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
**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

Data bits

Redundancy

Noisy channel

Reliability–Rate tradeoff

Possible

Impossible
Channel coding: principles

Data bits

Redundancy

Noisy channel

Decreasing $P_e$ further:

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

Reliability–Rate tradeoff

$n = 10$
Data bits

Redundancy

Noisy channel

Decreasing $P_e$ further:

1. More redundancy
   **Bad:** loses rate

2. Increase blocklength!
Channel coding: principles

Data bits

Redundancy

Noisy channel

Decreasing $P_e$ further:

1. More redundancy
   Bad: loses rate

2. Increase blocklength!

$n = 1000$
Channel coding: principles

Decreasing $P_e$ further:

1. More redundancy
   - Bad: loses rate
2. Increase blocklength!
**Channel coding: Shannon capacity**

**Shannon:** Fix $R < C$

$P_e \downarrow 0$ as $n \to \infty$
Channel coding: Shannon capacity

Shannon: Fix $R < C$

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

Question:

For what $n$ will $P_e < 10^{-3}$?
Channel coding: Gaussian approximation

Shannon: Fix $R < C$

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

Question:
For what $n$ will $P_e < 10^{-3}$?
Channel coding: Gaussian approximation

**Shannon:** Fix $R < C$

\[ P_e \downarrow 0 \text{ as } n \rightarrow \infty \]

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

**Answer:**
\[ n \gtrsim \text{const} \cdot \frac{V}{C^2} \]

**Reliability–Rate tradeoff**

**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 $\epsilon$, 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 \( \epsilon \))

**Note:** Exact value unknown (search is doubly exponential in \( n \)).
Fix $R < C$. What is the smallest blocklength $n^*$ needed to achieve

$$R^*(n, \epsilon) \geq R \quad ?$$
Fix $R < C$. What is the smallest blocklength $n^*$ needed to achieve

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

**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 $\epsilon = 10^{-3}$:

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

**Difficulty:** How to verify accuracy of this estimate?
 Bounds

- Bounds are implicit in Shannon’s theorem

\[
\lim_{n \to \infty} R^*(n, \epsilon) = C \iff \begin{cases}
R^*(n, \epsilon) \leq C + o(1), \\
R^*(n, \epsilon) \geq C + o(1).
\end{cases}
\]

(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:

\[
\epsilon \leq e^{-nE_r(R)} \quad \text{[Gallager’65]}
\]
\[
\epsilon \leq e^{-nE_r(R-o(1)) + O(\log n)} \quad \text{[Csizár-Körner’81]}
\]
Bounds

- Bounds are implicit in Shannon’s theorem

\[
\lim_{n \to \infty} R^*(n, \epsilon) = C \iff \begin{cases} 
R^*(n, \epsilon) \leq C + o(1), \\
R^*(n, \epsilon) \geq C + o(1). 
\end{cases}
\]

(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:

\[
\epsilon \leq e^{-nE_r(R)} \quad [\text{Gallager’65}]
\]

\[
\epsilon \leq e^{-nE_r(R-o(1)) + O(\log n)} \quad [\text{Csiszár-Körner’81}]
\]

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
  \[
  \lim_{n \to \infty} R^*(n, \epsilon) = C \iff \left\{ \begin{array}{l}
  R^*(n, \epsilon) \leq C + o(1), \\
  R^*(n, \epsilon) \geq C + o(1).
\end{array} \right.
  \]
  (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 \( n > \frac{19600}{\epsilon^{16}} \ldots (1) \)

- Solution: Derive bounds from scratch.
New achievability bound

**Theorem (RCU)**

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

\[
\epsilon \leq \sum_{t=0}^{\frac{n}{2}} \binom{n}{t} \delta^t (1-\delta)^{n-t} \min \left\{ 1, \sum_{k=0}^{t} \binom{n}{k} 2^{-n-nR} \right\}.
\]
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)

\[
\text{Hamming Space } \mathbb{F}_2^n
\]
Proof of RCU bound for the BSC

- Input space: $A = \{0, 1\}^n$
- Let $c_1, \ldots, c_M \sim Bern\left(\frac{1}{2}\right)^n$ (random codebook)
- Transmit $c_1$
Proof of RCU bound for the BSC

