Def: A (classical) channel $N$ is specified by:

- input alphabet $X$
- output alphabet $Y$
- a distribution $p(y|x)$ for each $x \in X$

$x \in X$ $\xrightarrow{N} y \in Y$ $wp$ $p(y|x)$

Aside: can write $N$ as a stochastic matrix.
**Example 1** Binary symmetric Channel (BSC)

\[ X = Y = \{0, 1, 2\}, \text{ Input } \{\text{Sent} \text{ wp } 1-p, \text{ flipped wp } p \} \]

\[
\begin{array}{c c c c}
X & \overset{1-p}{\rightarrow} & y & \overset{0}{\rightarrow} & 0 \\
0 & \overset{p}{\rightarrow} & 1 & \overset{p}{\rightarrow} & 0 \\
1 & \overset{1-p}{\rightarrow} & 1 & \overset{1-p}{\rightarrow} & 1 \\
\end{array}
\]

**Example 2** Erasure channel (Er)

\[ X = Y = \{0, 1, 2\}, \text{ Input } \{\text{Sent} \text{ wp } 1-p, \text{ replaced by } 2 \text{ wp } p \} \]

\[
\begin{array}{c c c c}
X & \overset{1-p}{\rightarrow} & y & \overset{0}{\rightarrow} & 0 \\
0 & \overset{p}{\rightarrow} & 1 & \overset{p}{\rightarrow} & 0 \\
1 & \overset{1-p}{\rightarrow} & 2 & \overset{0}{\rightarrow} & 1 \\
\end{array}
\]

\[
\begin{array}{c c c c}
& & 0 & 1 \\
0 & 1-p & 0 \\
1 & 0 & 1-p \\
2 & 0 & p \\
\end{array}
\]
eg 3. Pentagon channel
\[ X = Y = \{1, 2, 3, 4, 5\} \text{, input } \begin{cases} \text{sent up } & \text{1-}p \\ \text{shifted up } & \text{mod Supp} \end{cases} \]

\[
\begin{array}{c c c c c}
1 & 2 & 3 & 4 & 5 \\
1 & 2 & 3 & 4 & 5 \\
1 & 2 & 3 & 4 & 5 \\
1 & 2 & 3 & 4 & 5 \\
1 & 2 & 3 & 4 & 5 \\
\end{array}
\]
General assumptions:
1. Can use channel many (n) times
2. Each use identical & independent:
   \[
   \text{For inputs } x_1, x_2, \ldots, x_n \\
   \text{outputs } y_1, y_2, \ldots, y_n \text{ with } \prod_{i=1}^{n} P(y_i = x_i)
   \]
   "Called discrete memoryless channels DMCs"

Non DMCs:

eq 1 Time vary channel: the ith use is a BSC with prob error \( p_i \)

eq 2 Burst error: \( x_1, x_2, \ldots, x_n \longrightarrow x_1, x_2, \ldots, x_n \) missing a contiguous block in the output
   "Dg eats a page from your book!"
eg 3 \( x_1 x_2 \ldots x_i x_j \ldots x_n \)

\[ \downarrow \]

\( x_1 x_2 \ldots x_j x_i \ldots x_n \)

Symbols emerging in slightly wrong order

eg 4 \( x_1 x_2 \ldots x_n \)

\[ \downarrow \]

\( y_1 y_2 \ldots y_m \quad m < n \)

“Missing messages” - don’t know which ones.

Aside: quantum analogues and coding strategies?
DMC from now on

Dealing with noise by error correcting codes:

eq 1. repeat $k$ times

0 $\rightarrow$ 00 ... 0  \quad \text{majority decoder}

1 $\rightarrow$ 11 ... 1

\uparrow \quad \uparrow

messages a code word for each message

"The code = set of code words = subset of all possible inputs"
Eq 2. Hamming codes (e.g. encode 4 bits in 7, corrects up to 1 error)

Each codeword $x$ satisfies 3 parity constraints:

$x_1 \ x_2 \ \cdots \ x_7$

$P = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$, \hspace{1cm} PX = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$, \hspace{1cm}$\text{ie } x_1 \oplus x_3 \oplus x_5 \oplus x_7 = 0$

$x_2 \oplus x_3 \oplus x_6 \oplus x_7 = 0$

etc

What's cool: if $y_i = x_i + e_i$ and only $e_i = 1$

then $Py = Pe = i$th col of $P$,

decoding / identifying the error is easy!
Geometrically: (say $X = \gamma$)

- Code words

Can recover message if code words are sparse enough so that these spheres don't overlap.
Qn: For a fixed message size, to have smaller & smaller error prob, need bigger & bigger codes ..

