Quantum Channels and their capacities

Graeme Smith, IBM Research

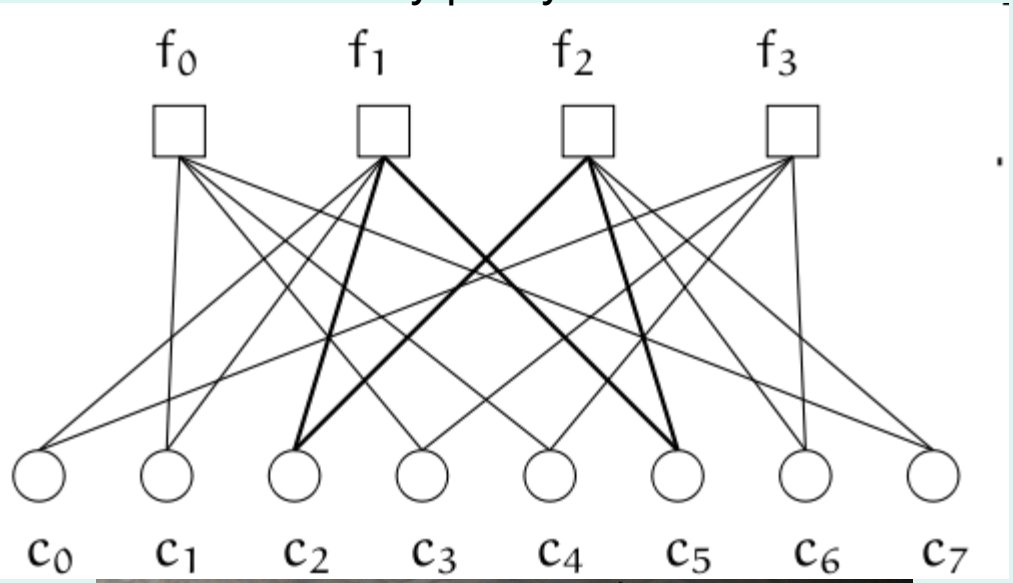
11th Canadian Summer School on Quantum Information

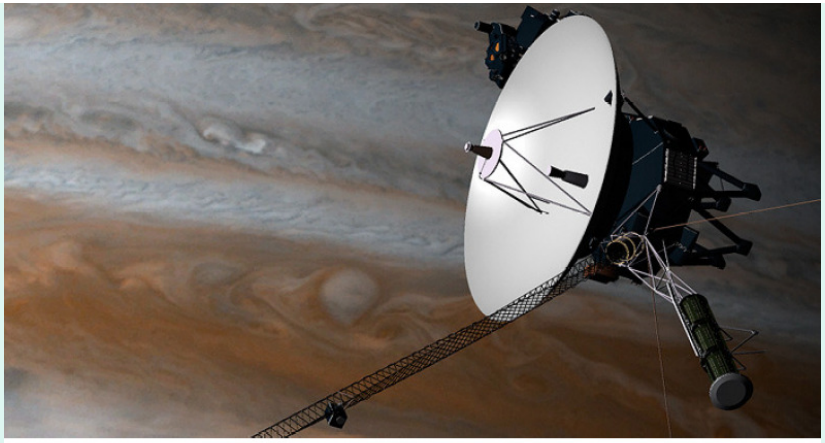
Information Theory

- “A Mathematical Theory of Communication”, C.E. Shannon, 1948
- Lies at the intersection of Electrical Engineering, Mathematics, and Computer Science
- Concerns the reliable and efficient storage and transmission of information.

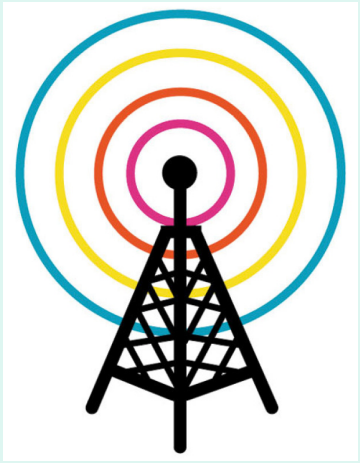
Information Theory: Some Hits

Low density parity check codes





Voyager (Reed Solomon codes)



Cell Phones

Lempel-Ziv compression (gunzip, winzip, etc)

Quantum Information Theory

When we include quantum mechanics (which was there all along!) things get much more interesting!

Secure communication, entanglement enhanced communication, sending quantum information,…

Capacity, error correction, compression, entropy..

Outline

- Lecture 1:
  Classical Information Theory
- Lecture 2:
  Quantum Channels and their many capacities
- Lecture 3:
  Advanced Topics--- Additivity, Partial Transpose, LOCC,Gaussian Noise,etc
- Lecture 4:
  Advanced Topics---Additivity, Partial Transpose, LOCC,Gaussian Noise,etc.

Example: Flipping a biased coin

Let's say we flip n coins.
They're independent and identically distributed (i.i.d):

$$\Pr(X_i = 0) = 1\text{-}p \qquad \Pr(X_i = 1) = p$$

$$\Pr(X_i = x_i, X_j = x_j) = \Pr(X_i = x_i) \Pr(X_j = x_j)$$

$$X_1 X_2 \ldots X_n$$

Q: How many 1's am I likely to get?

Example: Flipping a biased coin

Let's say we flip n coins.
They're independent and identically distributed (i.i.d):

$$\Pr(X_i = 0) = 1-p \qquad \Pr(X_i = 1) = p$$

$$\Pr(X_i = x_i, X_j = x_j) = \Pr(X_i = x_i) \Pr(X_j = x_j)$$

$$X_1 X_2 \ldots X_n$$

Q: How many 1's am I likely to get?

A: Around pn and, with very high probability between $(p-\delta)n$ and $(p+\delta)n$

Shannon Entropy

Flip n i.i.d. coins, Pr($X_i = 0$) = p, Pr($X_i = 1$) = 1-p
Outcome: $x_1 \ldots x_n$.

w.h.p. get ≈ pn 1's, but how many different configurations?

There are $\binom{n}{pn} = \frac{n!}{(pn)!((1-p)n)!}$ such strings.

