Quantum Information Theory and Cryptography

John Smolin, IBM Research

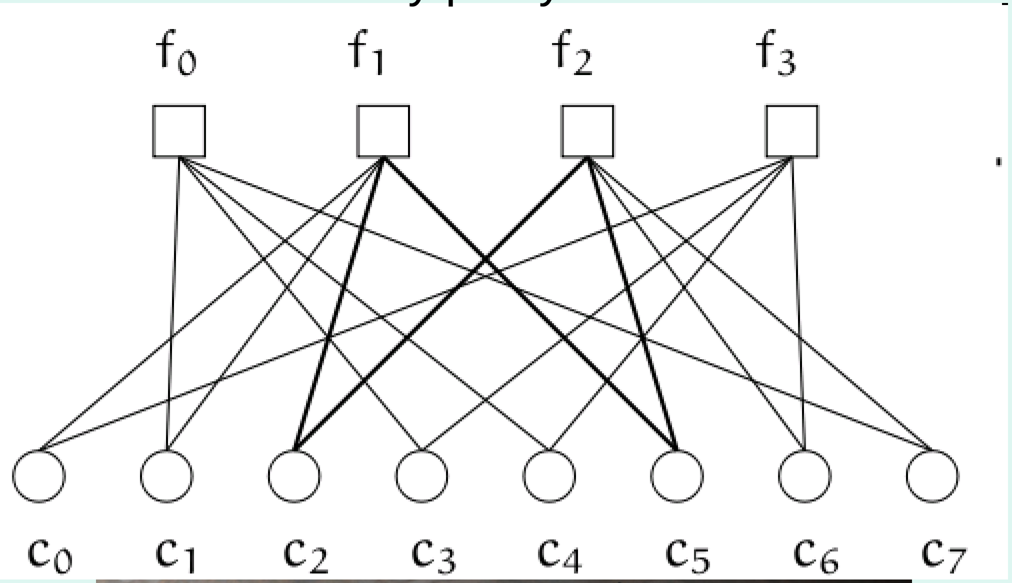
IPAM

Information Theory

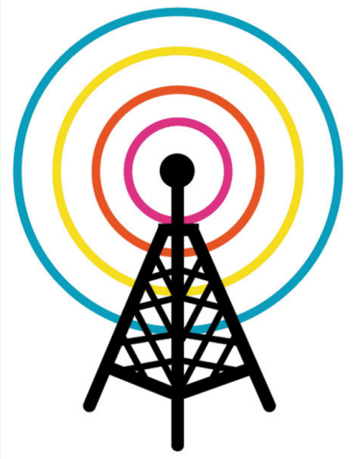
- “A Mathematical Theory of Communication”, C.E. Shannon, 1948
- Lies at the intersection of Electrical Engineering, Mathematics, and Computer Science
- Concerns the reliable and efficient storage and transmission of information.

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..

Example: Flipping a biased coin

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

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

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

$$x_1 x_2 \ldots x_n$$

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):

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

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

$$x_1 x_2 \ldots x_n$$

Q: How many 1’s am I likely to get?

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) =1- p, Pr(X_i = 1) = p
Outcome: $x_1 \ldots x_{n.}$

w.h.p. get approximately 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

$$\log \binom{n}{pn} \approx n \log n - n - pn \log(pn) + pn + (1-p)n \log((1-p)n) + (1-p)n$$

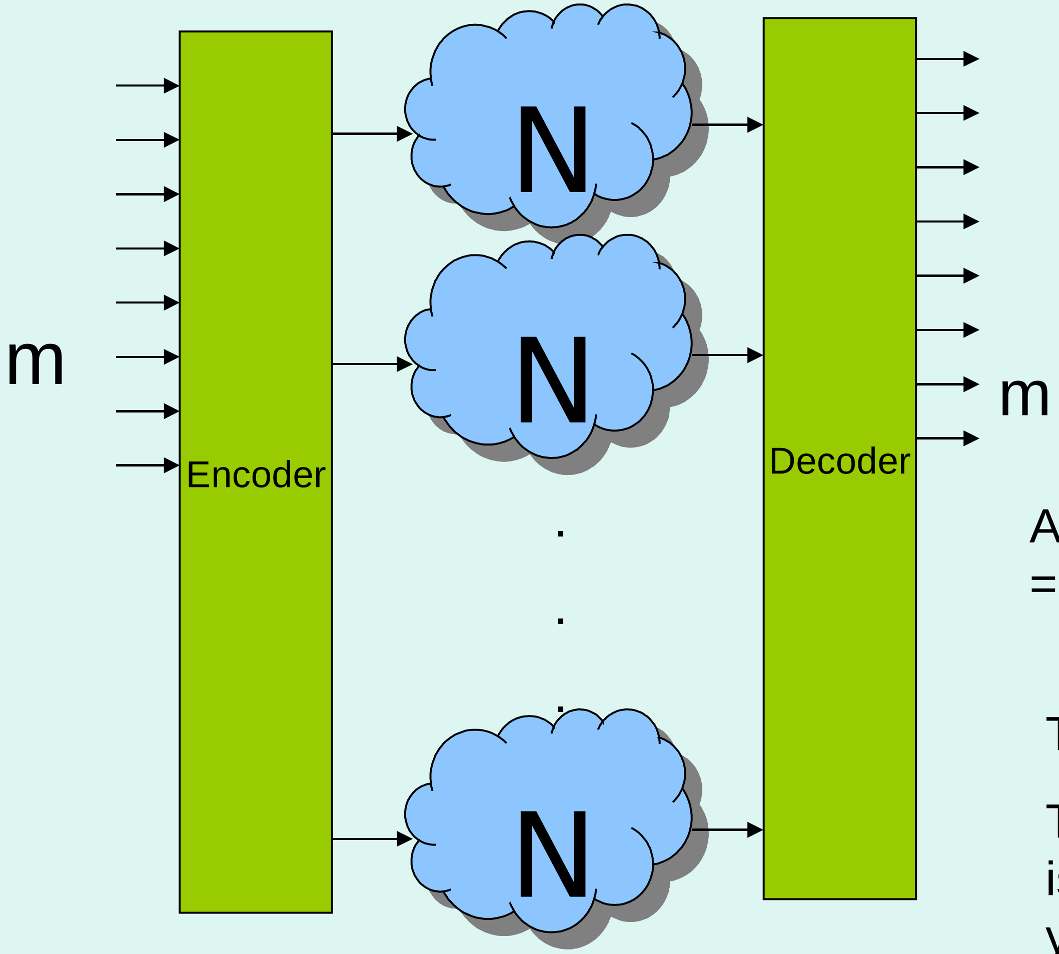
=n H(p)

Where H(p) = -p logp – (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: $H(X) = \sum_x -p(x) \log p(x)$

Channel Capacity



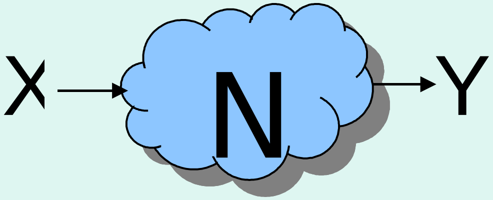
Given n uses of a channel, encode a message m 2 {1,…,M} to a codeword $x^n = (x_1(m),…, x_n(m))$

At the output of the channel, use $y^n = (y_1,…,y_n)$ to make a guess, m^0.

