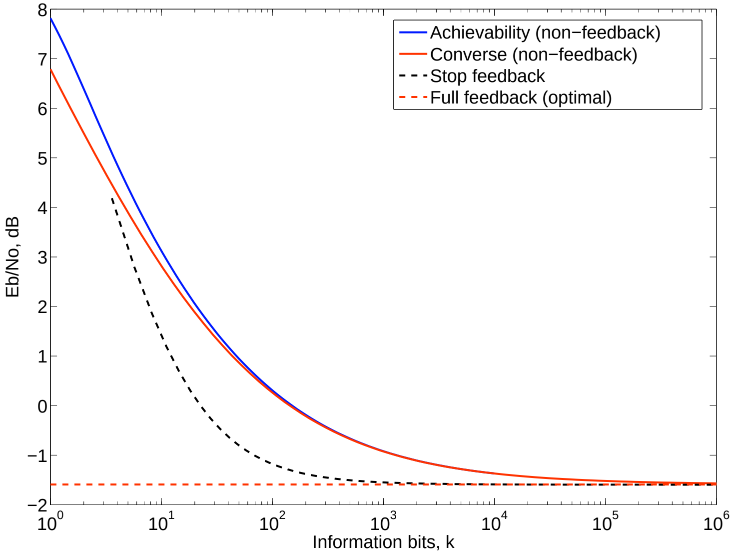
- ...and also **must** have $R \to 0$.

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



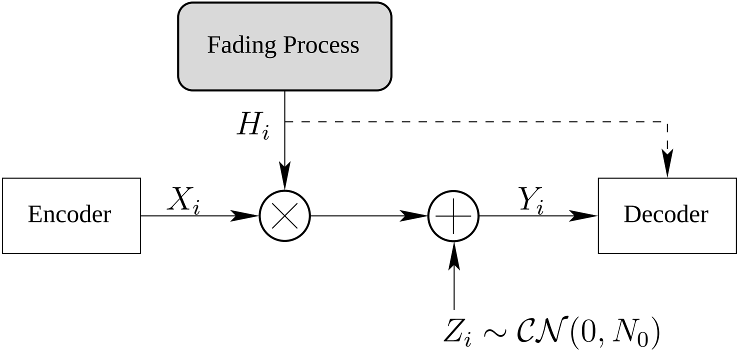
spectre

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



spectre

Energy per bit: Enter fading

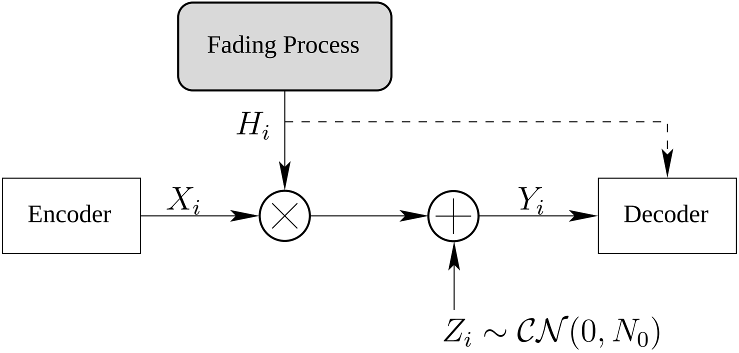


- AWGN channel with fluctuating SNR:

$$Y_i = H_i X_i + Z_i\,.$$

- Channel gain: known (CSIR) or unknown (noCSI)