- Input space: \( A = \{0, 1\}^n \)
- Let \( c_1, \ldots, c_M \sim Bern\left(\frac{1}{2}\right)^n \) (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 $c_1 \rightarrow Y$
- $Y = c_1 + Z$, $Z \sim Bern(\delta)^n$
- Decoder: find closest codeword to $Y$
Proof of RCU bound for the BSC

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

\( Y = c_1 + Z, \quad Z \sim \text{Bern}(\delta)^n \)

- Decoder: find closest codeword to \( Y \)
Proof of RCU bound for the BSC

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

$Y = c_1 + Z, \quad Z \sim Bern(\delta)^n$
- 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 \( c_1 \rightarrow Y \)
- \( Y = c_1 + Z, \quad Z \sim Bern(\delta)^n \)
- Decoder: find closest codeword to \( Y \)
Proof of RCU bound for the BSC

- Input space: $A = \{0, 1\}^n$
- Let $c_1, \ldots, c_M \sim \text{Bern}(\frac{1}{2})^n$ (random codebook)
- Transmit $c_1$
- Noise displaces $c_1 \rightarrow Y$
- $Y = c_1 + Z$, $Z \sim \text{Bern}(\delta)^n$
- Decoder: find closest codeword to $Y$
- Probability of error analysis:

\[
P[\text{error} | Y, \text{wt}(Z) = t] = P[\exists j > 1 : c_j \in \text{Ball}(Y, t)]
\leq \sum_{j=2}^{M} P[c_j \in \text{Ball}(Y, t)]
\leq 2^{nR} \sum_{k=0}^{t} \binom{n}{k} 2^{-n}
\]
... cont'd ...

- Conditional probability of error:

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

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

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

- Average \( \text{wt}(Z) \sim \text{Binomial}(n, \delta) \implies \text{Q.E.D.} \)

Note: Step \( \ast \) tightens Gallager’s \( \rho \)-trick:

\[
\mathbb{P} \left[ \bigcup_{j} A_j \right] \leq \left( \sum_{j} \mathbb{P}[A_j] \right)^\rho
\]
Sphere-packing converse (BSC variation)

Theorem (Elias’55)

For any $(n, M, \epsilon)$ code over the BSC$(\delta)$:

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

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

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

Note: Convexity of $f$ follows from general properties of $\beta_\alpha$ (below)
Sphere-packing converse (BSC variation)

Proof:
- Denote decoding regions $D_j$: $\bigcap D_j = \{0, 1\}^n$
- 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 $D_j$: $\bigcap D_j = \{0, 1\}^n$

- Probability of error is:
  \[ \epsilon = \frac{1}{M} \sum_j \mathbb{P}[c_j + Z \notin D_j] \geq \frac{1}{M} \sum_j \mathbb{P}[Z \notin B_j] \]

- $B_j = \text{ball centered at 0 s.t. } \text{Vol}(B_j) = \text{Vol}(D_j)$
Sphere-packing converse (BSC variation)

Proof:

- Denote decoding regions $D_j$: $\bigcup D_j = \{0, 1\}^n$
- 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. $\text{Vol}(B_j) = \text{Vol}(D_j)$
- Simple calculation:
  \[ \mathbb{P}[Z \notin B_j] = f(\text{Vol}(B_j)) \]
- $f$ – convex, apply Jensen:
  \[ \epsilon \geq f \left( \frac{1}{M} \sum_{j=1}^{M} \text{Vol}(D_j) \right) = f \left( \frac{2^n}{M} \right) \]
Bounds: example $BSC(0.11), \epsilon = 10^{-3}$
Bounds: example $BSC(0.11), \varepsilon = 10^{-3}$

![Graph showing bounds and converse for BSC(0.11)]

- **Capacity**
  - Delay required to achieve 90% of $C$: $2985 \leq n^* \leq 3106$
  - Error-exponent: $n^* \approx 5000$
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 < ϵ < 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 \( \epsilon \)!
Normal approximation: $BSC(0.11); \epsilon = 10^{-3}$

$$C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon) + \frac{1}{2} \log \frac{n}{n}$$
Normal approximation: $BSC(0.11); \epsilon = 10^{-3}$

To achieve 90% of $C$:

$$n^* \approx 3150$$

$$C - \sqrt{\frac{V}{n} Q^{-1}(\epsilon)} + \frac{1}{2} \log \frac{n}{n}$$
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) graph](graph)