Using $\log n! = n \log n - n + O(\log n)$ we get

$$\begin{aligned} \log \binom{n}{pn} &\approx n \log n - n - pn \log(pn) + pn - (1-p)n \log((1-p)n) + (1-p)n \\ &= nH(p) \end{aligned}$$

Where H(p) = -plogp – (1-p)log(1-p)

So, now, if I want to transmit x_1…x_n, I can just check which typical sequence, and report that! Maps n bits to nH(p)

Similar for larger alphabet: $H(X) = \sum_x -p(x) \log p(x)$

Shannon Entropy

Flip n i.i.d. coins, Pr($X_i = x_i$) = $p(x_i)$
$x_1 \ldots x_n$ has prob $P(x_1 \ldots x_n) = p(x_1) \ldots p(x_n)$

$$\log P(x_1 ... x_n) = \sum_{i=1}^{n} \log p(x_i)$$

Typically, $-\frac{1}{n} \log P(x_1 ... x_n) \approx -\langle \log p(x) \rangle = H(X)$

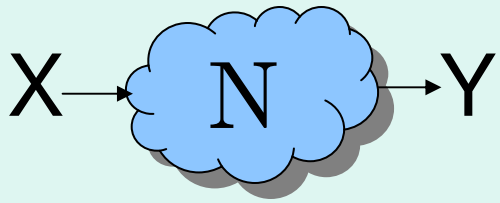
So, typical sequences have $P(x_1 ... x_n) \sim \frac{1}{2^{n(H(X) \pm \delta)}}$

More or less uniform dist over typical sequences.

Correlations and Mutual Information

H(X) – bits of information transmitted by letter from ensemble X

Noisy Channel: p(y|x)



(X,Y) correlated. How much does Y tell us about X?

After I get Y, how much more do you need to tell me so that I know X?

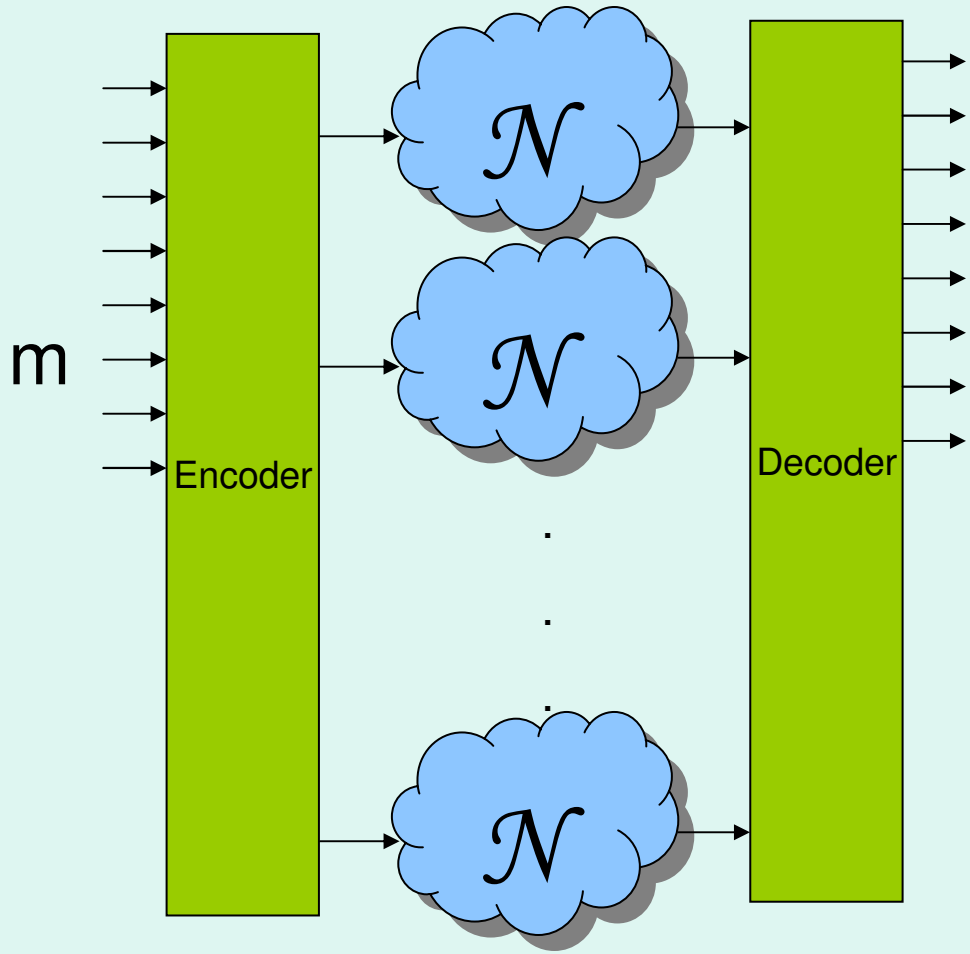
Given y, update expectations: $p(x|y) = \frac{p(y|x)p(x)}{p(y)}$

Only $H(X|Y) = \langle -\log p(x|y) \rangle$ bits per letter needed.

Can calculate $H(X|Y) = \langle -\log \frac{p(x,y)}{p(y)} \rangle = H(X,Y) - H(Y)$

Savings: H(X)–H(X|Y) = H(X) + H(Y) – H(X,Y) =: I(X;Y)

Channel Capacity



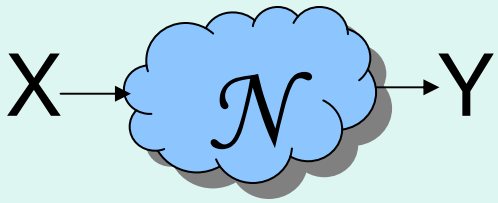
Given n uses of a channel, encode a message $m \in \{1,\ldots,M\}$ to a codeword $x^n = (x_1(m),\ldots, x_n(m))$

At the output of the channel, use $y^n = (y_1,\ldots,y_n)$ to make a guess, m'.

The rate of the code is $(1/n)\log M$.

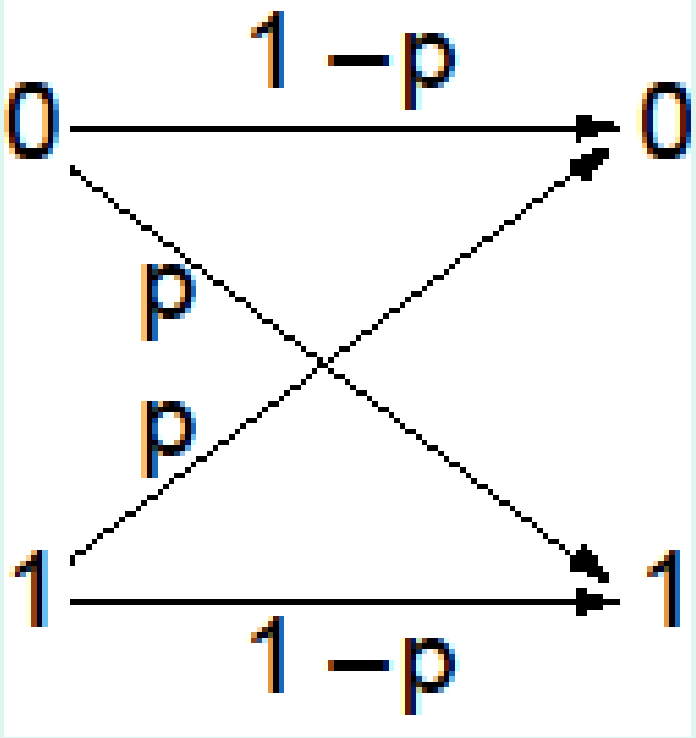
The capacity of the channel, C(N), is defined as the maximum rate you can get with vanishing error probability as $n \to \infty$