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

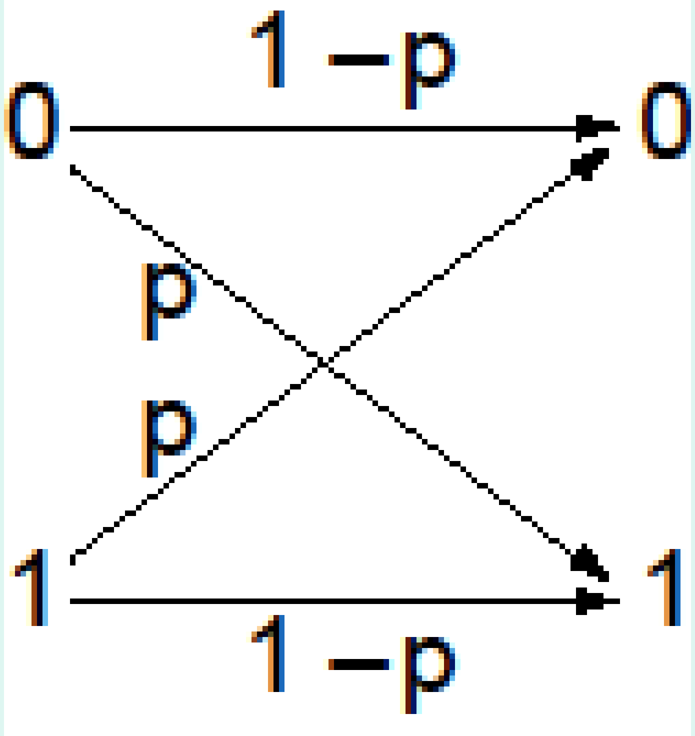
The capacity of the channel, C(N), is defined as the maximum rate you can get with vanishing error probability as n ! 1