- Less noise
- More noise
Delay required to achieve 90 % of capacity:

▶ Error-exponents:

$$n^* \approx 5000$$

▶ True value:

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

▶ Channel dispersion:

$$n^* \approx 3150$$
Converse Bounds
Notation

- Take a random transformation $A \overset{P_{Y|X}}{\longrightarrow} B$
  (think $A = \mathcal{A}^n$, $B = \mathcal{B}^n$, $P_{Y|X} = P_{Y^n|X^n}$)
- Input distribution $P_X$ induces $P_Y = P_{Y|X} \circ P_X$
  $P_{XY} = P_X P_{Y|X}$
- Fix code:
  $$W \overset{encoder}{\rightarrow} X \rightarrow Y \overset{decoder}{\rightarrow} \hat{W}$$
  $W \sim \text{Unif}[M]$ and $M = \#$ of codewords
  Input distribution $P_X$ associated to a code:
  $$P_X[\cdot] \triangleq \# \text{ of codewords } \in (\cdot) \over M$$
- Goal: Upper bounds on $\log M$ in terms of $\epsilon \triangleq \mathbb{P}[\text{error}]$
  As by-product: $R^*(n, \epsilon) \lesssim C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon)$
Fano’s inequality

**Theorem (Fano)**

For any code

\[
\begin{align*}
\text{encoder} & : W 
\rightarrow X \\
\quad & : P_{Y|X} 
\rightarrow Y \\
\text{decoder} & : Y 
\rightarrow \hat{W}
\end{align*}
\]

with \( W \sim \text{Unif}\{1, \ldots, M\} \):

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

Implies weak converse:

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

**Proof:** \( \epsilon \)-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 \((W, X, Y, \hat{W})\):

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

\[
\text{DAG: } W \rightarrow X \rightarrow Y \rightarrow \hat{W}
\]

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

\[
\text{DAG: } W \rightarrow X \rightarrow Y \rightarrow \hat{W}
\]

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 $D(\cdot \Vert \cdot)$

\[
\begin{align*}
D(P_A \Vert Q_A) & \geq D(P_B \Vert Q_B) \\
\end{align*}
\]
Data-processing for $D(\cdot \| \cdot)$

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

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

$D(P_{WXY\hat{W}} \| Q_{WXY\hat{W}}) \geq d(P[W = \hat{W}] \| Q[W = \hat{W}])$

$= d(1 - \epsilon \| \frac{1}{M})$

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

$D(\cdot \| \cdot)$ denotes the Kullback-Leibler divergence.
A proof of Fano via *channel substitution*

So far:

\[ D(P_{WXY\hat{W}} \parallel Q_{WXY\hat{W}}) \geq d(1 - \epsilon \parallel \frac{1}{M}) \]

Lower-bound RHS:

\[ d(1 - \epsilon \parallel \frac{1}{M}) \geq (1 - \epsilon) \log M - h(\epsilon) \]

Analyze LHS:

\[
D(P_{WXY\hat{W}} \parallel Q_{WXY\hat{W}}) = D(P_{XY} \parallel Q_{XY}) = D(P_X P_{Y|X} \parallel P_X Q_Y) = D(P_{Y|X} \parallel Q_Y | P_X)
\]

(Recall: \( D(P_{Y|X} \parallel Q_Y | P_X) = \mathbb{E}_{x \sim P_X} [D(P_{Y|X=x} \parallel Q_Y)] \))
A proof of Fano via *channel substitution*: last step

Putting it all together:

\[(1 - \epsilon) \log M \leq D(P_{Y|X} \parallel Q_Y | P_X) + h(\epsilon) \quad \forall Q_Y \quad \forall \text{code}\]

Two methods:

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

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

2. Take \( Q_Y = P_Y^* = \text{the caod} \) (capacity achieving output dist.) and recall

\[D(P_{Y|X} \parallel P_Y^* | P_X) \leq \sup_{P_X} I(X; Y) \quad \forall P_X\]

Conclude:

\[(1 - \epsilon) \log M \leq \sup_{P_X} I(X; Y) + h(\epsilon)\]

*Important*: Second method is particularly useful for FBL!
Question: How about replacing $D(\cdot \mid \cdot)$ with other divergences?