Binary Symmetric Channel

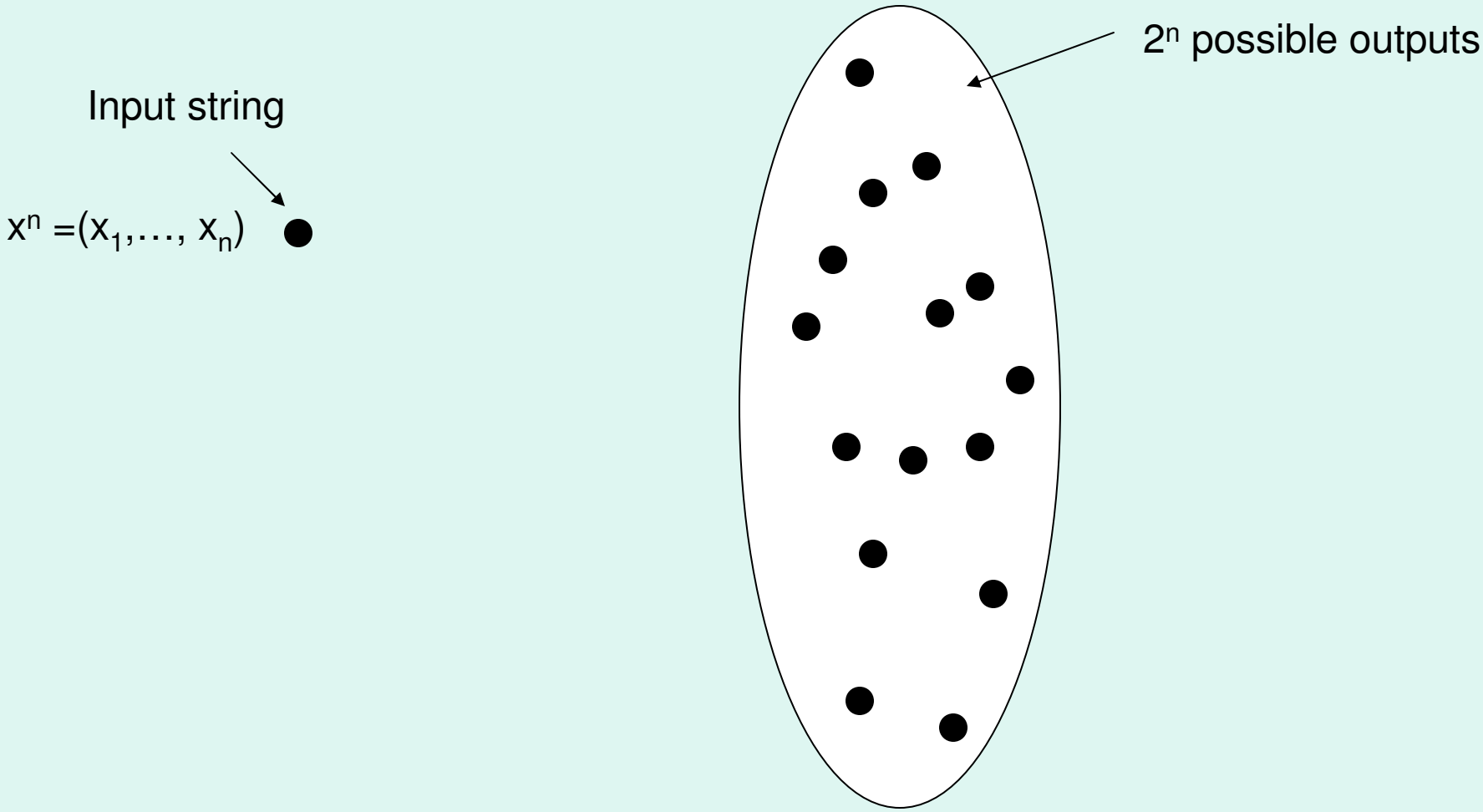


p(0|0) = 1-p p(1|0) = p

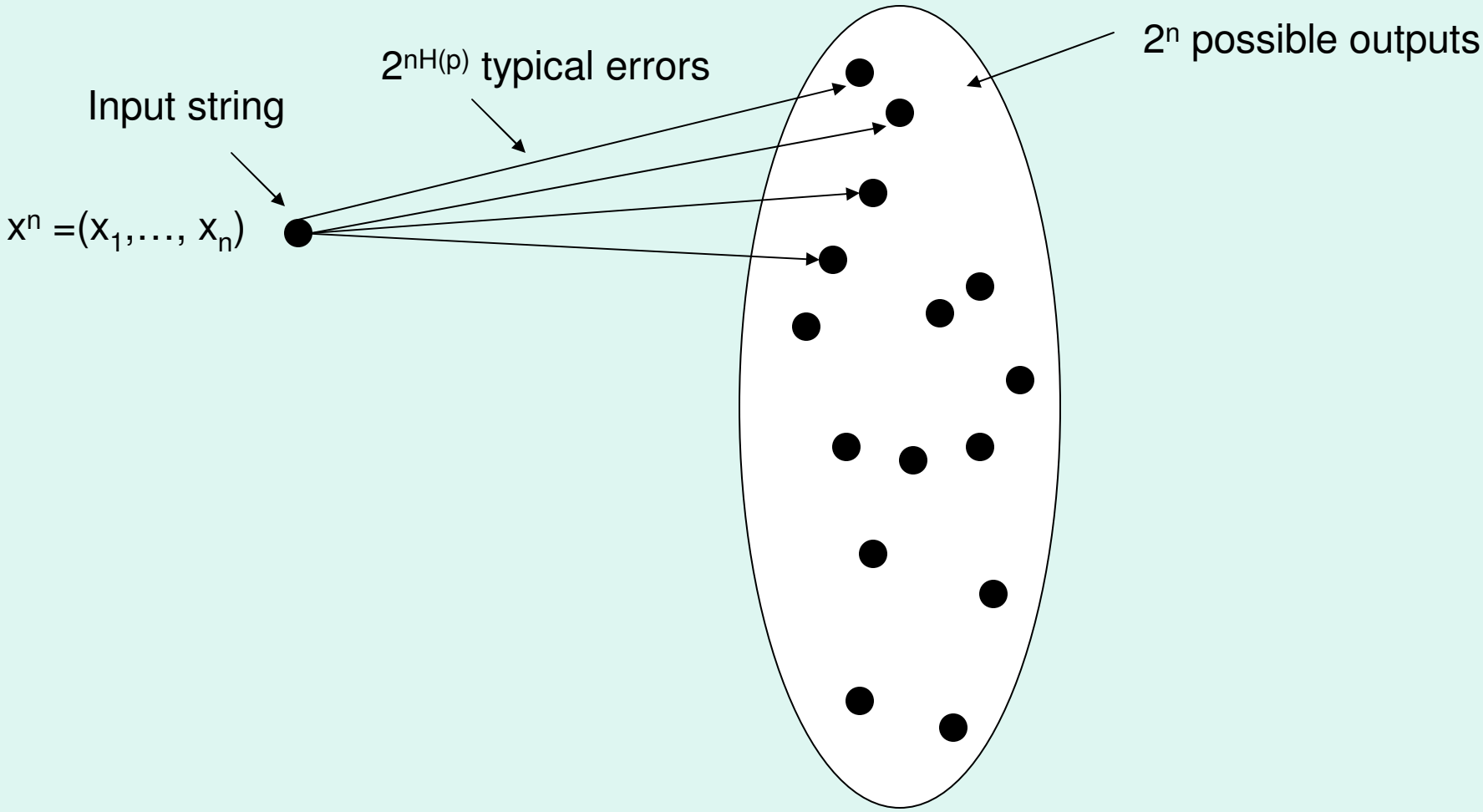
p(0|1) = p p(1|1) = 1-p



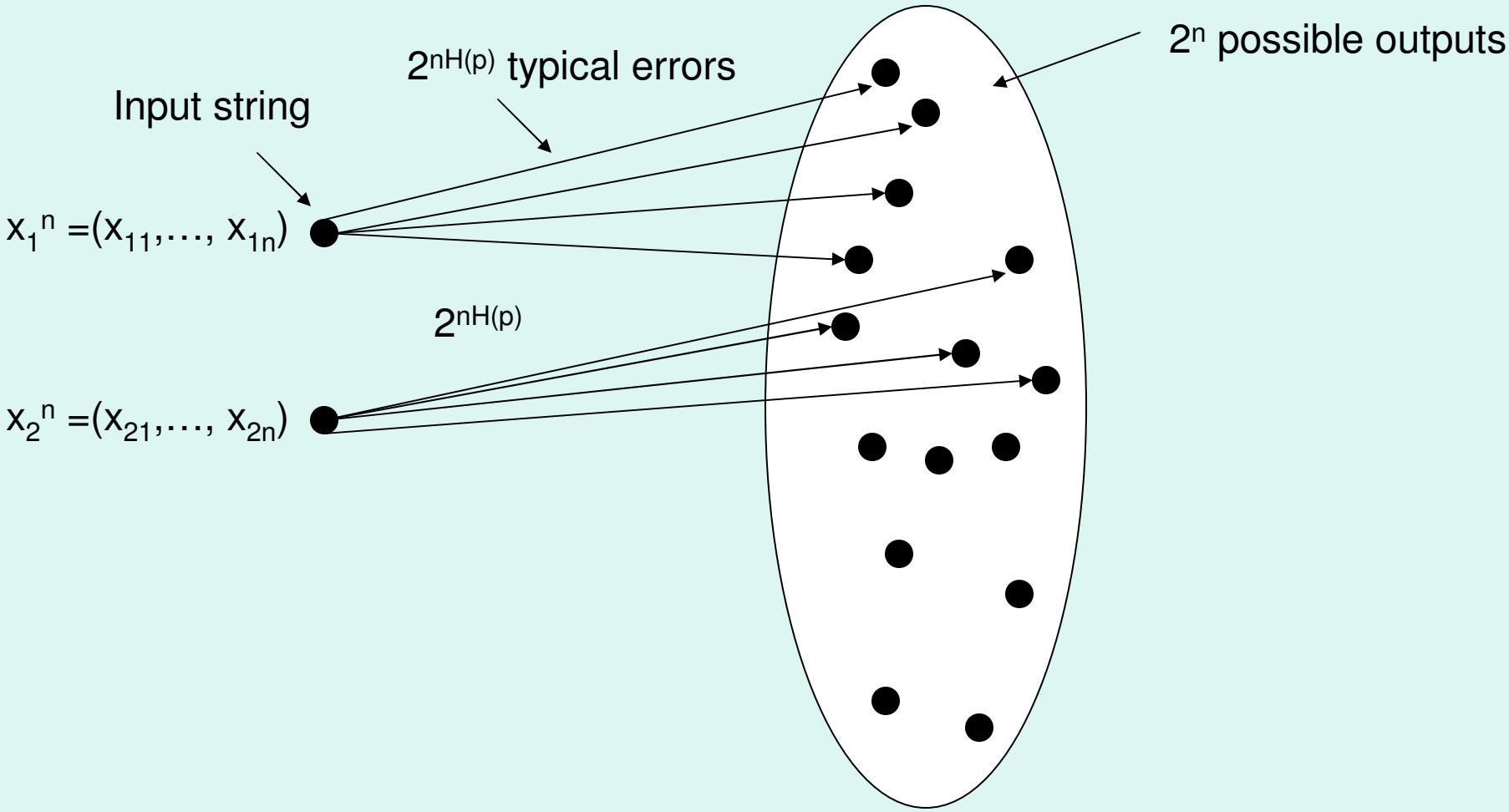