Energy per bit: Enter fading

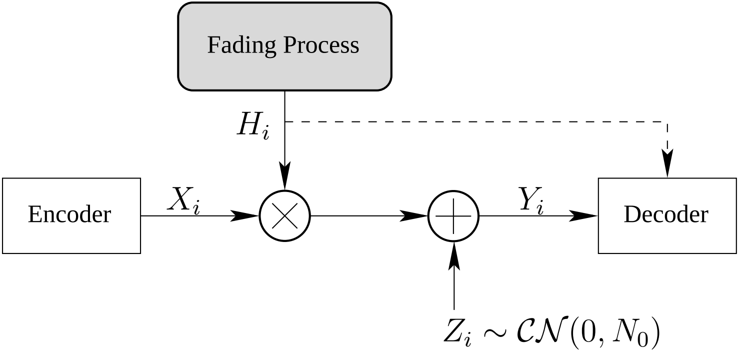


- [Jacobs'63], [Kennedy'69]: For both CSIR and noCSI (!)

$$\min\left(\frac{E_b}{N_0}\right) \to \log_e 2 = -1.59 \text{ dB} \qquad , k \to \infty \,.$$

- Achieved by very wideband (flash) signals

Energy per bit: Enter fading



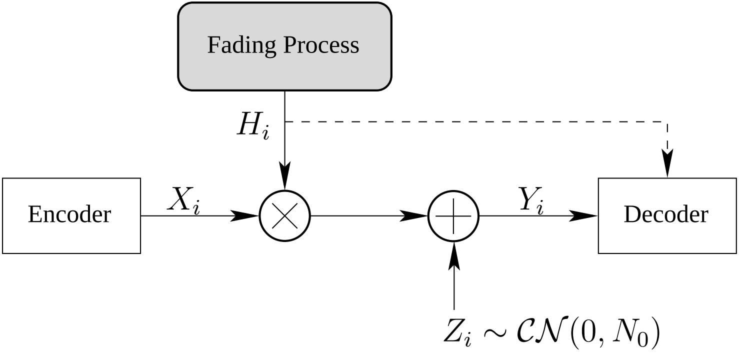
- [Verdú'02]:

$$\left(\frac{E_b}{N_0}\right) = (-1.59 \text{ dB}) + \epsilon \quad \Longrightarrow \quad R \approx e^{-\frac{1}{\epsilon}} \quad \text{(noCSI)}$$

$$\Longrightarrow \quad R \approx \epsilon \quad \text{(CSIR)}$$

- Note: $R = \frac{k}{n}$ – spectral efficiency

Energy per bit: Enter fading



- All of the above: $k \to \infty$.
- **Our work:** For finite k

$$\text{noCSI}: \quad E_b(k, \epsilon) \approx (-1.59\ dB) + c_\epsilon \sqrt[3]{\frac{\log k}{k}}$$

$$\text{w/ CSIR}: \quad E_b(k, \epsilon) \approx (-1.59\ dB) + c_\epsilon \sqrt{\frac{V}{k}}$$

noCSI fading: equivalent channel

$$Y_i = H_i X_i + Z_i, \qquad Y, H, X, Z \in \mathbb{C}$$

- Fading: assume $H_i \overset{iid}{\sim} \mathcal{CN}(0,1)$ (**Rayleigh (!)** fast-fading)
- Input: Only depends on $|X_i|^2$
- Output: Suff. statistic is $|Y_i|^2$
- Equivalent channel:

$$|Y_i|^2 = (1 + |X_i|^2) S_i, \qquad S_i \overset{iid}{\sim} \mathrm{Exp}(1)\,.$$

- Info density:

$$i(x; y) \triangleq \log \frac{dP_{Y^\infty | X^\infty = x}}{dP_{Y^\infty | X^\infty = 0}}(y)$$

- ...and its distribution given $X = x$:

$$i(x; Y) \overset{d}{=} \sum_i |x_i|^2 S_i \log e - \log(1 + |x_i|^2)$$

noCSI fading: information density

- Info density:

$$i(x;Y) \stackrel{d}{=} \sum_i |x_i|^2 S_i \log e - \log(1+|x_i|^2)$$

- Bit-per-energy: Maximize $(1-\epsilon)$-quantile of $\frac{i(x;Y)}{E}$ s.t.