Answer:

<table>
<thead>
<tr>
<th>$D(\cdot \mid \cdot)$</th>
<th>relative entropy (KL divergence)</th>
<th>weak converse</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_\lambda(\cdot \mid \cdot)$</td>
<td>Rényi divergence</td>
<td>strong converse</td>
</tr>
<tr>
<td>$\beta_\lambda(\cdot, \cdot)$</td>
<td>Neyman-Pearson ROC curve</td>
<td>FBL bounds</td>
</tr>
</tbody>
</table>

Next: What is $\beta_\lambda$?
Neyman-Pearson’s $\beta_\alpha$

**Definition**

For every pair of measures $P, Q$

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

**Important:** Like relative entropy $\beta_\alpha$ satisfies data-processing.
$\beta_\alpha = \text{binary hypothesis testing}$

Two types of errors:

$$
\mathbb{P}\left[\text{Tester says \"Q}_Y\"\right] \leq \epsilon \\
\mathbb{Q}\left[\text{Tester says \"P}_Y\"\right] \rightarrow \min
$$

Hence: Solve binary HT $\iff$ compute $\beta_\alpha(P^n_Y, Q^n_Y)$

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

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

But in fact $\log \beta_\alpha(P^n_Y, Q^n_Y)$ can also be computed exactly!
How to compute $\beta_\alpha$?

**Theorem (Neyman-Pearson)**

$\beta_\alpha$ is given parametrically by $-\infty \leq \gamma \leq +\infty$:

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

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

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

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

where

$$V(P||Q) = \text{Var}_P \left[ \log \frac{P(X)}{Q(X)} \right]$$
Back to proving converse

Recall two measures:

\[
\begin{align*}
\mathbb{P} : & \quad P_{WXY\hat{W}} = P_W \times P_{X|W} \times P_{Y|X} \times P_{\hat{W}|Y} \\
& \quad \text{DAG}: W \to X \to Y \to \hat{W} \\
& \quad \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} \\
& \quad \text{DAG}: W \to \underline{X} \to Y \to \hat{W} \\
& \quad \mathbb{Q}[W = \hat{W}] = \frac{1}{M}
\end{align*}
\]

Then by definition of \( \beta_\alpha \):

\[
\beta_{1-\epsilon}(P_{WXY\hat{W}}, Q_{WXY\hat{W}}) \leq \frac{1}{M}
\]

But \[
\log \frac{P_{WXY\hat{W}}}{Q_{WXY\hat{W}}} = \log \frac{P_X P_{Y|X}}{P_X Q_Y} \quad \Rightarrow \quad \log M \leq -\log \beta_{1-\epsilon}(P_X P_{Y|X}, P_X Q_Y) \quad \forall Q_Y \quad \forall \text{code}
\]
Meta-converse: minimax version

**Theorem**

Every \((M, \epsilon)\)-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 \((M, \epsilon)\)-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 \(\beta_\alpha\) 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 \((M, \epsilon)\)-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 \((M, \epsilon)\)-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 $(M, \epsilon)$-code for channel $P_{Y|X}$ satisfies

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

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 $Q_Y \approx \text{“center” of } P_{Y|X}$
- **Exercise**: Redo BSC.
Example: Converse for AWGN

The AWGN Channel

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

\[ X \rightarrow \oplus \rightarrow Y \]

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 \( n \rightarrow n + 1 \) equalize powers:

\[
\|c_j\|^2 = nP \quad \forall j \in \{1, \ldots, M\}
\]
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 \( n \rightarrow n + 1 \) equalize powers:
  \[ \|c_j\|^2 = nP \quad \forall j \in \{1, \ldots, M\} \]

- 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 \( \mathbb{P} \): \( Y^n = X^n + \mathcal{N}(0, \mathbf{I}_n) \)
  \( \implies \) distribution of LLR (CLT approx.)
  \[ \approx nC + \sqrt{nVZ} \quad Z \sim \mathcal{N}(0, 1) \]
  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.)
  \[ \approx nC + \sqrt{nV}Z, \quad Z \sim \mathcal{N}(0, 1) \]

- 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.)
  \[
  \approx nC + \sqrt{nV} Z, \quad Z \sim \mathcal{N}(0, 1)
  \]

- **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 \( \chi^2 \) 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:

$SNR = 0 \text{ dB}, \epsilon = 10^{-3}$
From one $Q_Y$ to many

Symmetric channel
inputs equidistant to $P_Y^*$
“distance” $= -\beta \alpha (\cdot, \cdot)$

cAoD $P_Y^*$

Symmetric channel: choice of $Q_Y$ is clear
From one $Q_Y$ to many

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

General channels: Caod is no longer equidistant to all inputs (read: analysis horrible!)
From one $Q_Y$ to many

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.: $Q_{Y|X=x} = \text{center of a cluster of } x$

- Prove converse bound for channel $Q_{Y|X}$

  E.g.: $Q[W = \hat{W}] \lesssim \frac{# \text{ of clusters}}{M}$

- Find $\beta_\alpha(P, Q)$, i.e. compare:

  \[
  P : P_{WX\hat{Y}} = P_W \times P_{X|W} \times P_{Y|X} \times P_{\hat{W}|Y}
  \]

  vs.

  \[
  Q : P_{WX\hat{Y}} = P_W \times P_{X|W} \times Q_{Y|X} \times P_{\hat{W}|Y}
  \]

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

  \[
  \beta_{1-P_e(P_{Y|X})} \leq 1 - P_e(Q_{Y|X}) \quad \forall \text{code}
  \]
Meta-converse theorem: point-to-point channels

**Theorem**

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

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

**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:** $Q[Y \in \cdot | W = w] = P[Y \in \cdot | W \neq w]$
  
  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 $\epsilon \triangleq P[\text{error}]$ and $\epsilon' \triangleq Q[\text{error}]$ 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
- 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 $\epsilon \triangleq P[error]$ and $\epsilon' \triangleq Q[error]$ 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.: $Q[W \in \{\text{list}\}] = \frac{|\{\text{list}\}|}{M}$
- 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:
  \[
P[\text{error}] = \epsilon, \quad Q[\text{error}] = \epsilon'.
\]

- **Same idea**: use code as a suboptimal binary HT: \(P_{Y|X}\) vs. \(Q_{Y|X}\)
- ... and compare to the best possible test:
  \[
  D(P_{XY} \parallel Q_{XY}) \geq d(1 - \epsilon \| 1 - \epsilon')
  \]
  \[
  \beta_{1-\epsilon}(P_X P_{Y|X}, P_X Q_{Y|X}) \leq 1 - \epsilon'
  \]
Example: MAC (weak-converse)

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

\[ \cdots \text{apply data processing of } D(\cdot \| \cdot) \cdots \]
\[ d(1 - \epsilon \| \frac{1}{M_1}) \leq D(P_{Y|X_1X_2} \| Q_{Y|X_1}P_{X_1}P_{X_2}) \]

Optimizing \( Q_{Y|X_1} \):

\[ \log M_1 \leq \frac{I(X_1; Y|X_2) + h(\epsilon)}{1 - \epsilon} \]

Also with \( X_1 \leftrightarrow X_2 \implies \text{weak converse (usual pentagon)} \)
Example: MAC (FBL?)

\[ P[\hat{W}_{1,2} = W_{1,2}] = 1 - \epsilon \quad \text{and} \quad Q[\hat{W}_{1,2} = W_{1,2}] = \frac{1}{M_1} \]

\[ \vdots \text{use } \beta_{\alpha}(\cdot, \cdot) \quad \vdots \]

\[ \beta_{1-\epsilon}(P_{X_1X_2Y}, P_{X_1QX_2Y}) \leq \frac{1}{M_1} \]

On-going work: This \( \beta_{\alpha} \) is highly non-trivial to compute.

[Huang-Moulin, MolavianJazi-Laneman, Yagi-Oohama]
Achievability Bounds
A random transformation $A \xrightarrow{P_{Y|X}} B$

$(M, \epsilon)$ codes:

$W \rightarrow A \rightarrow B \rightarrow \hat{W}$  
$W \sim \text{Unif}\{1, \ldots, M\}$

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

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

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

$\mathbb{E}[\iota_{X;Y}(X;Y)] = I(X;Y)$

$\text{Var}[\iota_{X;Y}(X;Y)|X] = V$

Memoryless channels: $\iota_{A^n;B^n}(A^n;B^n) = \text{sum of iid.}$

$\iota_{A^n;B^n}(A^n;B^n) \overset{d}{\approx} nI(A;B) + \sqrt{nVZ}$,  
$Z \sim \mathcal{N}(0,1)$

Goal: Prove FBL bounds.