Capacity of Binary Symmetric Channel



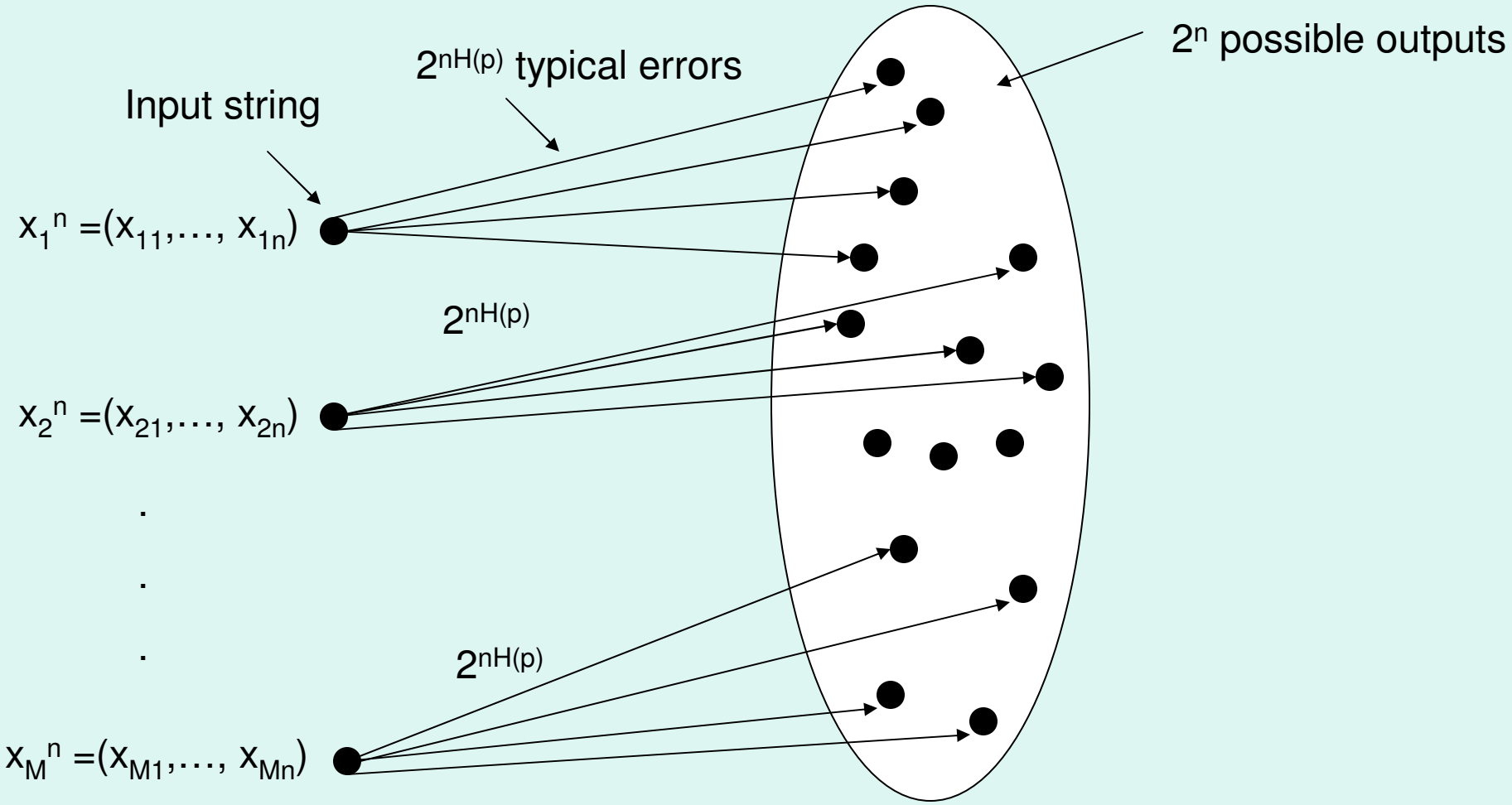
Capacity of Binary Symmetric Channel



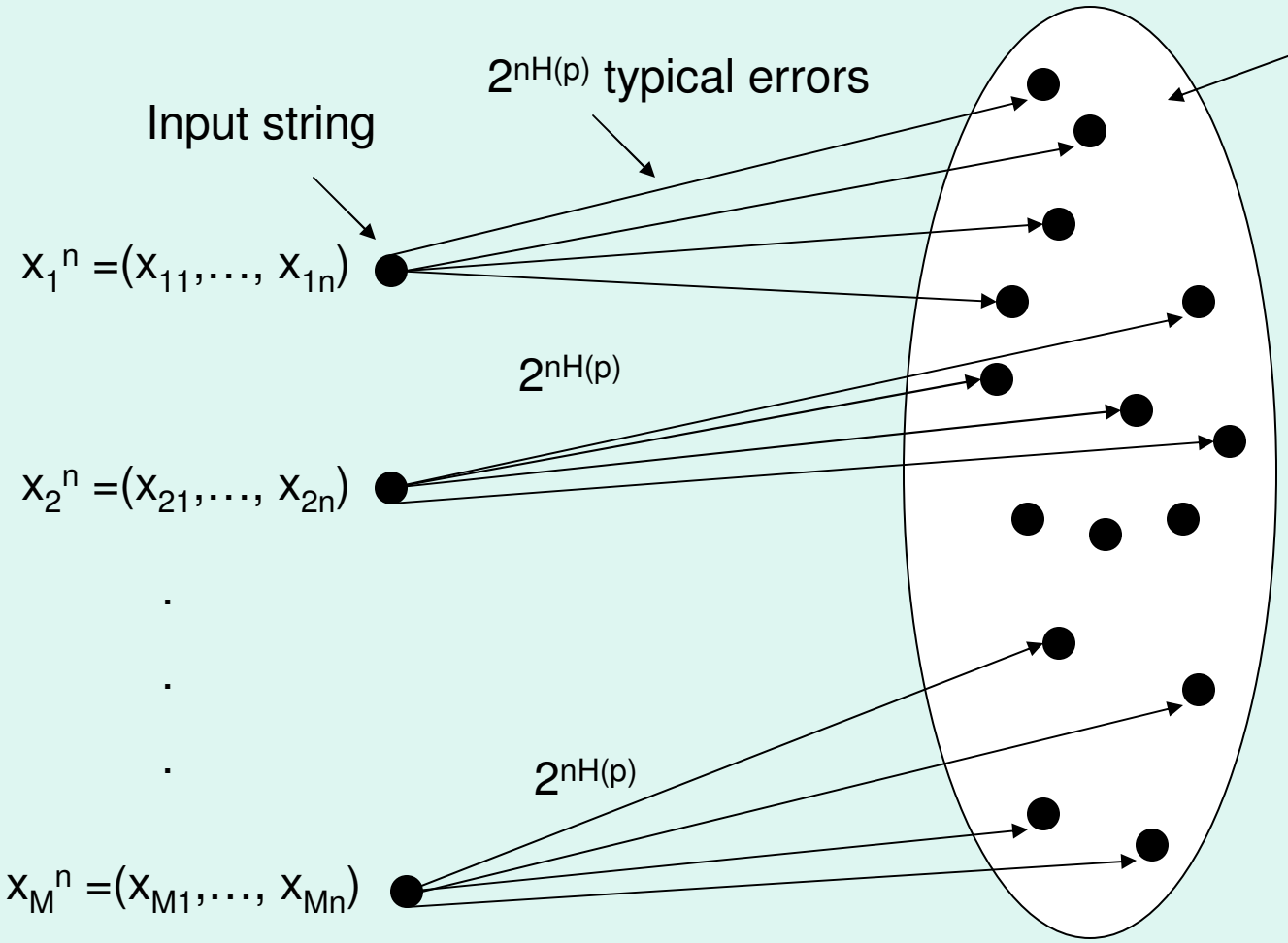
Capacity of Binary Symmetric Channel



Capacity of Binary Symmetric Channel



Capacity of Binary Symmetric Channel



2^n possible outputs

Each x_m^n gets mapped to $2^{nH(p)}$ different outputs.

If these sets overlap for different inputs, the decoder will be confused.