$$\sum_i |x_i|^2 \le E$$

- Classical asymptotics: Take $\mathbb{E}\left[i(x;Y)\right]$ and $E \to \infty$

$$\max \frac{\sum_i |x_i|^2 \log e - \log(1+|x_i|^2)}{\sum_i |x_i|^2} = \log e \quad \left(\frac{\text{bit}}{\text{energy}/N_0}\right)$$

Optimizer: $x_i \neq 0 \implies |x_i| \gg 1$ (flash!)

noCSI fading: information density

- Goal: Maximize $(1-\epsilon)$-quantile of $\frac{i(x;Y)}{E}$.

$$i(x;Y) \stackrel{d}{=} \sum_i |x_i|^2 S_i \log e - \log(1+|x_i|^2)$$

noCSI fading: information density

- Goal: Maximize $(1-\epsilon)$-quantile of $\frac{i(x;Y)}{E}$.

$$i(x;Y) \stackrel{d}{=} \sum_i |x_i|^2 S_i \log e - \log(1+|x_i|^2)$$

- Fundamental tension:
  1. To have LLN need lots of small atoms.
  2. To have $\frac{\sum_i \log(1+|x_i|^2)}{\sum_i |x_i|^2} \to 0$ need large atoms.

noCSI fading: information density

- Goal: Maximize $(1-\epsilon)$-quantile of $\frac{i(x;Y)}{E}$.

$$i(x;Y) \stackrel{d}{=} \sum_i |x_i|^2 S_i \log e - \log(1+|x_i|^2)$$

- Add extra constraint $|x_i|^2 \le A$: [Sethuraman-Hajek'05]

$$\max \frac{\sum_i |x_i|^2 \log e - \log(1+|x_i|^2)}{\sum_i |x_i|^2} = \left\{ \log e - \frac{\log(1+A)}{A} \right\}$$

- By CLT:

$$\frac{1}{E} i(x;Y) \sim \left\{ \log e - \frac{\log(1+A)}{A} \right\} + \sqrt{\frac{A \log^2 e}{E}} Z$$

- Optimize: $A^* \sim (Q^{-1}(\epsilon))^{-\frac{2}{3}} E^{\frac{1}{3}} \log^{\frac{2}{3}} E$

noCSI fading: information density

- Goal: Maximize $(1-\epsilon)$-quantile of $\frac{i(x;Y)}{E}$.

$$i(x;Y) \overset{d}{=} \sum_i |x_i|^2 S_i \log e - \log(1+|x_i|^2)$$

- Add extra constraint $|x_i|^2 \le A$: [Sethuraman-Hajek'05]

$$\max \frac{\sum_i |x_i|^2 \log e \ldots}{\sum_i \ldots} \ldots + A) \ldots A \Big\}$$

- By CLT:

$$\frac{1}{E} i(x;Y) \sim \Big\{ \ldots Z$$

This is only a heuristic!

- Optimize: $A^* \sim (Q^{-1}(\epsilon))^{-\frac{2}{3}} E^{\frac{1}{3}} \log^{\frac{2}{3}} E$

Achievability: non-asymptotic c.p.u.c. bound

Theorem (Verdú'90)

For a memoryless channel $P_{Y|X}$ the capacity-per-unit-cost is

$$C_{p.u.c.} = \sup_x \frac{D(P_{Y|X=x} \| P_{Y|X=0})}{\mathrm{cost}(x)}$$

Achievability: non-asymptotic c.p.u.c. bound

Theorem

There exists a k-bit code with $\mathbb{P}[\text{error}] = \epsilon$ and

$$2^k \geq \sup_{0<\tau<\epsilon} \frac{\tau}{\beta_{1-\epsilon+\tau}(P^{\otimes N}_{Y|X=x}, P^{\otimes N}_{Y|X=0})}$$

$$\frac{\textit{cost}}{\textit{bit}} = \frac{N \cdot \mathrm{cost}(x_0)}{k}$$

where $N \in \mathbb{Z}_+$ and $x_0 \in \mathcal{X}$.