Binary Symmetric Channel

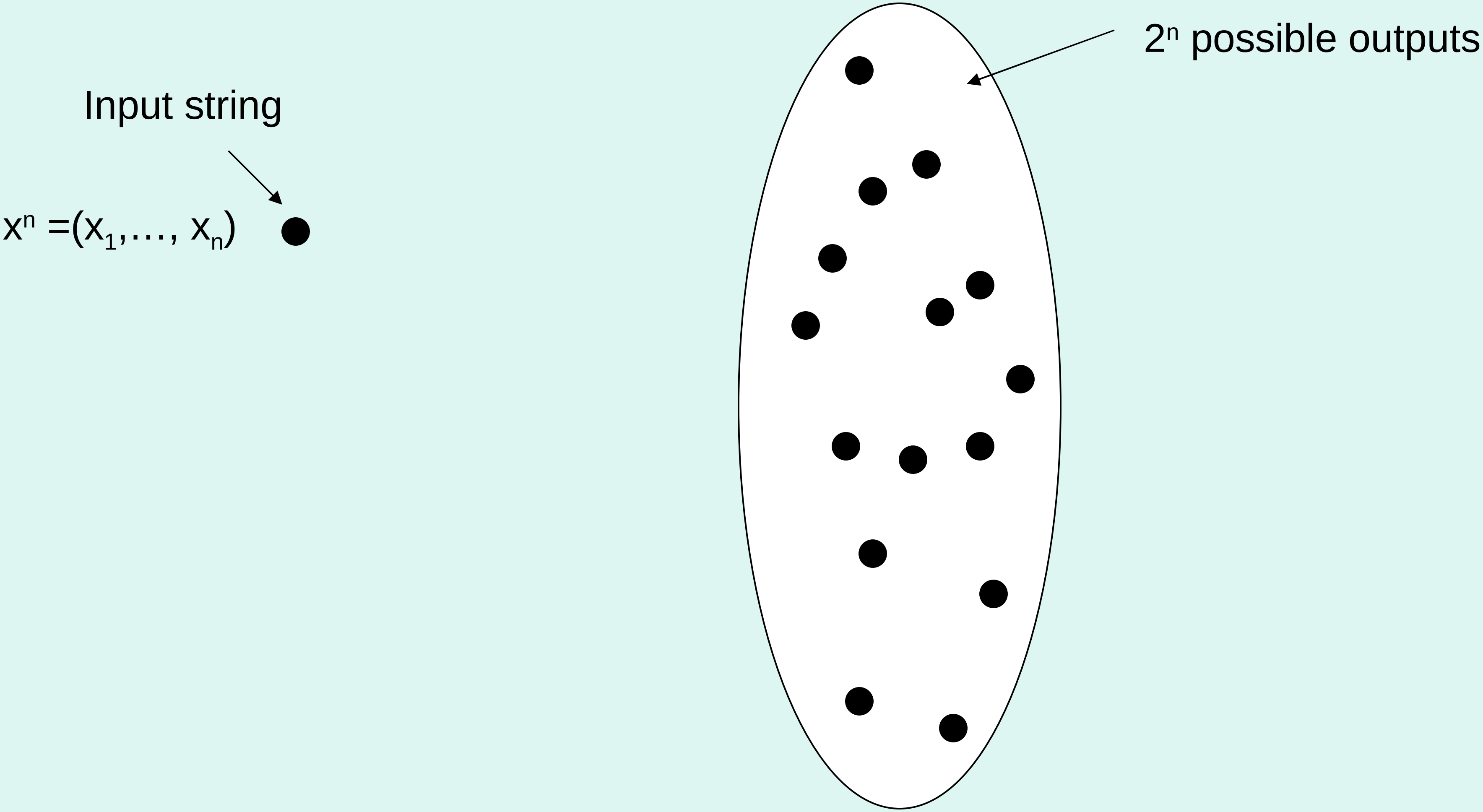


p(0|0) = 1-p p(1|0) = p

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



Capacity of Binary Symmetric Channel

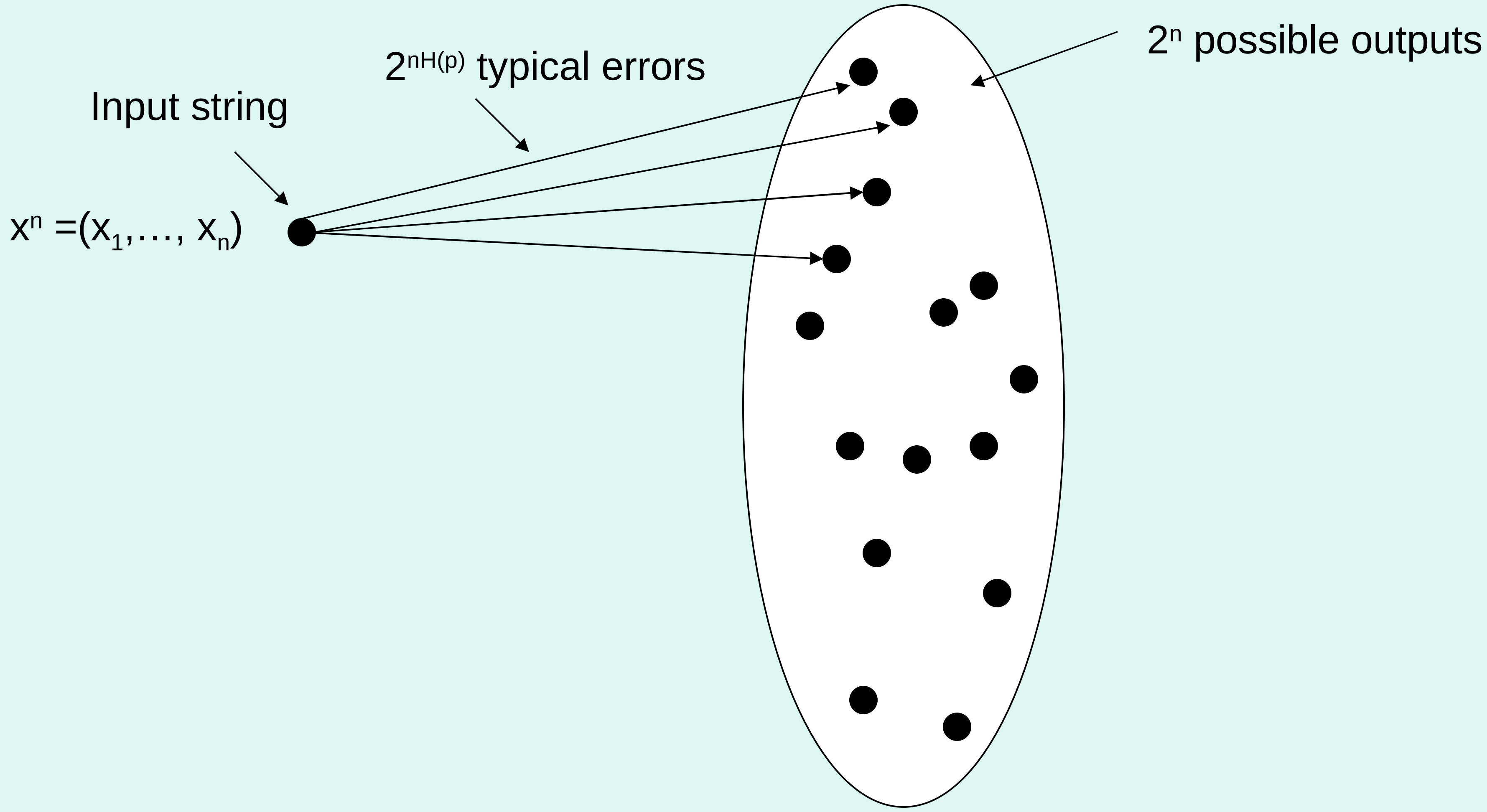


Input string

$x^n =(x_1,\ldots, x_n)$

2^n possible outputs

Capacity of Binary Symmetric Channel



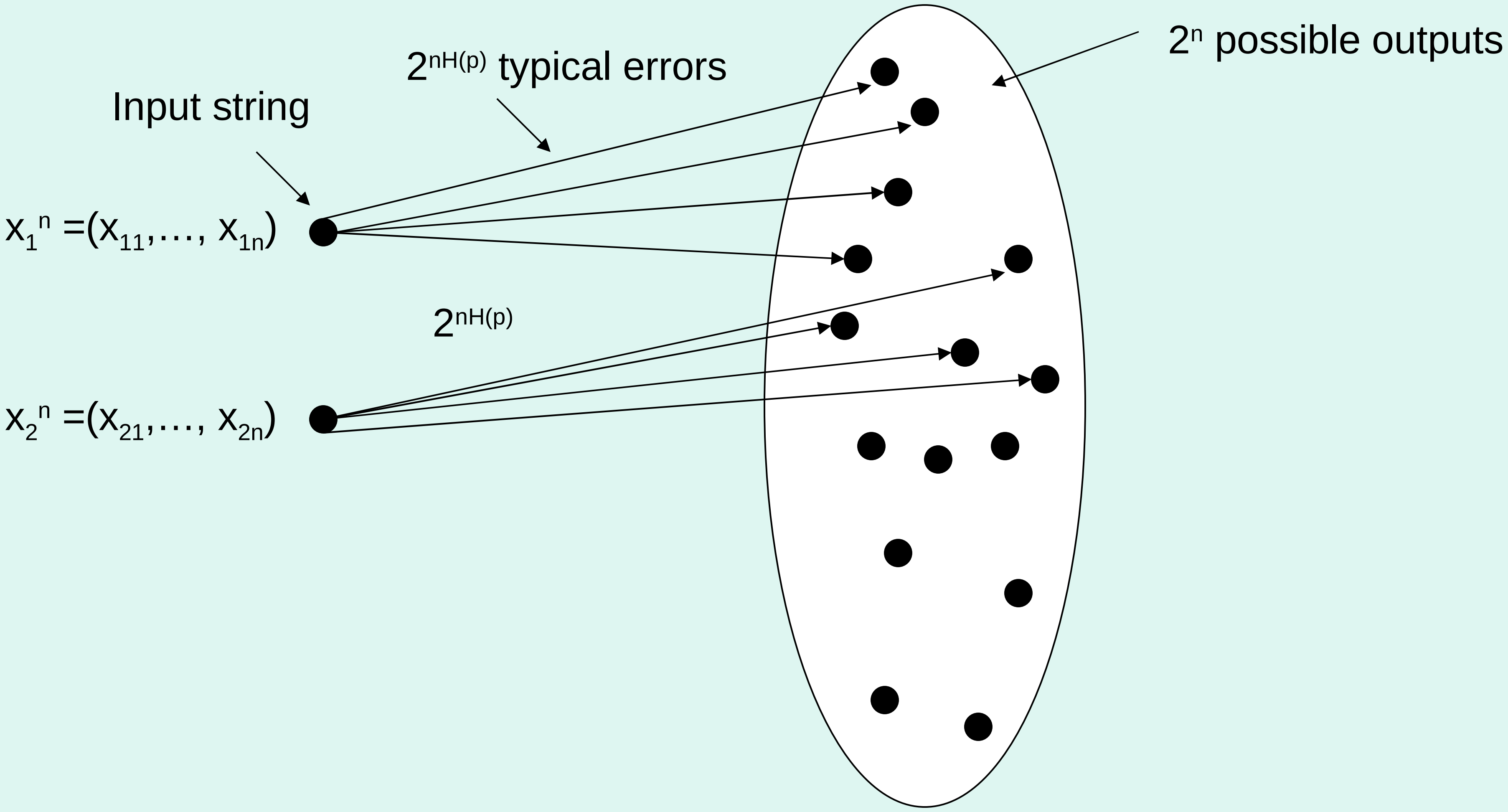
2^n possible outputs

$2^{nH(p)}$ typical errors

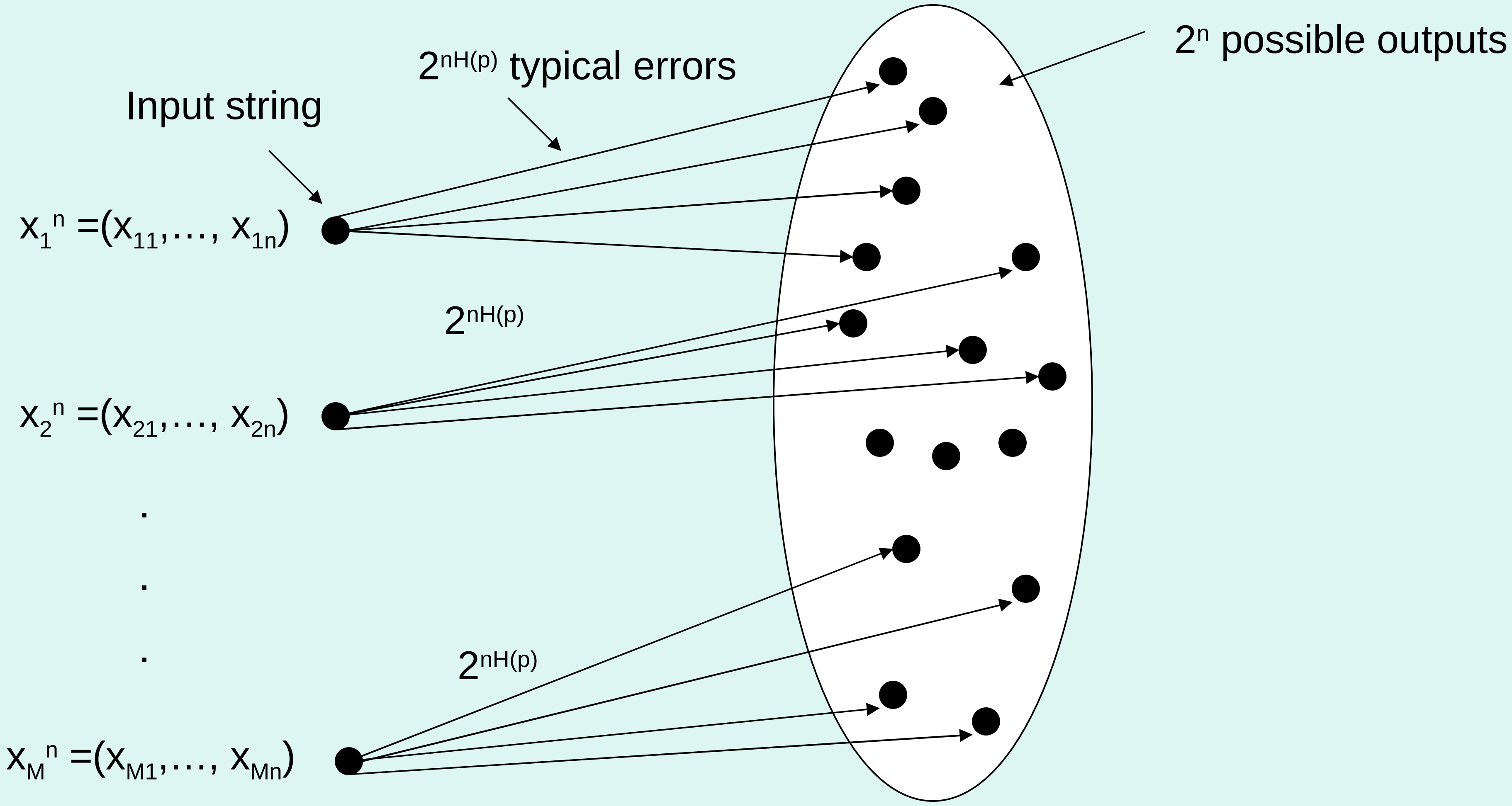