So, we need

$M\, 2^{nH(p)} \leq 2^n$, which implies

$(1/n)\log M \leq 1\text{-}H(p)$

Upper bound on capacity

Direct Coding Theorem: Achievability of 1-H(p)

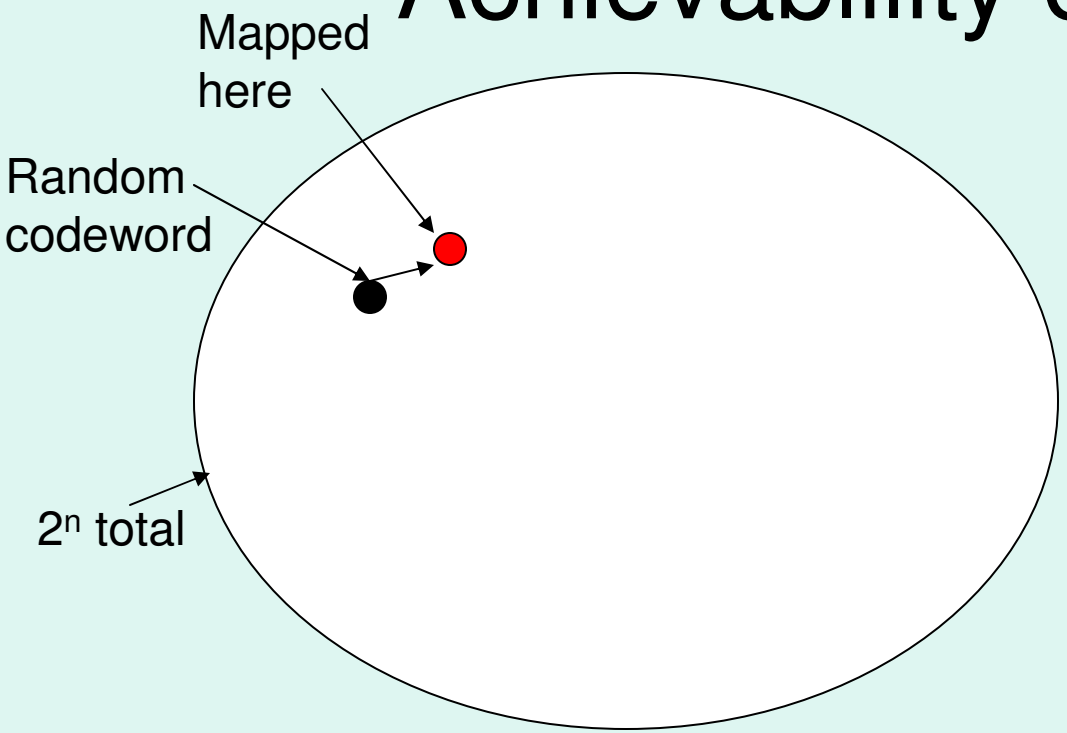
(1) Choose 2^{nR} codewords randomly according to X^n (50/50 variable)

(2) $x_m{}^n \rightarrow y^n$. To decode, look at all strings within $2^{n(H(p)+\delta)}$ bit-flips of y^n. If this set contains exactly one codeword, decode to that. Otherwise, report error.

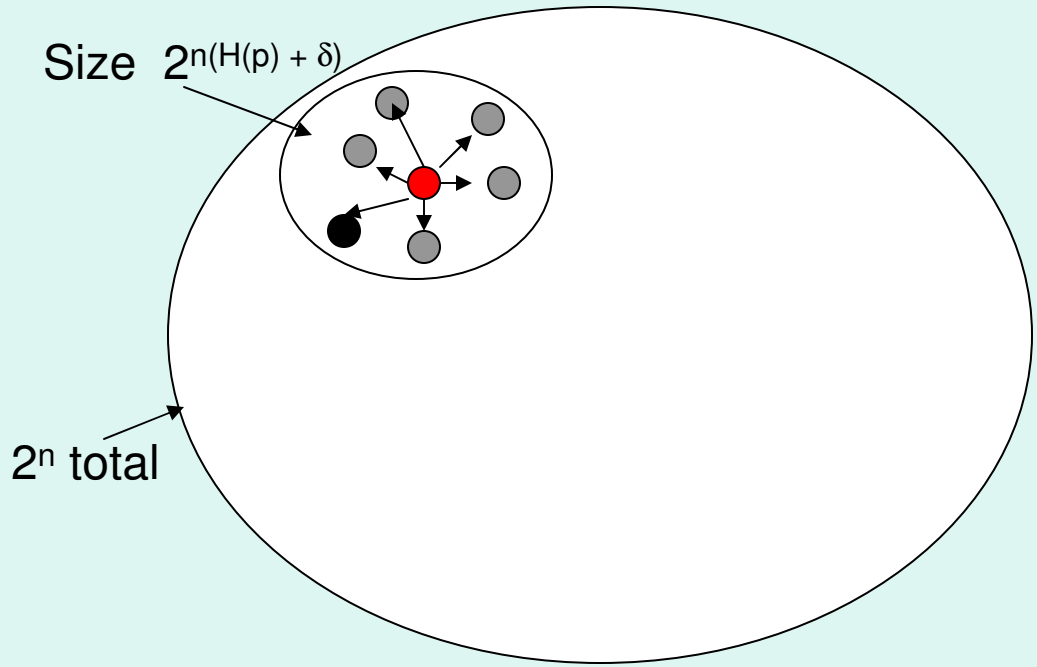
Decoding sphere is big enough that w.h.p. the correct codeword $x_m{}^n$ is in there.

So, the only source of error is if **two** codewords are in there. What are the chances of that???

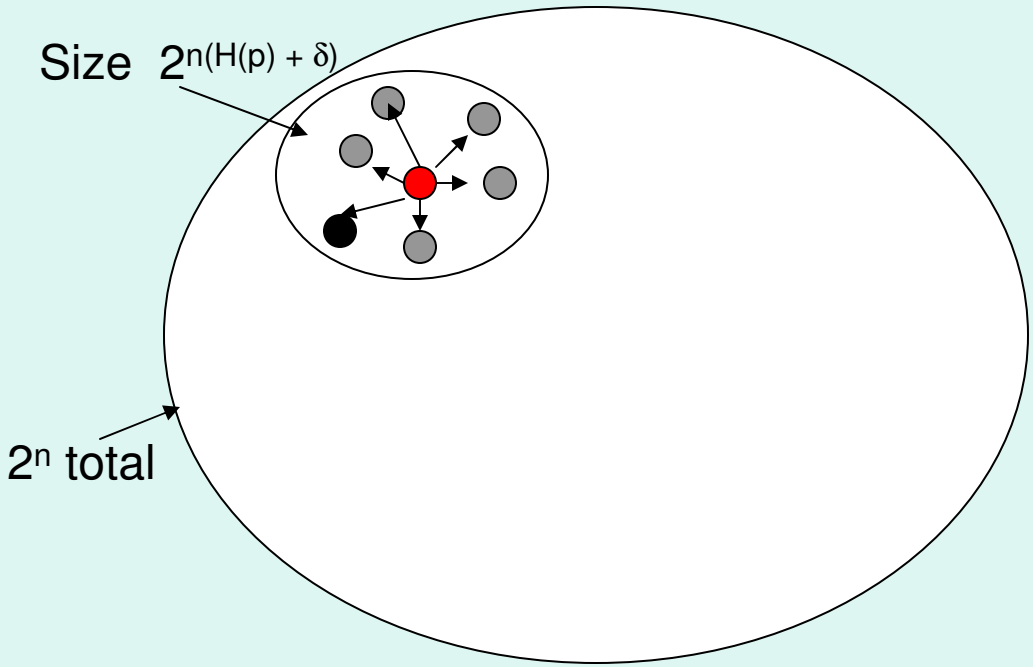
Direct coding theorem: Achievablility of 1-H(p)