Proof:

- As in [Verdú'90] use codewords:

$$c_j = [\underbrace{0, \ldots, 0}_{(j-1)N}, \underbrace{x_0, \ldots, x_0}_{N}, 0, \ldots] \qquad j \in [2^k]$$

- And run 2^k binary tests "$X = c_j$ vs $X = 0$"
- Can be seen as a version of $\kappa\beta$-bound from [PPV'10]

Asymptotics: noCSI fading channel

Theorem (Yang-Durisi-P.'15)

Max # of bits transmittable with energy budget E

$$k^*(E, \epsilon) = E \log_2 e - \mathrm{const} \cdot \left(Q^{-1}(\epsilon) E\right)^{\frac{2}{3}} \log_2^{\frac{1}{3}} E + \cdots$$

Proof:

- For any codeword $x \in \mathbb{C}^\infty$:

$$i(x; Y) \stackrel{d}{=} \sum_{n=1}^{\infty} |x_i|^2 S_i - \log(1 + |x_i|^2), \quad S_i \stackrel{iid}{\sim} \mathrm{Exp}(1)$$

- Analysis (tedious!) shows: ϵ-quantile is maximized by

$$x^* = [\underbrace{\sqrt{A}, \sqrt{A}, \ldots, \sqrt{A}}_{N\text{-times}}, 0, \ldots]$$

$$N = c \left(Q^{-1}(\epsilon) E \log^{-1} E\right)^{\frac{2}{3}}, \quad N \cdot A = E$$

What about perfect CSI at Receiver?

$$Y_i = H_i X_i + Z_i, \qquad Y, H, X, Z \in \mathbb{C}$$

Theorem

Fading with CSIR = non-fading AWGN (for energy-per-bit).

Small print: Achievability and asymptotics coincide.

- Simulate AWGN channel

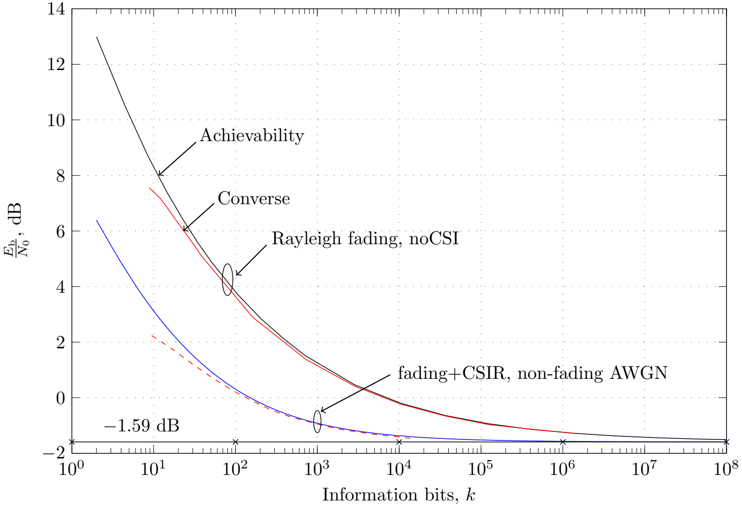
$$\text{Tx:} \qquad X_1 = \cdots = X_n = \frac{x}{\sqrt{n}}$$

$$\text{Rx:} \qquad \hat{Y} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \bar{H}_i Y_i$$