Input string

$x^n = (x_1, \ldots, x_n)$

Capacity of Binary Symmetric Channel


2^n possible outputs
$2^{nH(p)}$ typical errors
Input string
$x_1^n = (x_{11}, \ldots, x_{1n})$
$2^{nH(p)}$
$x_2^n = (x_{21}, \ldots, x_{2n})$

Capacity of Binary Symmetric Channel



2^n possible outputs

$2^{nH(p)}$ typical errors

Input string

$x_1^n = (x_{11}, \ldots, x_{1n})$

$2^{nH(p)}$

$x_2^n = (x_{21}, \ldots, x_{2n})$

.

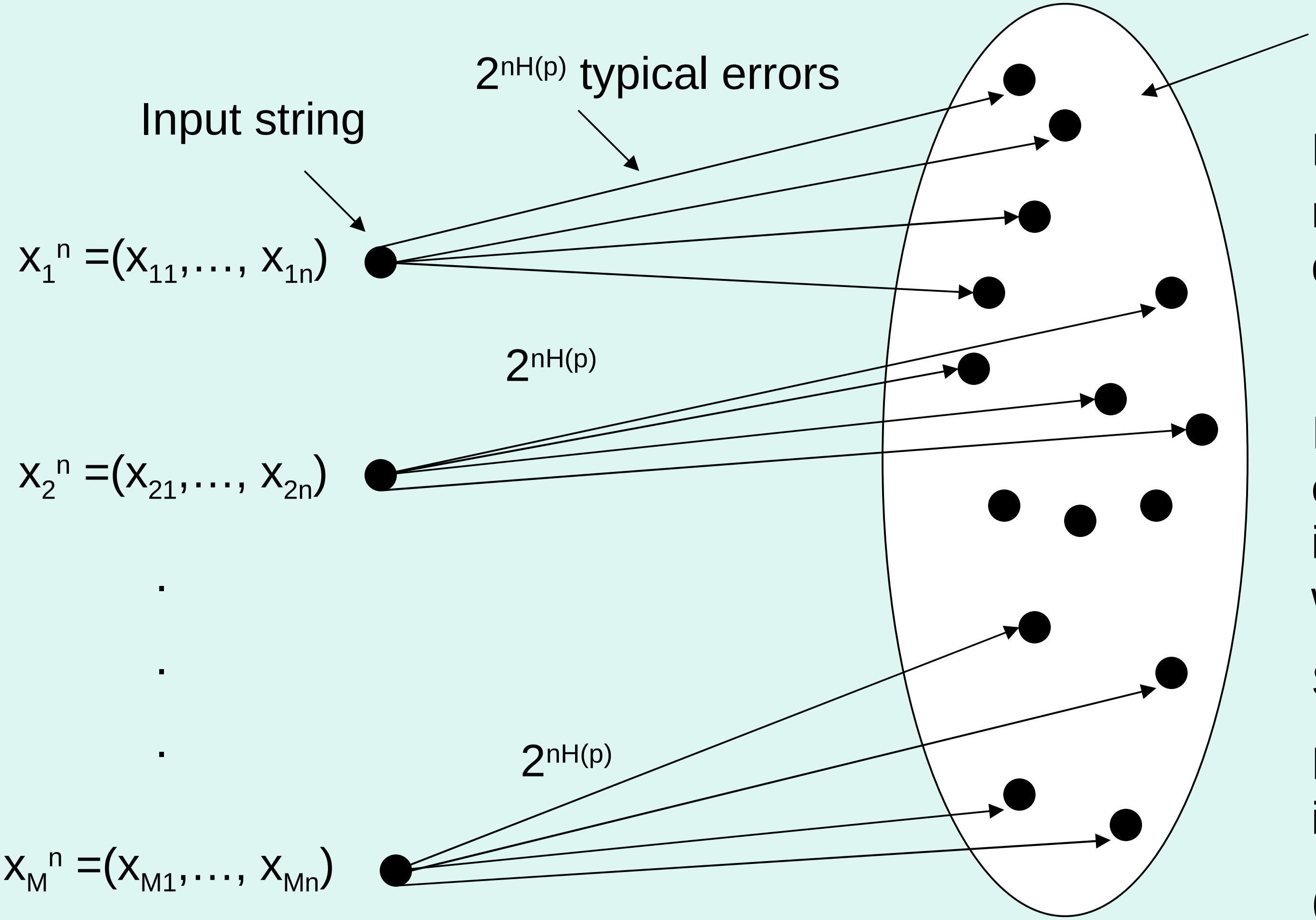
.

.