Direct coding theorem: Achievablility of 1-H(p)

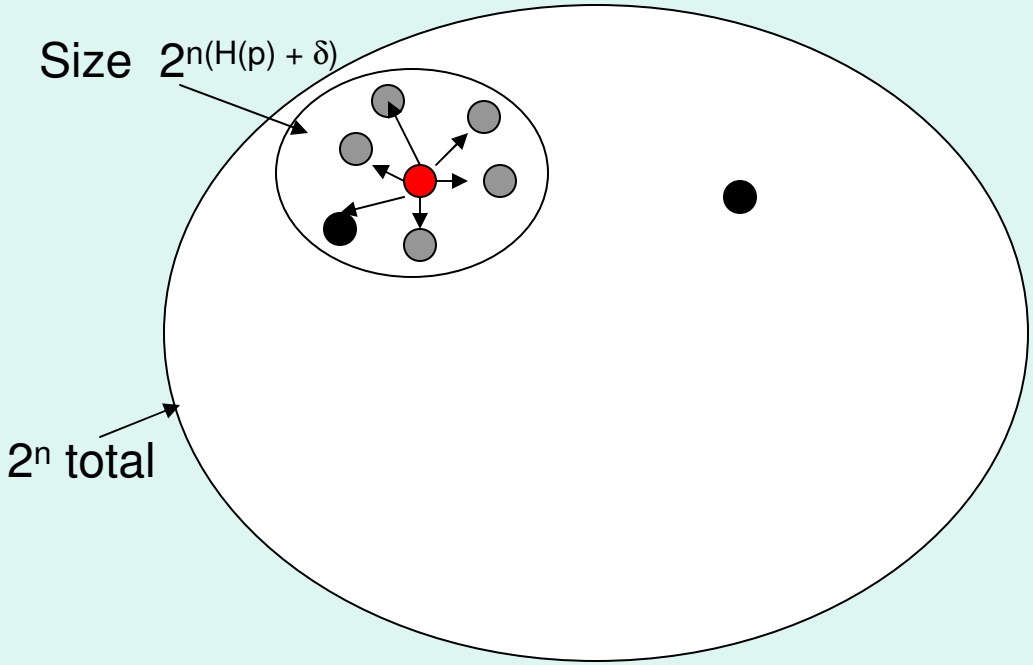


Direct coding theorem: Achievablility of 1-H(p)



If code is chosen randomly, what's the chance of another codeword in this ball?

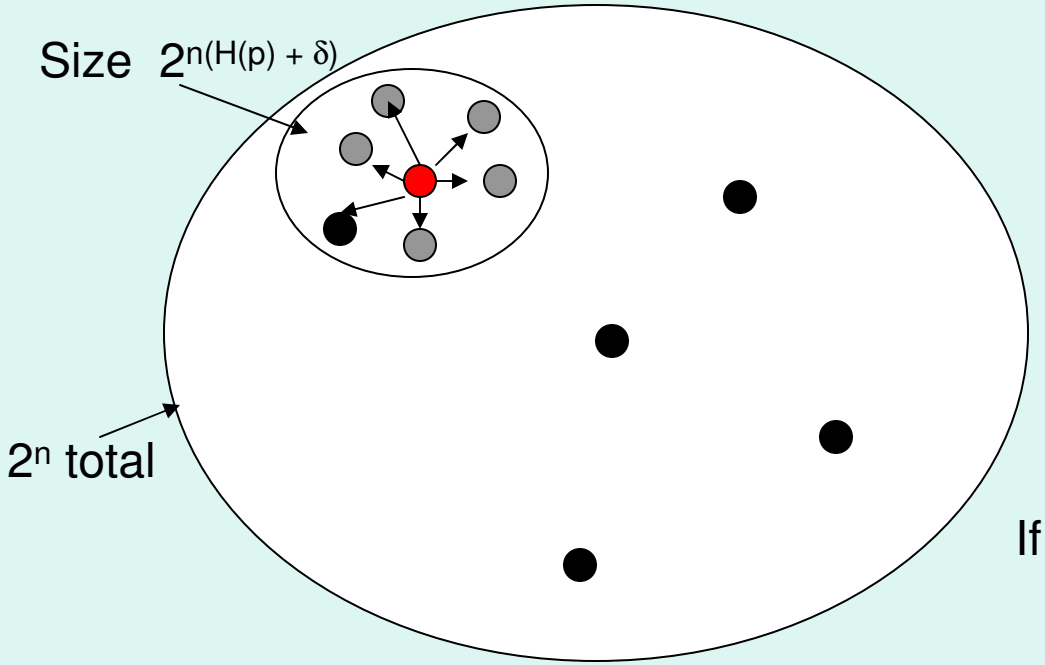
Direct coding theorem: Achievablility of 1-H(p)



If code is chosen randomly, what's the chance of another codeword in this ball?

If I choose one more word, the chance is $\frac{2^{n(H(p)+\delta)}}{2^n}$

Direct coding theorem: Achievablility of 1-H(p)