As by-product: $R^*(n, \epsilon) \gtrsim C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon)$
**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:
  \[
  M \geq \sup_{\gamma \geq 0} \left\{ \gamma (\epsilon - \mathbb{P}[n_X; Y(X; Y) \leq \log \gamma]) \right\}
  \]

- **Shannon’57 bound**: \( \exists (M, \epsilon) \)-code:
  \[
  \epsilon \leq \inf_{\gamma \geq 0} \left\{ \mathbb{P}[n_X; Y(X; Y) \leq \log \gamma] + \frac{M - 1}{\gamma} \right\}.
  \]

- **Gallager’65 bound**: \( \exists (n, M, \epsilon) \)-code over memoryless channel:
  \[
  \epsilon \leq \exp \left\{ -n E_r \left( \frac{\log M}{n} \right) \right\}.
  \]

- 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

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

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

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 $d_{Ham}(\cdot, \cdot) \leftrightarrow -\iota_{X;Y}(\cdot, \cdot)$
- For example ML decoder: $\hat{W} = \arg\max_j \iota_{X;Y}(C_j; Y)$
- Conditional prob. of error:

$$
\mathbb{P}[\text{error} | X, Y] \leq (M - 1)\mathbb{P}[\iota_{X;Y}(\bar{X}; Y) \geq \iota_{X;Y}(X; Y) | X, Y]
$$

- 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

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

$$\pi(a, b) = \mathbb{P} \left[ \mathbb{I}_{X;Y}(\tilde{X}; Y) \geq \mathbb{I}_{X;Y}(X; Y) \mid X = a, Y = b \right]$$

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

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

$$\epsilon \leq \mathbb{E} \left[ \exp \left\{ - \left| I_{XY} - \log \frac{M-1}{2} \right|^+ \right\} \right].$$

Highlights:

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

▶ Codebook: random $C_1, \ldots, C_M \sim P_X \text{ iid}$

▶ Feinstein decoder:

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

▶ $j$-th codeword’s probability of error:

$$\mathbb{P}[	ext{error} \mid W = j] \leq \mathbb{P}[\mathcal{I}_{X;Y}(X; Y) \leq \gamma] + (j - 1) \mathbb{P}[\mathcal{I}_{X;Y}(\bar{X}; Y) > \gamma]$$

In (a): $C_j$ too far from $Y$

In (b): $C_k$ with $k < j$ is too close to $Y$

▶ Average over $W$:

$$\mathbb{P}[	ext{error}] \leq \mathbb{P}[\mathcal{I}_{X;Y}(X; Y) \leq \gamma] + \frac{M-1}{2} \mathbb{P}[\mathcal{I}_{X;Y}(\bar{X}; Y) > \gamma]$$
DT bound: Proof

- Recap: for every $\gamma$ there exists a code with

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

- Key step: closed-form optimization of $\gamma$.

Note: $\rho_{X;Y} = \log \frac{dP_{XY}}{dP_{\bar{X}Y}}$

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

Bayesian dependence testing!

**Optimum threshold:** Ratio of priors $\implies \gamma^* = \log \frac{M-1}{2}$

- Change of measure argument:

$$P \left[ \frac{dP}{dQ} \leq \tau \right] + \tau Q \left[ \frac{dP}{dQ} > \tau \right] = \mathbb{E}_P \left[ \exp \left\{ - \left| \log \frac{dP}{dQ} - \log \tau \right|^+ \right\} \right].$$
Example: Binary Erasure Channel $BEC(0.5)$, $\epsilon = 10^{-3}$
Example: Binary Erasure Channel \( BEC(0.5), \epsilon = 10^{-3} \)

At \( n = 1000 \) best \( k \rightarrow n \) code:

\[
\begin{align*}
(DT) & \quad 450 \leq k \leq 452 \quad \text{(meta-c.)}
\end{align*}
\]

Converse: \( Q_{Y^n} = \text{truncated caod} \)
Theorem

For all $Q_Y$ and $\tau$ there exists an $(M, \epsilon)$-code inside $F \subset A$