$2^{nH(p)}$

$x_M^n = (x_{M1}, \ldots, x_{Mn})$

Capacity of Binary Symmetric Channel


$2^{nH(p)}$ typical errors
Input string
$x_1^n = (x_{11}, \ldots, x_{1n})$
$2^{nH(p)}$
$x_2^n = (x_{21}, \ldots, x_{2n})$
.
.
.
$2^{nH(p)}$
$x_M^n = (x_{M1}, \ldots, x_{Mn})$
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

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

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

Upper bound on capacity

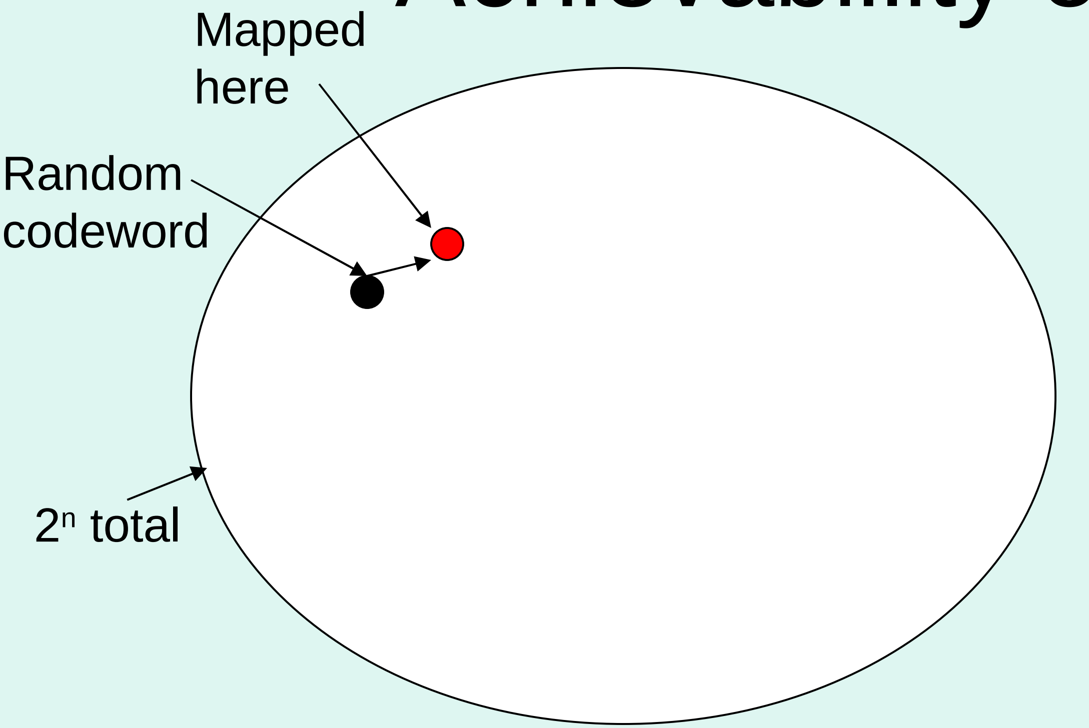
Direct Coding Theorem: Achievability of 1-H(p)

