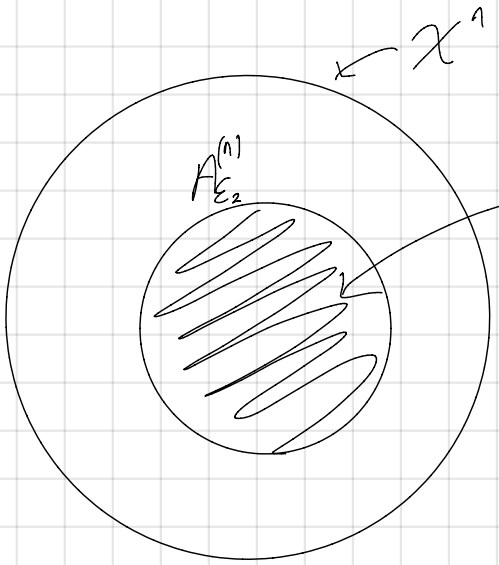


$$|\mathcal{X}^n| = |\mathcal{X}|^n = 2^{n \log |\mathcal{X}|}$$

$$|A_{\varepsilon_2}^{(n)}| \leq 2^{n(H(X)+\varepsilon)}$$

As long as $R > H(X) + \varepsilon_2$ is achievable

→ You can turn this into 0-error scheme by putting an indicator bit: If typical indicator bit is 1 and we do AEP encoding. If indicator is 0 then we encode the whole sequence x^n with $n \log |\mathcal{X}|$ and Average length: $\mathbb{P}[A_\varepsilon^{(n)}](n(H(X)+\varepsilon)) + (1 - P(A_\varepsilon^{(n)}))(n \log |\mathcal{X}|)$

What about a converse result?

Given n, enc., dec. s.t. $\mathbb{P}[X^n \neq \hat{X}^n] < \varepsilon$ for ε arbitrarily small.

$$X^n - M - \hat{X}^n$$

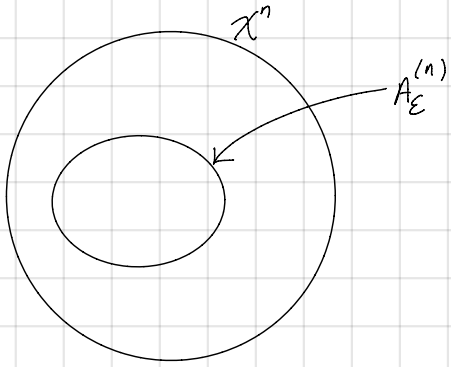
$$nR \geq H(M) \geq I(X^n; M) \geq I(X^n; \hat{X}^n) = H(X^n) - H(X^n|\hat{X}^n) = nH(X) - H(X^n|\hat{X}^n) \geq nH(X) - H(\varepsilon) - \varepsilon n \log |\mathcal{X}|$$

(Data Proc. ineq.)

$$\Rightarrow R \geq H(X) - \varepsilon \log |\mathcal{X}| - \frac{H(\varepsilon)}{n} \quad \text{as } \varepsilon \to 0 \quad R \geq H(X)$$

09/29/2016
Thursday

Recall:

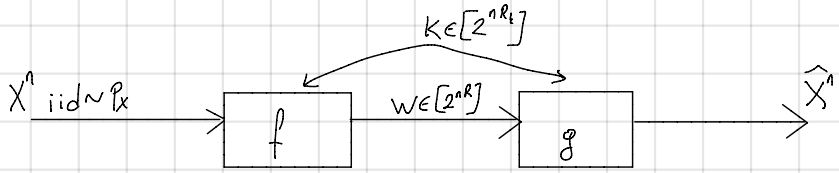


Contains all the probability
Size $2^{nH(X)}$

Let $B_\delta^{(n)}$ be the smallest set $\mathbb{P}[B_\delta^{(n)}] > 1 - \delta$.

Let $\varepsilon_n \to 0$ $\quad |B_\delta^{(n)}| \doteq |A_\varepsilon^{(n)}|$ $\quad a^n \doteq b^n$ if $\lim_{n\to\infty} \frac{1}{n} \log \frac{a_n}{b_n} = 0$

Shannon Cipher with Stochastic Source and Small Error.