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

where

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

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 \( y \).
- Test \( y \) against each codeword \( c_i, i = 1, \ldots M \):
  
  Run optimal binary HT for:
  
  \[
  \mathcal{H}_0 : P_{Y|X=c_i} \\
  \mathcal{H}_1 : Q_Y
  \]

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

- First test that returns \( \mathcal{H}_0 \) becomes the decoded codeword.
- If all \( \mathcal{H}_1 \) – declare error.
**k/β bound: idea**

**Decoder:**

- Take received $y$.
- Test $y$ against each codeword $c_i, i = 1, \ldots M$:
  
  Run optimal binary HT for:
  
  $\mathcal{H}_0 :$ \( P_{Y|X=c_i} \)
  
  $\mathcal{H}_1 :$ \( Q_Y \)

  \[
  \begin{align*}
  \mathbb{P}[	ext{detect } \mathcal{H}_0] &= 1 - \epsilon + \tau \\
  \mathbb{Q}[	ext{detect } \mathcal{H}_0] &= \beta_1 - \epsilon + \tau (P_{Y|X=x}, Q_Y)
  \end{align*}
  \]

- 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 $\tau$-disjoint: $\mathbb{P}[Y \in B_x \cap \text{others}|x] \leq \tau$
- Key step: Cannot pick more codewords $\Rightarrow$
  $M \bigcup_{j=1}^{M} \{j\text{-th decoding region}\}$ is a composite HT:
  \[
  \mathcal{H}_0 : \quad P_{Y|X=x} \quad x \in F
  \]
  \[
  \mathcal{H}_1 : \quad Q_Y
  \]
- Performance of the best test:
  \[
  \kappa_{\tau} = \inf_{\{E: P_{Y|X}[E|x] \geq \tau \quad \forall x \in F\}} Q_Y[E].
  \]
- Thus:
  \[
  \kappa_{\tau} \leq Q[\text{all } M \text{ “balls”}]
  \leq M \sup_x \beta_{1-\epsilon+\tau}(P_{Y|X=x}, Q_Y)
  \]
Hierarchy of achievable bounds (no cost constr.)

- Arrows show logical implication

> Gallager

> Feinstein

DT

RCU
Hierarchy of achievability bounds (no cost constr.)

- Arrows show logical implication
- Performance ↔ computation rule of thumb.
Hierarchy of achievability bounds (no cost constr.)

- Arrows show logical implication
- Performance ↔ computation rule of thumb.
- ISIT’2013: Haim-Kochman-Erez [WeB4]
Channel Dispersion
Connection to CLT

Recap:

- Let $P_{Y^n|X^n} = P_{Y|X}^n$ 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):

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

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

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

This section: Study $\sqrt{n}$-term.
General definition of channel dispersion

Definition

For any channel we define channel dispersion as

\[ V = \lim_{\epsilon \to 0} \lim_{n \to \infty} n \frac{(C - R^*(n, \epsilon))^2}{2 \ln \frac{1}{\epsilon}} \]

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

\[ V = \lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{n(C - R^*(n, \epsilon))^2}{2 \ln \frac{1}{\epsilon}} \]

**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]:**
  \[ V = \text{Var}[i_X; Y(X; Y)] \quad X \sim \text{capacity-achieving} \]

- **AWGN channel [PPV’08]:**
  \[ V = \frac{\log^2 e}{2} \left[ 1 - \left( \frac{1}{1 + \text{SNR}} \right)^2 \right] \]

- **Parallel AWGN [PPV’09]:**
  \[ V = \sum_{j=1}^{L} V_{\text{AWGN}} \left( \frac{W_j}{\sigma^2_j} \right) \quad \{ W_j \} - \text{waterfilling powers} \]

- **DMC with input constraints [Hayashi’09, P’10]:**
  \[ V = \text{Var}[i_X; Y(X; Y)|X] \quad X \sim \text{capacity-achieving} \]
Dispersion of channels with memory

From [PPV’09, PPV’10, PV’11]:

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

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

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

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

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

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

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

- **ISIT’13**: Quasi-static fading channels: $V = 0$ (!)