(1) that brings more and more errors too

(2) will the rate $\rightarrow 0$ ?

(3) for growing message size,
   will prob(every part correct) $\rightarrow 0$?

Usually (1) not a problem if prob error small enough to start with, but (2), (3) can happen, say, with the repetition code.

Will see, we can do much much better
   -- magic: iid channel use + large $n$
Sending messages through \( n \) uses of a noisy channel:

An \((M,n)\) code consists of

1. index set \( M = \{1, \ldots, M^2\} \)
2. an encoding function \( E_n : M \rightarrow X^{\otimes n} \)
3. a decoding function \( D_n : Y^{\otimes n} \rightarrow M \)

The codewords are \( E_n(1), E_n(2), \ldots, E_n(M) \).
For message $m$, there's an error if

$$M' = D_n \circ N^\omega \circ \Xi_n (m) \neq m$$

Say, happens w.p. $Pe(m)$

Define $\bar{Pe} = \text{worse case prob of error} = \max_{m \in M} Pe(m)$

$$\bar{E}_{\bar{Pe}} = \text{average} = \frac{1}{M} \sum_{m=1}^{M} Pe(m)$$

Rate of an $(M, n)$ code: $\frac{1}{n} \log M$
Def: For a channel $N$, a rate $R$ is achievable if $\exists$ sequence of $(M=\lceil 2^{nR} \rceil, n)$ codes st. $P_e^n \to 0$ as $n \to \infty$

Def: Capacity of $N$, $C(N) = \sup R$ over achievable rates

NB If $C > 0$, the entire message, longer & longer (as $n$) comes out correctly almost surely!
Thm (Shannon's noisy coding theorem)

\[ C(N) = \max_{p(x)} I(X; Y) \]

NB1. \( p(xy) = p(x) p(y|x) \)

\[ \max \text{ over specified by } N \]

NB2. Expression involves only 1 copy of \( p(xy) \)
but \( C(N) \) has an asymptotic definition.

NB3. Works in worse case, no distribution of message
"p(x)" in the max has meaning TBD.

NB4. Every channel (but one) has \( C > 0 \)!
Example 1: BSC

\[ I(X;Y) = H(Y) - H(Y|X) \]

\[ \uparrow \]

\[ H(p) \text{ instead of } p(x) \]

Max this by making \( y \) random possible when \( p(0) = p(1) = \frac{1}{2} \).

\[ \therefore \text{Capacity of BSC} = 1 - H(p) \]

Example 2: Erasure Channel

\[ I(X;Y) = H(X) - H(Y|X) = (1-p) H(X) \]

Again optimal \( p H(X) \)

\[ p(x) = p(0) = p(1) = \frac{1}{2} \]

Same rate as if where the erasures are known up-front

Capacity of erasure channel = (1-p)
Eq 3. Pentagon channel (with $p = \frac{1}{2}$)

$$I(X:Y) = H(Y) - H(Y|X)$$

Always $s = 1$

Again optimal $p(x)$ uniform,

$$C(\bigcirc) = \log s - 1 = 1 \log \left(\frac{s}{2}\right)$$

Eq 1-3 very symmetric

This's optimal $p(x)$ uniform

```
\begin{tikzpicture}
    % Diagram code here
\end{tikzpicture}
```

NB If we demand $p_e = 0$, but allowing many uses, we're
studying the "Zero-error capacity" (lower bold for $C(N)$)

Eq. The BSC & erasure channel has 0 zero-error capacity

That of $\bigcirc$ is $\log s$,
That of eq 4 is 1.

Comparing $\bigcirc$ with $E_{10^{-10}}^{5}$ (erasure channel with $1X1=5$, $p = 10^{-10}$)

$C(E_{10^{-10}}^{5}) \approx \log 5 > C(\bigcirc)$
But zero error capacity of $E_{10^{-10}}^{5} = 0$

< that of $\bigcirc$
Back to $C(N) = \max \sum_{x=\gamma} \rho(x)$