Assumptions: $X^n \perp\!\!\!\perp K$, and $K \sim$ Uniform

Design: Block length n

Encoder: $f: \mathcal{X}^n \times \mathcal{K} \to \mathcal{W}$
Decoder: $g: \mathcal{W} \times \mathcal{K} \to \mathcal{X}^n$
} Randomized functions are allowed

$X^n - (K, W) - \hat{X}^n$

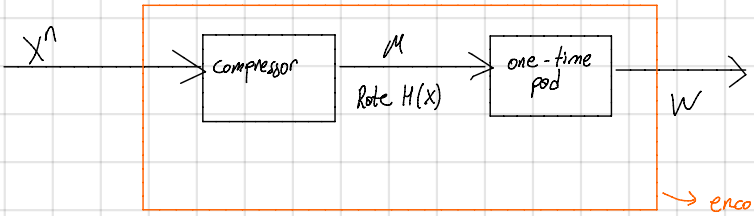
Goal: Reliability: $\mathbb{P}[X^n \neq \hat{X}^n]$ small.
Secrecy: $W \perp\!\!\!\perp X^n$ (Perfect secrecy assumption)

Achievable rate: (R, R_k) is achievable if $\forall \varepsilon > 0$ $\exists n, f, g$ (at rates (R, R_k)) such that $\mathbb{P}[X \neq \hat{X}] < \varepsilon$ and $W \perp\!\!\!\perp X^n$.

Theorem: The closure of the set of achievable rates (R, R_k) is

$$\left\{ (R, R_k): \begin{array}{l} R \geq H(X) \\ R_k \geq H(X) \end{array} \right.$$

Proof (Achievability:) First compress to rate $H(X) + \delta$, then use one-time pad.



(Converse:) Assume (R, R_k) is achievable. For arbitrary $\varepsilon > 0$, $\exists n, f, g$ gives objective.

$$\begin{aligned} nR &\geq H(W) && \text{(uniform dist. bd)} \\ &\geq H(W|K) \\ &\geq I(X^n; W|K) && \text{(} +I(X^n;K) = 0 \text{)} \\ &= I(X^n; W, K) \\ &\geq I(X^n; \hat{X}^n) \\ &= H(X^n) - H(X^n|\hat{X}^n) \\ &= nH(X) - H(X^n|\hat{X}^n) && (X^n \text{ i.i.d}) \end{aligned}$$

$(H(X^n|\hat{X}^n) \leq \varepsilon n \log|\mathcal{X}| + H(\varepsilon))$

$$\Rightarrow R \geq H(X) - \varepsilon \log|\mathcal{X}| + H(\varepsilon)$$

Let $\varepsilon \to 0$ $\qquad \delta(\varepsilon) \to 0$ as $\varepsilon \to 0$

$$R \geq H(X)$$

$$
\begin{aligned}
nR_k &= H(K) \\
&\geq H(K|W) \\
&\geq I(X^n; K|W) \\
&= I(X^n; K, W) - I(X^n; W) \\
&\quad \vdots \\
&\Rightarrow R_k \geq H(X)
\end{aligned}
$$

$X^n \perp\!\!\!\perp W$ because of secrecy assumption

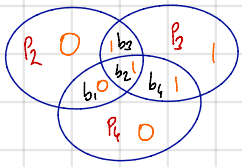
Note: Instead of $W \perp\!\!\!\perp X^n$, we could ask for a looser secrecy criterion i.e. $\frac{1}{n} I(X^n; W) < \varepsilon$. The same proof still holds.

aka. "Information leakage rate"

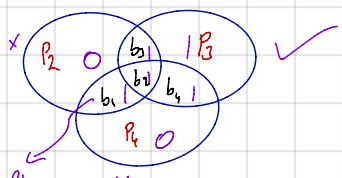
Channel Capacity:

AM, FM radio, Digital

Before Shannon: Hamming (4, 7)
(4 info bits, 7 code bits)



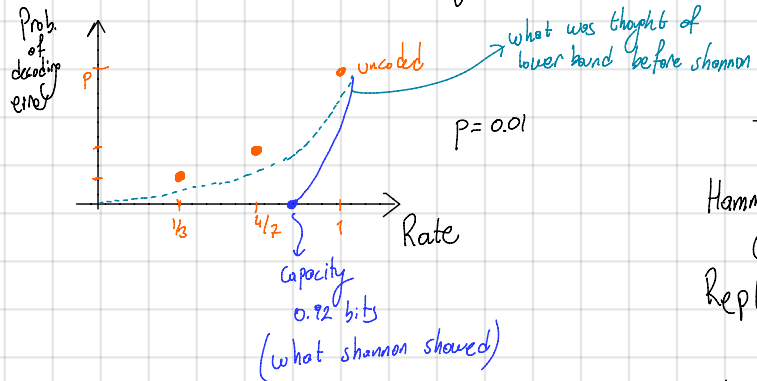
b_1	b_2	b_3	b_4	P_1	P_2	P_3	
0	1	1	1	0	0	1	you send
1	1	1	1	0	0	1	you get



→ provided that there is no more than single bit flip Hamming code can always correct itself.