  Yang-Durisi-Koch-P. in [WeA4]
Dispersion: product vs generic channels

- Relation to alphabet size:
  \[ V \leq 2 \log^2 \min(|A|, |B|) - C^2. \]

- Dispersion is additive:
  \[
  \begin{align*}
  \{ &A_1 \to \text{DMC}_1 \to B_1, \\
  &A_2 \to \text{DMC}_2 \to B_2 \}\n  = &A_1 \times A_2 \to \text{DMC} \to B_1 \times B_2
  \\
  C &= C_1 + C_2, \quad V_\epsilon = V_{1,\epsilon} + V_{2,\epsilon}
  \\
  \implies &\text{product DMCs have atypically low dispersion.}
  \end{align*}
  \]
**Dispersion and normal approximation**

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

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

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

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

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 $O(n^{-2/3})$.*
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.
  
  $\implies$ 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.

  $\Rightarrow$ define rate-invariant metric:

**Definition (Normalized rate)**

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

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

- 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)

Normalized rates of code families over BIAWGN, $P_e=0.0001$

- Turbo $R=1/3$
- Turbo $R=1/6$
- Turbo $R=1/4$
- Voyager
- Galileo HGA
- Turbo $R=1/2$
- Cassini/Pathfinder
- Galileo LGA
- Hermitian curve $[64,32]$ (SDD)
- BCH (Koetter–Vardy)
- Polar+CRC $R=1/2$ (List dec.)
- ME LDPC $R=1/2$ (BP)
Performance of short algebraic codes (BSC, $\epsilon = 10^{-3}$)

![Performance of short algebraic codes](chart.png)
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. $\epsilon$) maximizes throughput?
Optimizing ARQ systems

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

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

- **Question:** Given $k$ bits what rate (equiv. $\epsilon$) 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

$$\epsilon \approx 10^{-3} \ldots 10^{-2}$$
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$(\delta)$

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

where

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

$$\tau \triangleq \inf\{n \geq 0 : S_n \geq \gamma\},$$

$$S_n \triangleq n \log(2 - 2\delta) + \log \frac{\delta}{1 - \delta} \cdot \sum_{k=1}^{n} Z_k,$$

$$Z_k \sim i.i.d. \text{Bernoulli}(\delta).$$
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

\begin{equation}
W \xrightarrow{X_t} \text{Encoder} \xrightarrow{Y_t} \text{Channel} \xrightarrow{Y_t} \text{Decoder} \xrightarrow{\hat{W}}
\end{equation}

\begin{equation}
W \xrightarrow{z^{-1}} Y_{t-1} \xrightarrow{z^{-1}} W
\end{equation}

\begin{equation}
\text{STOP}
\end{equation}
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

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

\[ n \lesssim 20 \]
Delay to achieve 90% of the capacity of the BSC(0.11):

- **No feedback:**
  \[ n \approx 3100 \quad \text{penalty term:} \quad O\left(\frac{1}{\sqrt{n}}\right) \]

- **Stop feedback + variable-length:**
  \[ n \lesssim 200 \quad \text{penalty term:} \quad O\left(\frac{\log n}{n}\right) \]

- **Feedback + variable-length + termination:**
  \[ n \lesssim 20 \quad \text{penalty term:} \quad O\left(\frac{1}{n}\right) \]
Gaussian channel: Energy per bit

\[ Z_i \sim \mathcal{N}(0, \frac{N_0}{2}) \]

\[ X_i \rightarrow \bigoplus \rightarrow Y_i \]

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

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

**Asymptotically:** [Shannon’49]

\[
\min \left( \frac{E_b}{N_0} \right) \rightarrow \log 2 = -1.6 \text{ dB} \quad , \quad k \rightarrow \infty.
\]
Energy per bit vs. number of information bits ($\epsilon = 10^{-3}$)

![Graph showing the relationship between energy per bit and the number of information bits for different feedback scenarios.](image-url)
Energy per bit vs. # of information bits ($\epsilon = 10^{-3}$)
Summary

Classical: \((n \to \infty)\)

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

Optimal coding

\(C\)

\(\mathcal{N}(C, \frac{V}{n})\)
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]
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


References: multi-terminal compression


References: Channel coding (1948-2008)


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


References: Channel coding (2008-)


References: Channel coding (2008-)


References: Channel coding (multi-terminal)


References: Lossy compression, joint source-channel coding


Thank you!

Do not hesitate to ask questions!

Yury Polyanskiy <yp@mit.edu>
Sergio Verdú <verdu@princeton.edu>