Today

- Converse Coding Theorem for BSC
- Error-Correcting codes:
  - Parameters of interest
  - Greedy codes: Gilbert bound
  - Linear codes: Dual; Vanshamov bound.

Recall Coding Theorem

For BSC (p), let \( C = C(p) = 1 - H(p) \)

Then \( R < C \), suff. large \( n \), \( E \in E, D \)

mapping \( k = 8n \) bits to \( n \) bits & back s.t.
Deoding error Prob. \( \leq \exp(-n) \).
Converse: if $R > C$ then for sufficiently large $n,$

\[ \forall E, D \text{ mapping } k = Rn \text{ bits to } n \text{ bits,} \]

\[ \text{Prob. } [ \text{Decoding Error } ] \geq 1 - \exp(-n) . \]

**Proof:** Fix $R, n, E, D$

\[ \text{uniform} \]

Transmit $E(m)$ for random $m$

Receive $E(m) + Y$ for random $Y$.

\[ \text{Bsc}(p) \]

**Bad Events**

\[ E_1 : \text{Too few error } \text{wt}(Y) \leq (p - \frac{\epsilon}{2})n \]

\[ E_2 : \exists x \text{ st. } E_2(x) \text{ where} \]
\( E_2(x) : \)

(i) \( \omega + (E(\theta(x)) - x) \geq (p - \frac{\xi}{2})n \)

(ii) \( m = \theta(x) \)

(iii) \( Y = E(\theta(x)) \)

\( \sum x \)

\( A \)

\[ \Pr[Y \mid E_1] \to \exp(-\eta) \quad \text{[Chernoff]} \]

\( B \)

\[ \Pr[Y \mid E_2(x)] \]

\[ = 2^{-k} \cdot 2^{-H(p) \cdot \eta} \]

\[ \Pr[(\text{ii})] \quad \Pr[(\text{iii}) \ \text{given (i)}] \]

\[ \Pr[ E_2 ] \leq \sum_x \Pr[ E_2(x) ] \]

\[ = 2^n \cdot 2^{-k} \cdot 2^{-H(p) \cdot \eta} \to \exp(-\eta) \]
Neither $E1$ nor $E2 \Rightarrow$ Decoding failure:

Proof: Assume $T \subseteq E1$, $T \subseteq E2$, and $\text{Correct Decoding}$.

Let $x = E(m) + y$

\[ \frac{\text{Correct Decoding}}{\Rightarrow} D(x) = m \]

$7E1 \Rightarrow wt(y) \geq (p - \frac{\xi}{2})n$

$\downarrow$

$wt(x - E(m)) \geq (p - \frac{\xi}{2})n$

$\downarrow$

$E2(x) \in \begin{cases} 
(i) & wt(x - E(D(x))) \geq (p - \frac{\xi}{2})n \\
(ii) & D(x) = m \\
(iii) & y = x - E(D(x)) 
\end{cases}$

Contradiction! \( \Box \)
Notes on Shannon Theory

1. Theory much broader (than just BSC) ... but no crisp characterization of when it holds.

2. Even in cases where it holds, not clear how to compute $C$.
   
   Example: DELETION channel (drops bit with prob. $\frac{1}{2}$).

3. Proofs not constructive.

   Rest of course: Effort to seek constructive versions.
Contrast between Hamming & Shannon

- Surprisingly disjoint emphasis.
  - Shannon: $E, D$ but not what makes pair work.
  - Hamming: $C = \text{image}(E)$ but no mention of $E, D$ themselves.

- Different Error Models:
  - Shannon: random errors
  - Hamming: bounded & worst-case errors.
  (Why?)
We'll use Hamming theory: Why?

bit more "constructive"

- Can at least "prove" C is not good (in all cases)

- Can we "math" to prove C is good.
Codes & Parameters

A code $C$ over alphabet $\mathbb{F}_q$ is an $(n, k, d)_q$ code if

(i) $q = |\mathbb{F}_q|$

(ii) $C \subseteq \mathbb{F}_q^n$

(iii) $|C| \geq q^k$

(iv) $\Delta(C) = \min \{ \Delta(x, y) \mid x \neq y, x, y \in C, \Delta(x, y) \geq d \}$

Note: Four basic parameters
- Two too many for 2-d plots.
Asymptotics

- Fix $q$ [today $q = 2$]
- Study $R = \frac{k}{n}$ vs. $S = \frac{d}{n}$ as $n \to \infty$

- Need to fill this 2-d plot

Prior to Shannon: Probably didn't think $R > 0$ & $S > 0$ possible ...
**Greedy Code**

(Achieve $R, S > 0$)

Fix $S$; take large $n$; let $d = 8n$.