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

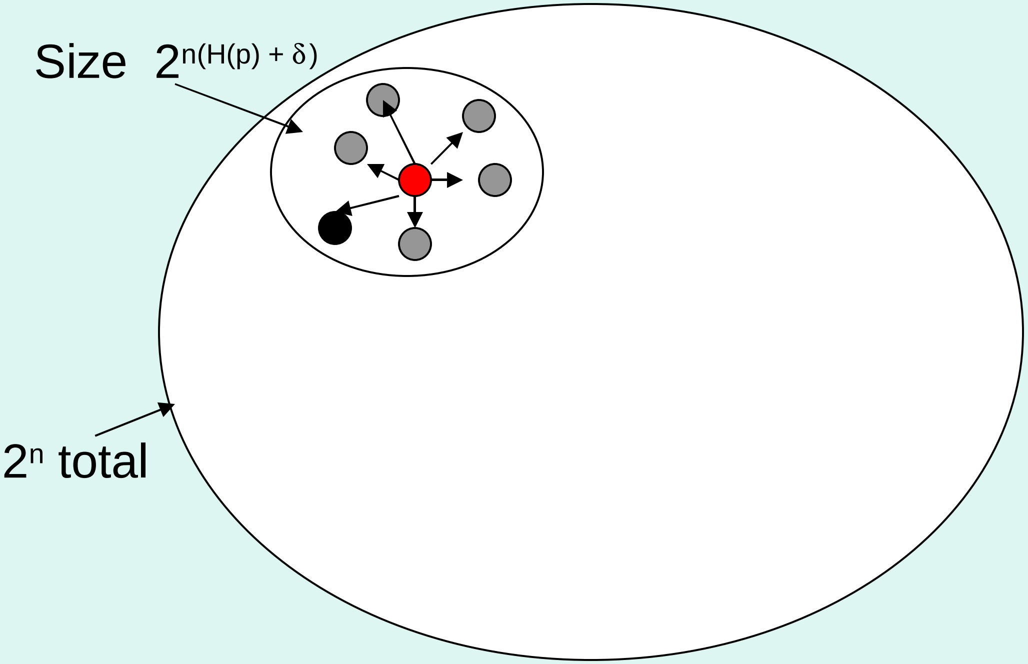


Mapped here

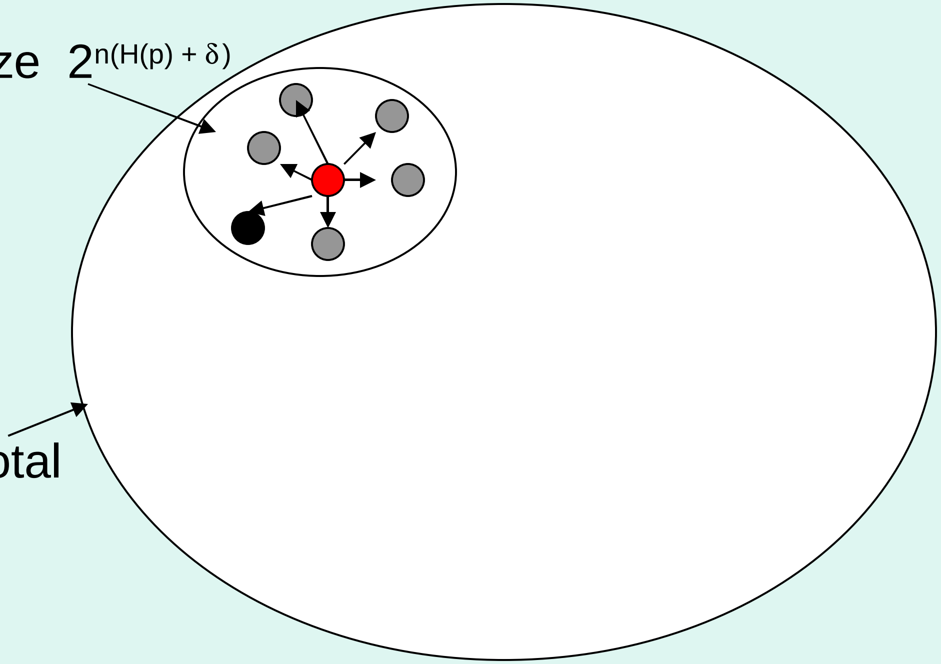
Random codeword

2^n total

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



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

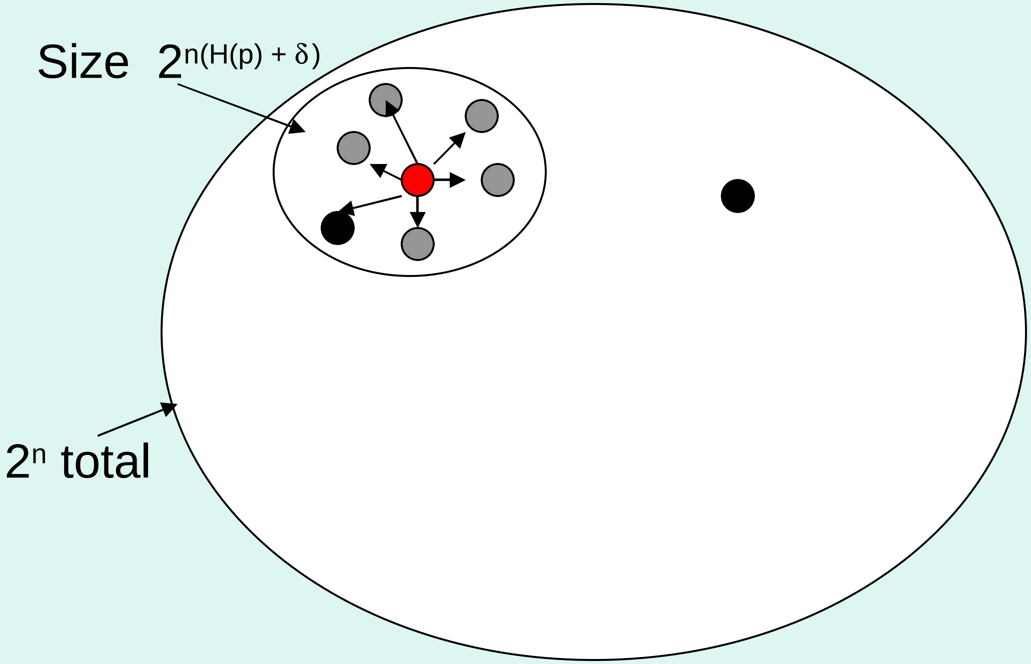


Size $2^{n(H(p) + \delta)}$

2^n total

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

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



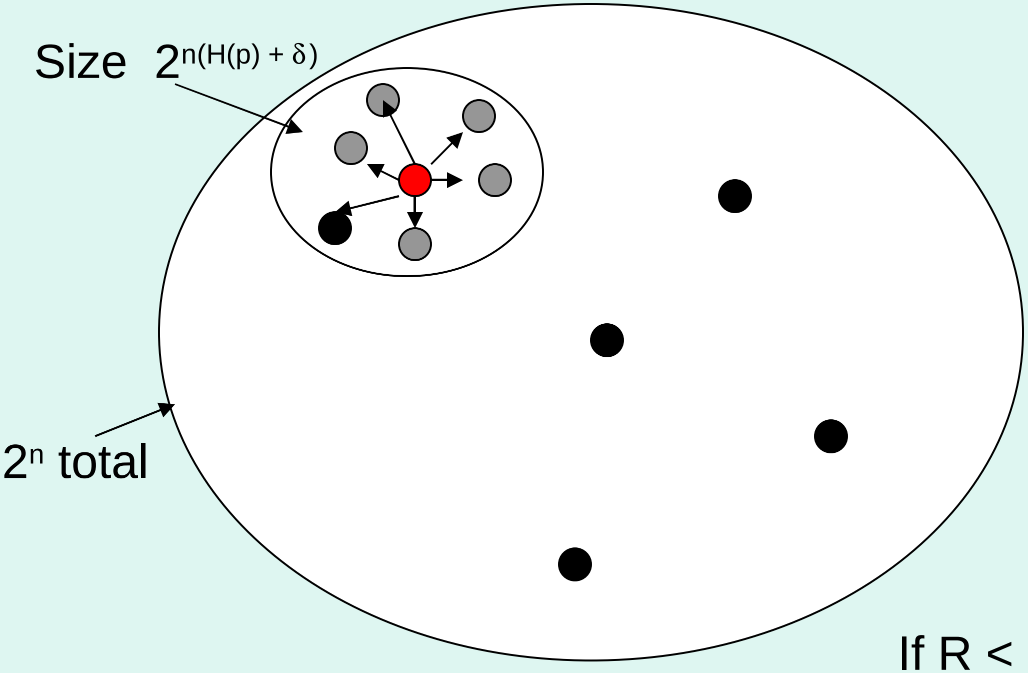
Size $2^{n(H(p) + \delta)}$

2^n total

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 $\dfrac{2^{n(H(p)+\delta)}}{2^n}$

Choose 2^{nR} more, the chance is

$$\frac{2^{n(H(p)+R+\delta)}}{2^n}$$

If R < 1 – H(p) – d, this approaches 0 as n goes to infinity

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).

Shannon's Theorem: Capacity for a general channel

Mutual Information: $I(X;Y)=H(X)+H(Y)-H(X,Y)$

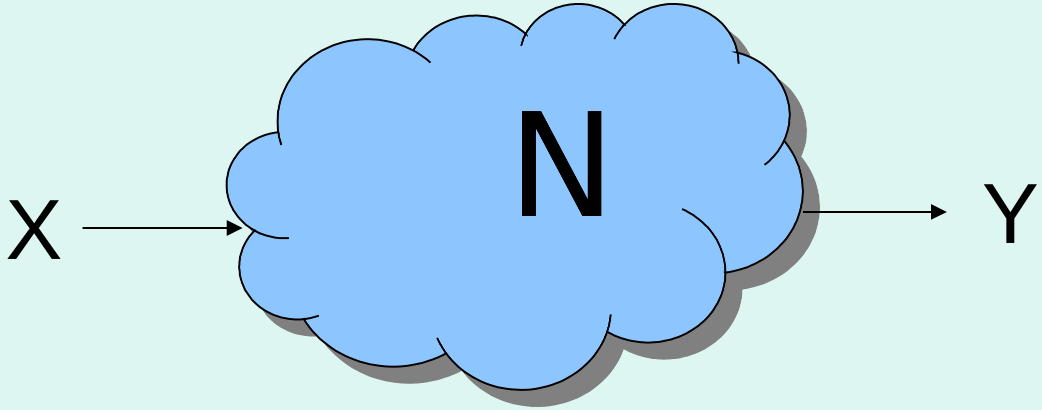
$=H(X)-H(X|Y)$

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

- For any input distribution $p(x)$, given by $p(y|x)$, 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.
- You can’t do any better. (Converse)

The many capacities of a quantum channel

Channel Capacity



p(y|x)

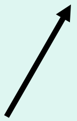
Capacity: bits per channel use in the limit of many channels

$$C = \max_X I(X;Y)$$

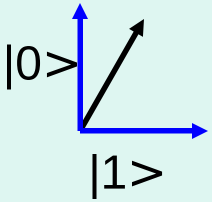
I(X;Y) is the mutual information

Pure Quantum States

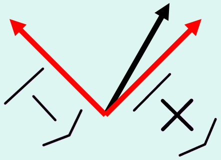
- Qubit: |ψi = α |0> + β|1> , α,β complex and $|\alpha|^2 + |\beta|^2 = 1$.



- If you measure |ψi in the |0>,|1> basis, you get 0 with prob. $|\alpha|^2$ and 1 with prob. $|\beta|^2$



- You could use some other basis, though. Like $|+\rangle, |-\rangle$, with $$|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$$



For a d-level system |ψ> is a unit vector in C^d

Mixed Quantum States

- Pure states are the minimum uncertainty states in quantum mechanics.
- We can also have mixed states:
  $\rho = \sum_i p_i |\psi_i X\psi_i|$ with p_i positive and
  $\sum_i p_i = 1$
- Can think of it as a bipartite pure state with one part traced out: $\rho_B = Tr_A |\psi_{AB} X \psi_{AB}|$
- A pure whole can have mixed parts