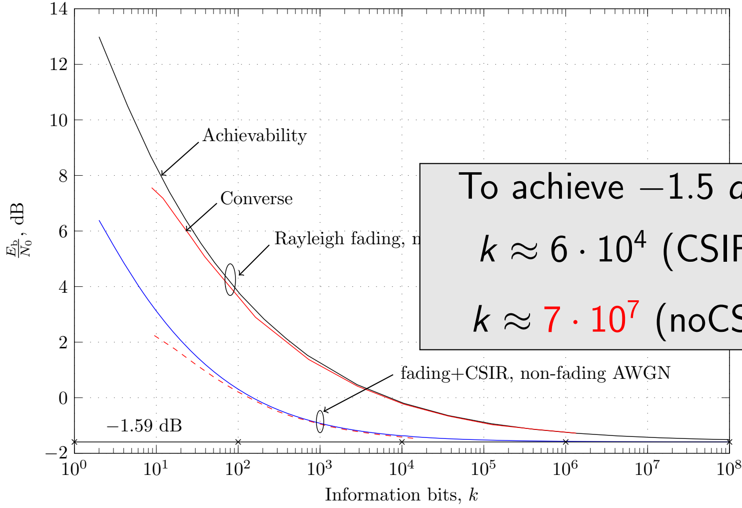
- As $n \to \infty$ we have by LLN and CLT: $(\mathbb{E}\,|H|^2 = 1)$

$$\hat{Y} = x \cdot \frac{1}{n} \sum_{i=1}^{n} |H_i|^2 + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \bar{H}_i Z_i \quad \to \quad x + Z\,.$$

Energy-per-bit: noCSI vs CSIR



Energy-per-bit: noCSI vs CSIR



To achieve $-1.5\ dB$:

$k \approx 6 \cdot 10^4$ (CSIR)

$k \approx 7 \cdot 10^7$ (noCSI)

spectre

Extensions: non-iid fading

Non-iid fading process:

- Block-fading: $H_1 = \cdots = H_{n_c}$.
- Achievability: Simulate iid H (use one symbol per block)
- Converse: info density

$$i(x; Y) \stackrel{d}{=} \sum_{i=1}^{\infty} \|x_i\|^2 S_i - \log(1 + \|x_i\|^2)$$

  where $x_i \in \mathbb{C}^{n_c}$ is input to i-th fading block.
  Proof: Apply unitary mtx to each x_i.
- $\implies$ Achievability/Converse/Asymptotics **same** as for iid H
- $\implies$ Learning the channel is futile (!)

Extensions: Multiple antennas

$$\underbrace{\overset{n_c}{\longleftrightarrow}}{} \quad m_r \left[\; Y_i \;\right] = \mathbb{H}_i \left[\; X_i \;\right]_{m_t}^{n_c} + Z_i$$

Multiple antennas:

- m_t Tx antennas: Apply $U(m_t) \times U(n_c)$ symmetries to reduce

$$X_i \sim \begin{bmatrix} x_{i,1} & 0 & 0 & 0 \\ & \ddots & & \\ 0 & 0 & x_{i,m} & 0 \end{bmatrix}$$

$\implies$ equivalent to iid SIMO fading.

- m_r Rx antennas: just reduce noise from N_0 to $\frac{N_0}{m_r}$
- $\implies$ Achievability/Converse/Asymptotics **same** as for SISO iid H
- $\implies$ STBC, V-BLAST etc – useless (!)

Energy-per-bit: conclusion

Punchline:

- AWGN and fading channels with CSIR:

$$E_b(k,\epsilon) \approx (-1.59\ dB) + \sqrt{\frac{V}{k}}Q^{-1}(\epsilon)$$

- Rayleigh fading noCSI:

$$E_b(k,\epsilon) \approx (-1.59\ dB) + \sqrt[3]{\frac{\log k}{k}\left(Q^{-1}(\epsilon)\right)^2}$$

- Unchanged under MIMO and/or block-fading
  (Optimal design: 1 Tx antenna and no pilots.)

Finite blocklength IT codebase:

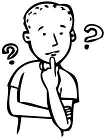
http://GitHub.com/yp-mit/spectre

(MATLAB, many contributors)

Finite blocklength IT codebase:

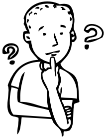
http://GitHub.com/yp-mit/spectre

(MATLAB, many contributors)



Ask questions (now or later)!

Yury Polyanskiy <yp@mit.edu>



Thank you!