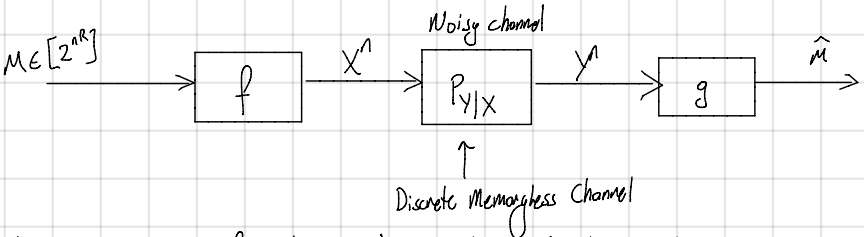
Repetition code:
Repetition (1,3) : 111
000

Bits are flipped with probability p.



	P_{error}	Rate
Hamming (4,7)	$21p^2 + O(p^3)$	4/7
Rep(1,3)	$3p^2 + p^3$ $12p^2 + O(p^3)$	1/3
Uncoded	p	1

General Setting for Channel Coding



We assume finite alphabets $|\mathcal{X}|, |\mathcal{Y}|$

$\forall i,\ P_{Y_i|X^i Y^{i-1}}(y_i | x^i, y^{i-1}) = P_{Y_i|X_i}(y_i|x_i)$

(Note that we are not saying $P_{Y^n|X^n} = \prod P_{Y|X}$ for memoryless channel in general)

For a channel w/out feedback: Memorylessness $\Leftrightarrow P_{Y^n|X^n} = \prod_{i=1}^{n} P_{Y_i|X_i}$

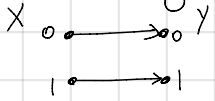
Let M be uniformly distributed

Defn: R is achievable if $\forall \epsilon > 0$, $\exists n, f, g$ at rate R such that $\mathbb{P}[M \neq \hat{M}] < \epsilon$

Defn (Capacity): Let the capacity C be the supremum of achievable rates R.

Theorem: $C = \max_{P_X} I(X;Y)$ ← with respect to $P_X P_{Y|X}$ (Problem statement gives this)

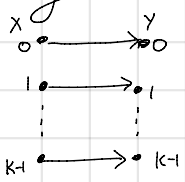
Example: Binary noiseless channel



$I(X;Y) = H(X)$

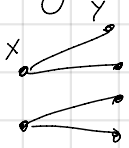
$C = \max_{P_X} I(X;Y) = 1$ bit

K-ary noiseless channel



$C = \max_{P_X} I(X;Y) = \max_{P_X} H(X) = \lg K$ bits

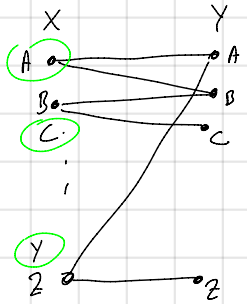
Non-overlapping Noise



$C = \max_{P_X} H(X) = \lg |\mathcal{X}|$

Example 2:

Noisy Typewriter:



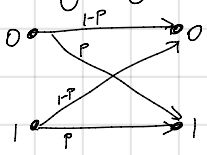
$$C = \max_{P_X} I(X;Y) = \max_{P_X} \big(H(Y) - \underbrace{H(Y|X)}_{\text{1-bit}}\big)$$

$$= \max_{P_X} H(Y) - 1$$

$$= \log_2 26 - 1 = \log_2 13 \text{ bits}$$

Example 3:

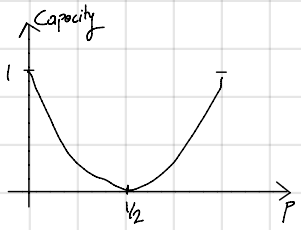
Binary Symmetric Channel



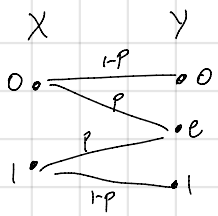
$$C = \max_{P_X} \big(H(Y) - H(Y|X)\big)$$

$$= \max_{P_X} H(Y) - H(p)$$

$$= 1 - H(p)$$



Example 4: Binary Erasure Channel



$$C = \max_{P_X} \big(H(Y) - \underbrace{H(Y|X)}_{H(p)}\big)$$

$$= \max_{P_X} \big(\underbrace{H(X)}_{1} - \underbrace{H(X|Y)}_{p\,\underbrace{H(X|Y=e)}_{H(P_X)}}\big)$$

$$= (1-p) \max_{P_X} H(X)$$

$$= 1-p$$