Entropy and Typical Spaces

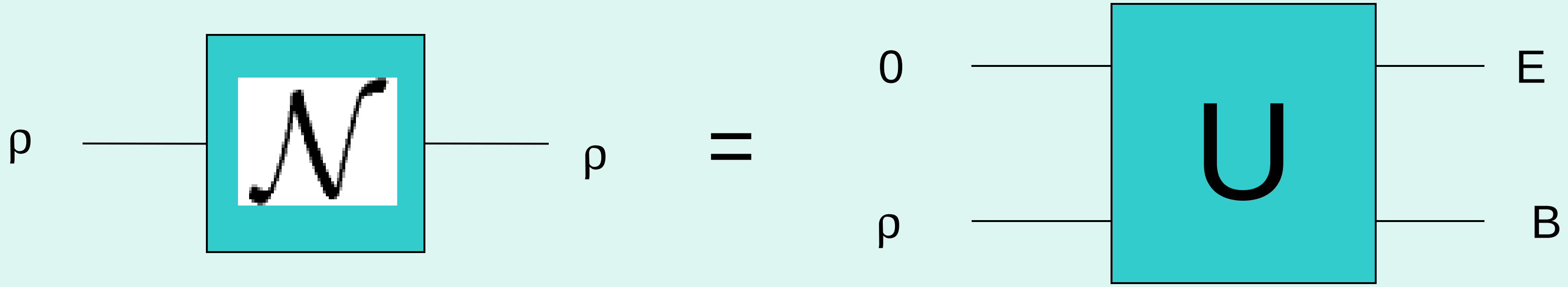
- Any mixed state can be written as
- $\rho_A =_B |\psi\rangle\langle\psi|_{AB}$
- $S(\rho_A) = -\rho_A \log \rho_B$ is the entropy
- It measures the uncertainty in A
- Given n copies of ρ_B we can reversibly map A to a space of dimension $2^{nS(\rho_B)}$ This is the “typical space”
- Analogous to “typical sets” of classical information theory. $2^{nH(p)}$ strings

Noisy Quantum Channels

- Noiseless quantum evolution: $\rho \longrightarrow U\rho U^\dagger$

  Unitary satisfies $U^\dagger U = I$

- Noisy quantum evolution: unitary interaction with inaccessible environment



$$\rho \longrightarrow \mathrm{Tr}_E U(\rho \otimes |0\rangle\langle 0|_E)U^\dagger$$

Classical Capacity of Quantum Channel

We can understand coding schemes for classical information in terms of the Holevo Information:
$\chi(\mathcal{N}) = \max_{\{p_X, \rho_X\}} I(X;B)$
where $I(X;B) = H(X) + H(B) - H(XB)$ uses von Neumann entropy and is evaluated on the state $\sum_x p_x |x\rangle\langle x| \otimes \mathcal{N}(\rho_x)$

Random coding arguments show that $\chi(\mathcal{N})$ is an achievable rate, so C(N) >= $\chi(\mathcal{N})$. Furthermore,

n uses

$$C(\mathcal{N}) = \lim_{n\to\infty} (1/n)\chi(\mathcal{N} \otimes \ldots \otimes \mathcal{N})$$

(see Holevo 73, 79, 98, Schumacher-Westmoreland 97)

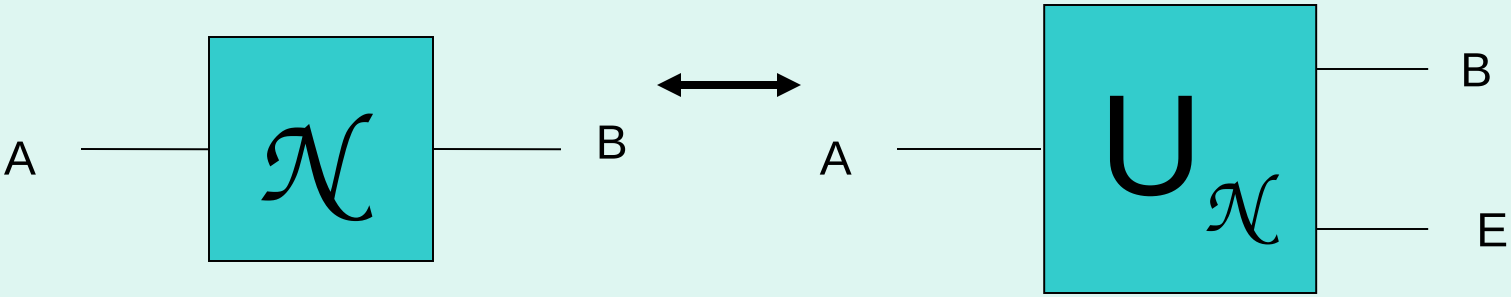
Alternative form of Holevo quantity

$$\chi(\mathcal{N}) = \max_{p_x,\rho_x} \; S(\sum_x p_x \rho'_x) - \sum_x p_x S(\rho'_x)$$

$$\text{where } \rho'_x = \mathcal{N}(\rho_x)$$

(Sometimes people refer to the Holevo quantity without the maximization as well)

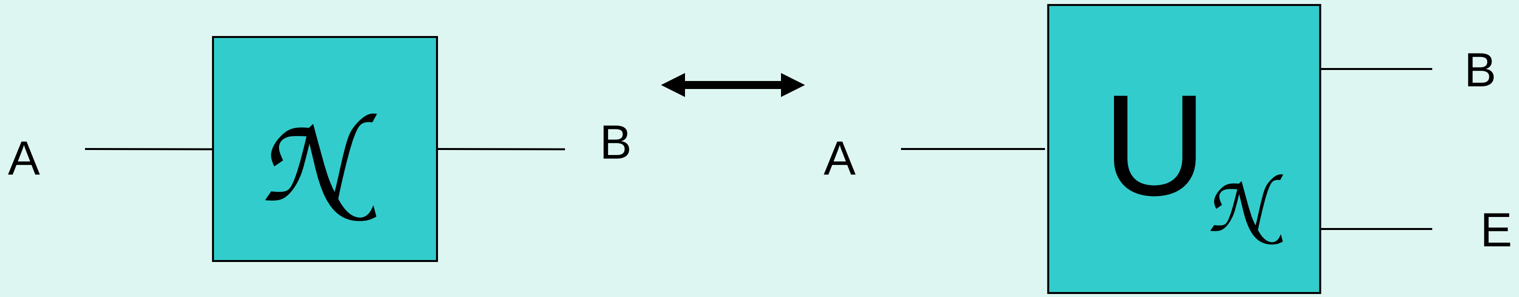
Private Classical Capacity



- Quantum channel has one sender, two receivers.
- Best rate for classical messages from A to B while E learns nothing is the **private capacity**. Call it P($\mathcal{N}$).
- Related to quantum key distribution---the fact than by analyzing the map from A to B we can infer the map from A to E allows unconditional security that is impossible classically.*

