Today: Concatenated Codes
    Start Polar Codes.

**Concatenated Codes**

Let us consider $BSC(p)$

$$C = 1 - h(p).$$

Shannon's Theorem implies: (after modifications)

\[
\begin{cases}
\text{Fix } p, \varepsilon > 0, \forall n \geq \Omega(\frac{1}{\varepsilon^2}), \\
\exists \text{ linear code given by generator matrix } G \in \{0,1\}^{n \times k} \text{ s.t. } k = (1 - h(p) - \varepsilon)n,
\end{cases}
\]

\[
\forall u, \quad \Pr[\text{Gu is closest to } \text{BSC}_p(Gu)] \geq 1 - 2^{-\Omega(\varepsilon^3 n)}.
\]

**Concatenated Codes:** Use Shannon's existential code for small lengths & combine with explicit longer codes.

"Outer"

We take $C_{\text{outer}} \subseteq \{0,1\}^{n_0}$, also linear.

\[
C: \\
\text{adversarially} \text{ corrupt up to } \gamma n_0 \text{ bits.} \quad \text{but } \gamma \text{ small.}
\]

\[
y = c + e \quad \downarrow \text{Decoder}
\]

Correct $c$. 

Such explicit codes are known. Assume we know it.

(i.e., deterministic poly(n0) time construction).

The rate of $C_\circ$ can be $1 - \delta(\gamma)$, and it can correct worst-case $\gamma n_0$ errors. ($\delta(\gamma) \rightarrow 0$ as $\gamma \rightarrow 0$).

We can take $\delta(\gamma) = \gamma^{1/3}$.

$$m \xrightarrow{\text{Encoder}} e \in C_\circ$$

if $|m| = k \Rightarrow |c| \leq \frac{k}{1 - \sqrt[3]{\delta}}$.

One possibility would be to re-encode $e$ with a Shannon BSC(p) code. But that’s too inefficient.

We decompose the task into blocks, and encode each block with Shannon code.

$$b \approx O\left(\frac{1}{\varepsilon^2} \log \frac{1}{\delta}\right)$$

rate $= \log(b(p|x)) + 8$

$1 - h(p) - \frac{3}{2}$.
Find the "inner" Shannon code by brute force in time

$$\text{poly}(\frac{1}{\varepsilon}).$$

Overall rate

$$= (1 - \frac{3}{\sqrt{\gamma}})(1 - h(p)) - \varepsilon \frac{1}{2}$$

$$\geq 1 - h(p) - \varepsilon$$ for \(\gamma\) small enough.

**Decoding:**

1. Decode inner codes first by brute force.

By Shannon's theorem, for each block of \(b'\) bits, the error probability is small.

2. Decode the resulting string by the outer code decoder.

*Runtime = \text{poly}(n_0 \cdot 2^{b'}) + \text{poly}(n_0) = \text{poly}(n_0 \cdot 2^{\frac{1}{2}\varepsilon^2}).$

*Correctness: Need to make sure that the output of 1 has \(\leq \gamma n_0\) errors (w.h.p).

* For each block, \(\Pr(\text{error}) \leq 2^{-\Omega(\varepsilon^2 b') \leq \gamma^2}$

with constant in \(b\) large enough.

* The errors between blocks are independent \(\Rightarrow\) we can use Chernoff's bound to make sure that the overall error fraction is \(\leq \gamma\) w.h.p.
* Concatenated Codes is by Forney '1966.

**Polar Codes.**

Let us focus on $\text{BEC}_\alpha$, $\text{Cap} = 1 - \alpha'$.

$$k = R \cdot n \approx (1 - \alpha')N.$$ 

Say $G$ is capacity achieving.

$$G : \begin{bmatrix} u_0 \\ u_{k-1} \end{bmatrix} \rightarrow \begin{bmatrix} x_0 \\ \vdots \\ x_N \end{bmatrix}$$

BEC: $$\begin{bmatrix} y_0 \\ \vdots \\ y_{N-1} \end{bmatrix}$$

$U$ is a r.v. denoting $u$.

$X := \cdots x$. 

Claim: Because $G$ allows recovery of $y$ from $y_0$ w.h.p,

$$H(U_0, U_{k-1} \mid Y_0, Y_{N-1}) \rightarrow 0.$$ 

Chain rule:

$$H(U_0 \mid Y_0^{N-1}) + H(U_1 \mid U_0, Y_0^{N-1}) + \cdots + H(U_{k-1} \mid U_0^{k-2}, Y_0^{N-1}) \rightarrow 0.$$
Now, think of an invertible matrix

\[ G_N := N \left[ \begin{array}{cc} A & G \end{array} \right] \]

\[ \text{rk} \left( G \right) = k. \]

Say: Columns of \( A \) + \( G \) = basis of \( \{0,1\}^N \).

Note: \( G_N \) is invertible. Now,

\[ \begin{bmatrix} G_N \\ U_1 \\ \vdots \\ U_k \end{bmatrix} \rightarrow \begin{bmatrix} X \\ \vdots \\ \vdots \end{bmatrix} \rightarrow \text{BEC}_E \rightarrow \begin{bmatrix} Y_1 \\ \vdots \\ Y_d \end{bmatrix} \]

\[ H(U_0^{N-1} \mid Y_0^{N-1}) = H(X_0^{N-1} \mid Y_0^{N-1}) = N \cdot H(X_0 \mid Y_0) \]

\[ \text{independence} \]

\[ = N \cdot \alpha. \quad \text{Chain rule again:} \]

\[ H(U_0 \mid Y_0^{N-1}) + H(U_1 \mid U_0 Y_0^{N-1}) + \ldots + H(U_{N-1} \mid U_0^{N-2} Y_0^{N-1}) \]

\[ = \alpha \cdot N. \]

We claim the way the sum adds up to \( \alpha \cdot N \) is that an \( \alpha \) fraction is close to 1 and the rest are close to 0.
Claim:\n\[ H(U_{N-k}^{N-1} | Y_0^{N-1}, U_0^{N-k-1}) \rightarrow 0. \]

Easy to see from the fact that \( G \) is a good code.

\( \Rightarrow \) The last \( k \) terms in the summation are nearly 0.
\( \Rightarrow \) The rest are nearly 1.

This is called Polarization!

This works for any memoryless channel and any good code \( G \).

This is due to Arikain. (2009).