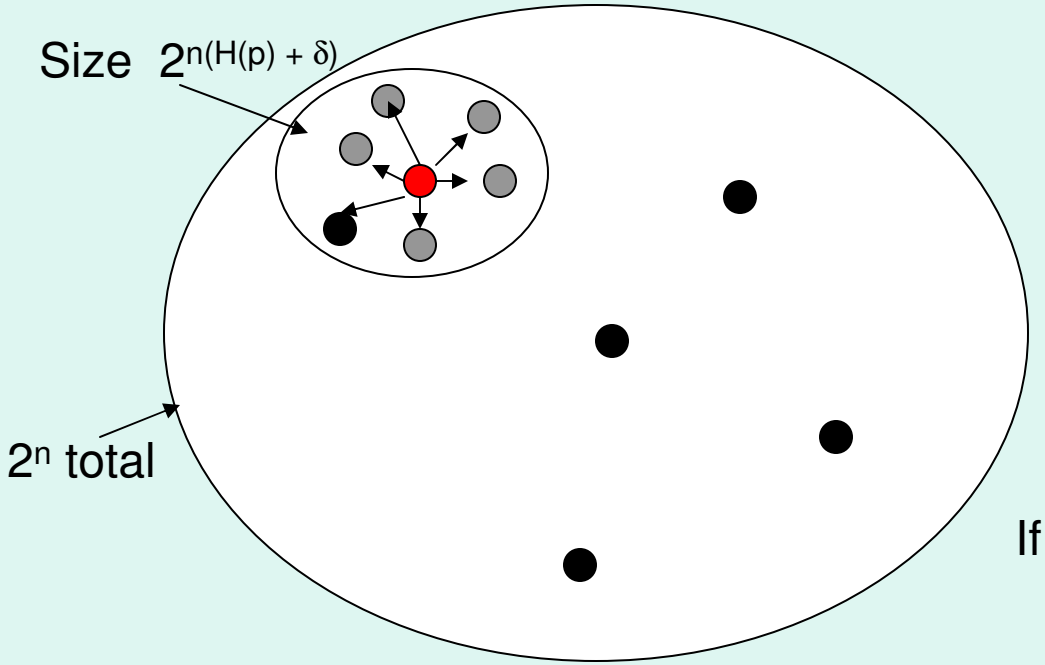
If code is chosen randomly, what's the chance of another codeword in this ball?

If I choose one more word, the chance is $\frac{2^{n(H(p)+\delta)}}{2^n}$

Choose 2^{nR} more, the chance is $\frac{2^{n(H(p)+R+\delta)}}{2^n}$

If $R < 1 - H(p) - d$, this $\to 0$ as $n \to \infty$

Direct coding theorem: Achievablility of 1-H(p)



If code is chosen randomly, what's the chance of another codeword in this ball?

If I choose one more word, the chance is $\frac{2^{n(H(p)+\delta)}}{2^n}$

Choose 2^{nR} more, the chance is $\frac{2^{n(H(p)+R+\delta)}}{2^n}$

If $R < 1 - H(p) - d$, this $\to 0$ as $n \to \infty$

So, the average probability of decoding error (averaged over codebook choice and codeword) is small.

As a result, there must be **some** codebook with low prob of error (averaged over codewords).

Low worst-case probability of error

Showed: there's a code with rate R such that

$$\frac{1}{2^{nR}} \sum_{i=1}^{2^{nR}} P_i < \epsilon$$

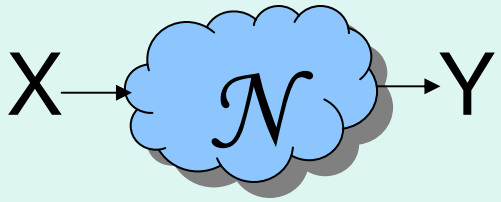
Let $N_{2\varepsilon}$ = #{i | P_i > 2ε}. Then,

$$\frac{2\epsilon N_{2\epsilon}}{2^{nR}} \leq \frac{1}{2^{nR}} \sum_{i \ st \ P_i > 2\epsilon} P_i \leq \frac{1}{2^{nR}} \sum_{i=1}^{2^{nR}} P_i < \epsilon$$

So that $N_{2\varepsilon} < 2^{nR - 1}$. Throw these away.

Gives a code with rate R – 1/n and $P_i < 2\varepsilon$ for all i.

Coding theorem: general case



- $2^{nH(X)}$ typical x^n
- For typical y^n, $\approx 2^{n(H(X|Y) + \delta)}$ candidate x^n's in decoding sphere. If unique candidate, report this.
- Each sphere of size $2^{n(H(X|Y)+\delta)}$ contains a fraction $\frac{2^{n(H(X|Y)+\delta)}}{2^{nH(X)}}$ of the inputs.
- If we choose 2^{nR} codewords, prob of accidentally falling into wrong sphere is $\frac{2^{nR}}{2^{n(H(X)-H(X|Y)-\delta)}} = \frac{2^{nR}}{2^{nI(X;Y)-\delta}}$

- Okay as long as $R < I(X;Y)$

Capacity for general channel

- We showed that for any input distribution p(x), given p(y|x), we can approach rate R = I(X;Y). By picking the best X, we can achieve C(N) = $\max_X$I(X;Y). This is called the “direct” part of the capacity theorem.
- In fact, you can’t do any better. Proving there’s no way to beat $\max_X$I(X;Y) is called the “converse”.

Converse

For homework, you'll prove two great results about entropy:

1. $H(X,Y) \leq H(X)+ H(Y)$
2. $H(Y_1Y_2|X_1X_2) \leq H(Y_1|X_1) + H(Y_2|X_2)$

We're going to use the first one to prove that no good code can transmit at a rate better than $\max_X I(X;Y)$

Converse

- Choose some code with 2^{nR} strings of n letters.
- Let $\mathbf{X^n}$ be uniform distribution on the codewords (codeword i with prob $1/2^{nR}$)
- Because the channels are independent, we get $p(y_1 \ldots y_n|x_1 \ldots x_n) = p(y_1|x_1) \ldots p(y_n|x_n)$
- As a result, the conditional entropy satisfies $H(\mathbf{Y^n}|\mathbf{X^n}) = \langle -\log p(y_1 \ldots y_n|x_1 \ldots x_n) \rangle = \sum_\iota \langle -\log p(y_i|x_i) \rangle = \sum_\iota H(\mathbf{Y_i}|\mathbf{X_i})$.
- Now, $I(\mathbf{Y^n};\mathbf{X^n}) = H(\mathbf{Y^n}) - H(\mathbf{Y^n}|\mathbf{X^n}) \leq \sum_\iota H(\mathbf{Y_i}) - H(\mathbf{Y_i}|\mathbf{X_i}) = \sum_\iota I(\mathbf{Y_i};\mathbf{X_i}) \leq n \max_X I(X;Y)$

- Furthermore, $I(\mathbf{Y^n};\mathbf{X^n}) = H(\mathbf{X^n}) - H(\mathbf{X^n}|\mathbf{Y^n}) = nR - H(\mathbf{X^n}|\mathbf{Y^n}) \leq n \max_X I(X;Y)$.
- Finally, since $\mathbf{Y^n}$ must be sufficient to decode $\mathbf{X^n}$, we must have $(1/n)H(\mathbf{Y^n}|\mathbf{X^n}) \to 0$, which means $R \leq \max_X I(X;Y)$.

Recap

- Information theory is concerned with efficient and reliable transmission and storage of data.
- Focused on the problem of communication in the presence of noise. Requires error correcting codes.
- Capacity is the maximum rate of communication for a noisy channel. Given by $\max_X I(X;Y)$.
- Two steps to capacity proof: (1) direct part: show by randomized argument that there exist codes with low probability of error. Hard part: error estimates. (2) Converse: there are no codes that do better---use entropy inequalities.

Coming up:

Quantum channels and their various capacities.

- Homework is due at the **Beginning** of next class.
- Fact II is called Jensen's inequality
- No partial credit