Converse of the Noisy Coding Theorem (NOT in MacKay)

"If $\tilde{R} > C(Q)$: $\exists \delta > 0$ $\exists N_0$ $\forall N \geq N_0$: $\nexists$ code with $\frac{K}{N} \geq \tilde{R}$ & $p_B \leq \delta$"

Tools: ① **Data Processing Inequality** (DPI) for $A \to B \to C$ Markov chain:

$$I(B:C) \geq I(A:C) \quad \& \quad H(A|B) \leq H(A|C)$$

Same for $A \to C$ interchanged! ie- $P(a,b,c) = P(a) P(b|a) P(c|b) = P(b) P(a|b) P(c|b)$

② If X^N arbitrary and Y^N channel output: ie- $P(x^N, y^N) = P(x^N) Q(y_1|x_1) \cdots Q(y_N|x_N)$

$$I(X^N : Y^N) \leq \sum_{i=1}^{N} I(X_i : Y_i) \leq N \cdot C(Q)$$ $\boxed{HW 5}$

③ **Fano's inequality** for $S \to T \to \hat{S}$ Markov chain, $p = \Pr(S \neq \hat{S})$

$$H(\{p, 1-p\}) + p \cdot \log \#\mathcal{A}_S \geq H(S|\hat{S}) \underset{①}{\geq} H(S|T)$$

Proof of the Converse: Consider (N, K)-code with $\frac{K}{N} \geq \tilde{R} > C$.

Let $S \in \{1, \dots, 2^K\}$ uniform. Recall: $S \to X^N \to Y^N \to \hat{S}$.

Then:

$*$ $H(S|Y^N) = H(S) - I(S:Y^N) \underset{\text{DPI ①}}{\geq} H(S) - I(X^N : Y^N) \underset{②}{\geq} K - N \cdot C$

  $S \to X^N \to Y^N$ Markov chain

$*$ $H(S|Y^N) \underset{\text{Fano ③}}{\leq} 1 + \Pr(\hat{S} \neq S) \cdot \log \#\mathcal{A}_S = 1 + p_B \cdot K$

  $S \to Y^N \to \hat{S}$ Markov chain

$\Rightarrow$ $K - N \cdot C \leq 1 + p_B \cdot K$

$\Rightarrow$ $p_B \geq \frac{1}{K}(K - N \cdot C - 1) = 1 - \frac{N \cdot C}{K} - \frac{1}{K} \geq \underset{>0}{1 - \frac{C}{\tilde{R}}} - \underset{\to 0}{\frac{1}{N\tilde{R}}}$ $\square$

Can never go below this for large enough N

Are we happy? What questions does Shannon's theorem leave unaddressed? algorithmics, large N, ... how to even compute C?

Shannon's Theorem vs. Practice (§11)

- Need large block size N for joint typicality ~~vs. fixed packet size~~ → HW 5
- Codebook $X^N(1),\ldots, X^N(2^K)$ exponentially large in N (if $R>0$)
- Random codes vs predictable performance

A family of codes is "<u>very good</u>" if $\frac{K}{N} \to C$ & $P_B \to 0$
"<u>good</u>" if $\frac{K}{N} \geq \tilde{R}$ & $P_B \to 0$ for some $\tilde{R} > 0$
"<u>bad</u>" otherwise

...and "**practical**" if $\exists$ efficient encoder + decoder

Often run in embedded devices (cell phone, satellite, TV, ...) !

<u>In practice</u>:

* most codes are <u>linear</u> (X^N **linear** function of S^K) ← do you know what this means?

* "easy" to come up with "plausible" encoders — but **optimal** decoding is in general (NP) hard! ← unlike for compression!

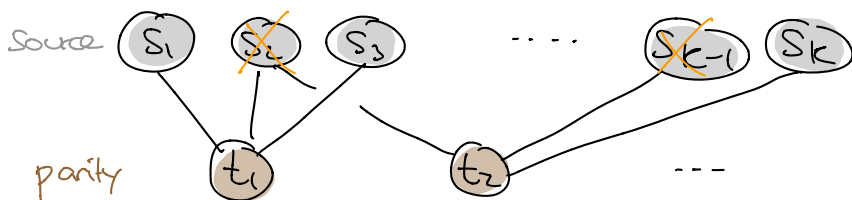$$\sigma_{opt}(y^N) = \arg\max_{\hat{S}} P(\hat{S}|y^N)$$

Why? If $P(S)$ arbitrary prior, want to choose σ to maximize $\Pr(\hat{S}=S)$
$$= \sum_{y^N} \Pr(\hat{S}=\sigma(y^N), T^N=y^N)$$
choose $S=\sigma(y^N)$ that maximizes $P(S|y^N) \propto P(S|y^N)$

e.g. imagine the following (LDPC) code:

e.g. 4 bits per parity constraint, each bit in 3 parity constraints

For **erasure channel**:
$S_1 \oplus \cancel{S_2} \oplus S_3 \oplus t_1 = 0$
$S_2 \oplus \ldots \oplus \cancel{S_{K-1}} \oplus t_2 = 0$
---

* types of decoders: "algebraic" vs. "iterative"

<u>Types of codes</u>:

* block codes: e.g. Hamming, Reed-Solomon, LDPC codes
  Storage, bar codes, Sat comm
  TODAY   WiFi, DVB, ...

* Convolutional: e.g. turbo codes
  3G/4G/LTE, Sat. comm.
  ~ linear streaming codes