$C \leftarrow \emptyset$; $S = \{0,1\}^n$

while $S \neq \emptyset$ do

Pick $x \in S$ arbitrarily;

$C \leftarrow C \cup \{x\}$;

$S \leftarrow S - \text{Ball}(x, d-1)$;

endwhile

Output $C$;

Claim: $\Delta(C) \geq d$ (Obvious)

Claim: $|C| \geq \frac{2^n}{|\text{Ball}(\emptyset, d-1)|} \approx 2 \cdot n(1-H(S))$
Theorem: \exists \text{ codes with } R = 1 - H(\delta)

Problem Set 2:

• Prove random code does not achieve this.

• Prove random code + deletion does achieve this.
Linear Codes

$C \subseteq \{0,1\}^n$ is linear if

$\forall x,y \in C, \quad x+y \in C.$

**Fact**: $C$ linear

$\iff \exists G \ k \times n$ matrix s.t.

\[
C = \{ \sum x \cdot G \mid x \in \{0,1\}^k \}
\]

$\iff \exists H \ n \times (n-k)$ s.t.

\[
C = \{ y \in \{0,1\}^n \mid y \cdot H = 0 \}
\]

**Note**: Definition extends to $C = F_q^n$ (finite field on $q$ elements)

**Notation**: Linear $(n,k,d)_q$ denoted $[n,k,d]_q$. 
Fact: $H$ is parity check of $[n,k,d]_q$ code iff every subset of $d-1$ rows linearly independent.

Varshamov’s Greedy (Linear) Code:
- Add rows to $H$ greedily;
- Initially $H$ empty;
- While $\exists v \in \{0,1\}^{n-k}$ s.t. $v = h_1 \ldots h_\ell$ for any (every) $h_1, \ldots, h_\ell \in H$
  $\ell \leq d-2$
- $H \leftarrow H \cup \{v\}$
- Output $H$;
Claim: $C = \{ y \mid y_1 + = 0 \}$ has $\Delta(c) \geq d$.

Claim: Get code of length $n$ with $2^k$ codewords provided

$$2^{n-k} > \left| \text{Ball}(n-1, d-2) \right|$$

(i.e., $2^k < \frac{2^n}{\left| \text{Ball}(n-1, d-2) \right|}$ achievable)

Note: Matches Hamming for $d = 3$

[Gilbert]Greedy $|C| = \Omega\left( \frac{2^n}{n^2} \right)$

[Varshamov]Greedy $|C| = \Omega\left( \frac{2^n}{n} \right)$
Gilbert-Vanshamov Plot

Still not constructive
Next Few Lectures

Giv bound: $R \geq 1 - H(S)$

Best known for binary codes

Some conjecture optimal.
(Non-asymptotic Improvements)

\[ \text{Given: } 2^k \geq \frac{2^n}{\text{Vol}(n,d-2)} \quad [\text{Vol}(n,d) = |\text{Ball}(0,d)|] \]

\[ [\text{BCH}] \quad 2^k \geq \Omega \left( \frac{2^n}{(d-1)^{d/2}} \right) \quad (\text{Will see later}) \]

\[ [\text{Jiang-Vardy}] : \quad 2^k \geq \Omega(d) \frac{2^n}{\text{Vol}(n,d-2)} \]

**Proof Sketch:**

- Let \( G \) be a graph with vertices \( \{0,1\}^n \).
- Edge \((x,y)\) if \( \Delta(x,y) \leq d-1 \).
- \( C \subseteq \{0,1\}^n \) has \( \Delta(C) \geq d \) if \( C \) independent set of \( G \).
$D \triangleq$ degree of vertices of $G$

\[ V_1(n, d-1) \]

**Turan's Theorem**: $G$ has ind. set of size

\[ |V(G)|/(D+1) \Rightarrow G_{UV} \text{ bound.} \]

- But # triangles in $G$ small

\[ \# \Delta \text{'s} = 2^n \cdot D^{2-\varepsilon} \text{ (for } \varepsilon > 0) \]

- [Ajtai Komlos Szemeredi] ...

\[ \Rightarrow \# \Delta \text{'s} = O \left( \frac{\log D \cdot 2^n}{D} \right) \]

... $\Box$