* "Stupid" private capacity without back-communcation

Private Classical Capacity



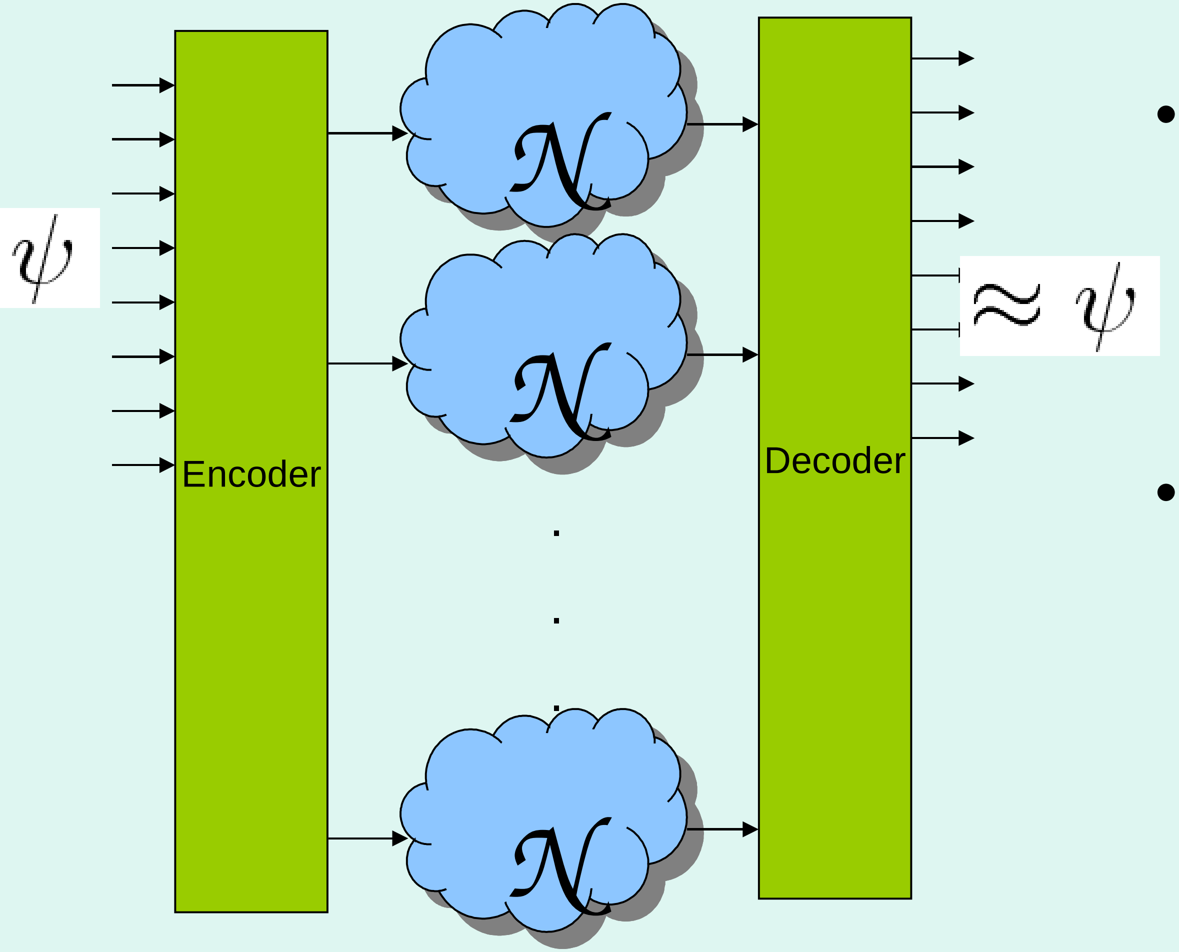
- Let $P^1(\mathcal{N}) = \max_{p_v,\phi_v} I(V;B) - I(V;E)$, with mutual informations evaluated on $\sum_v p_v |v\rangle\langle v| \otimes U\phi_v U$
- Random coding and privacy amplification shows $P(\mathcal{N}) >= P^1(\mathcal{N})$ and, in fact we can get

$$P(\mathcal{N}) = \lim_{n\to\infty} (1/n) P^1(\mathcal{N} \otimes \ldots \otimes \mathcal{N})$$

See Devetak 03

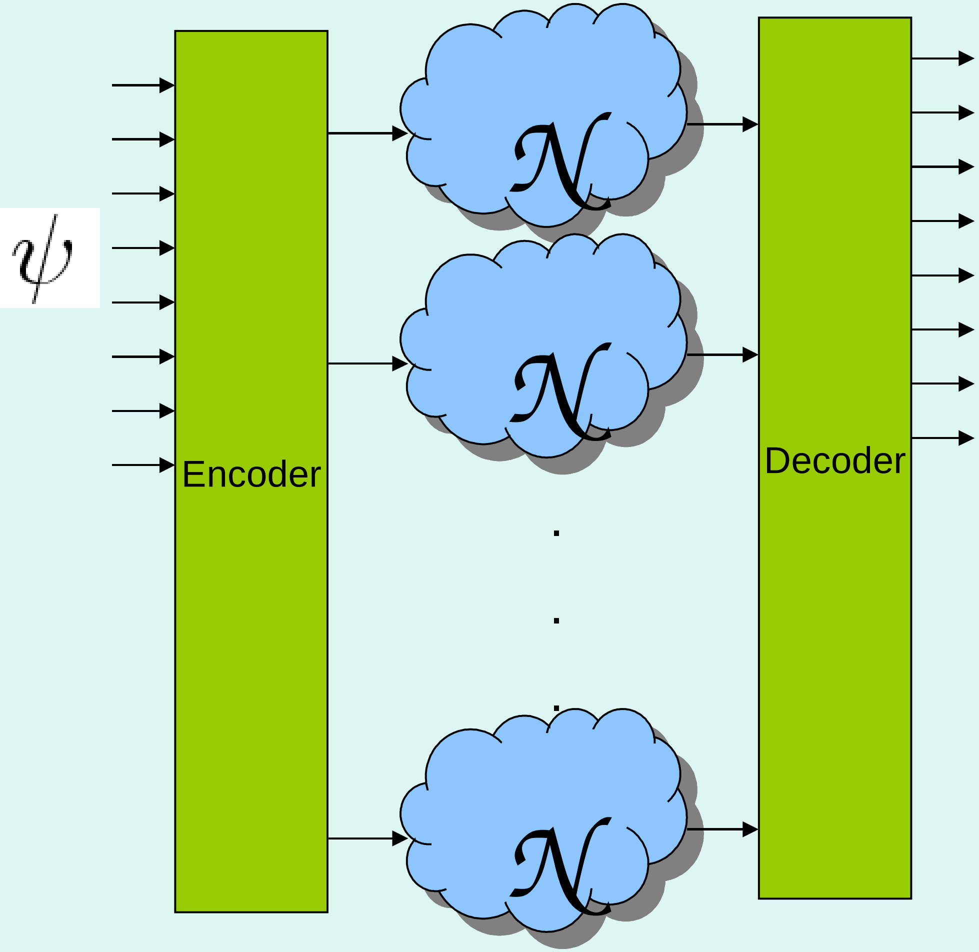
n uses

Quantum Capacity



- If we try to transmit an arbitrary quantum state, we arrive at the quantum capacity, Q($\mathcal{N}$).
- The quantum capacity, measured in qubits per channel use, characterizes the ultimate limit on quantum error correction.

Quantum Capacity



$\approx \psi$

Something like how much more B knows than E

$$Q^1 = \max_{\phi} H(B) - H(E)$$

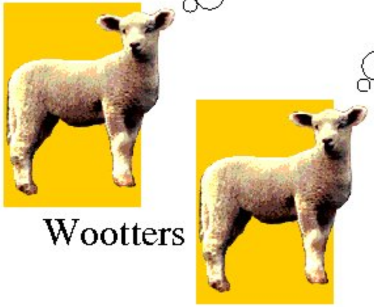
Evaluate entropies on $U\phi U^\dagger$

$$Q(\mathcal{N}) \geq Q^1(\mathcal{N})$$

$$Q(\mathcal{N}) = \lim_{n\to\infty} (1/n) Q^1(\mathcal{N} \otimes \ldots \otimes \mathcal{N})$$

Quantum Information cannot be cloned.

Bah!



Wootters

Zurek

$$|\psi_0\rangle = |\alpha\rangle \otimes |0\rangle$$
$$|\psi_1\rangle = |\beta\rangle \otimes |0\rangle$$

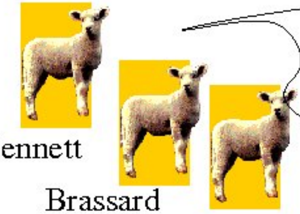
input states "blank paper"

Copied states

$$|\psi_0^c\rangle = |\alpha\rangle \otimes |\alpha\rangle$$
$$|\psi_1^c\rangle = |\beta\rangle \otimes |\beta\rangle$$

$$\langle\psi_1|\psi_0\rangle = \langle\beta|\alpha\rangle$$
$$\langle\psi_1^c|\psi_0^c\rangle = \langle\beta|\alpha\rangle^2$$

Cannot be done unitarily



Bennett

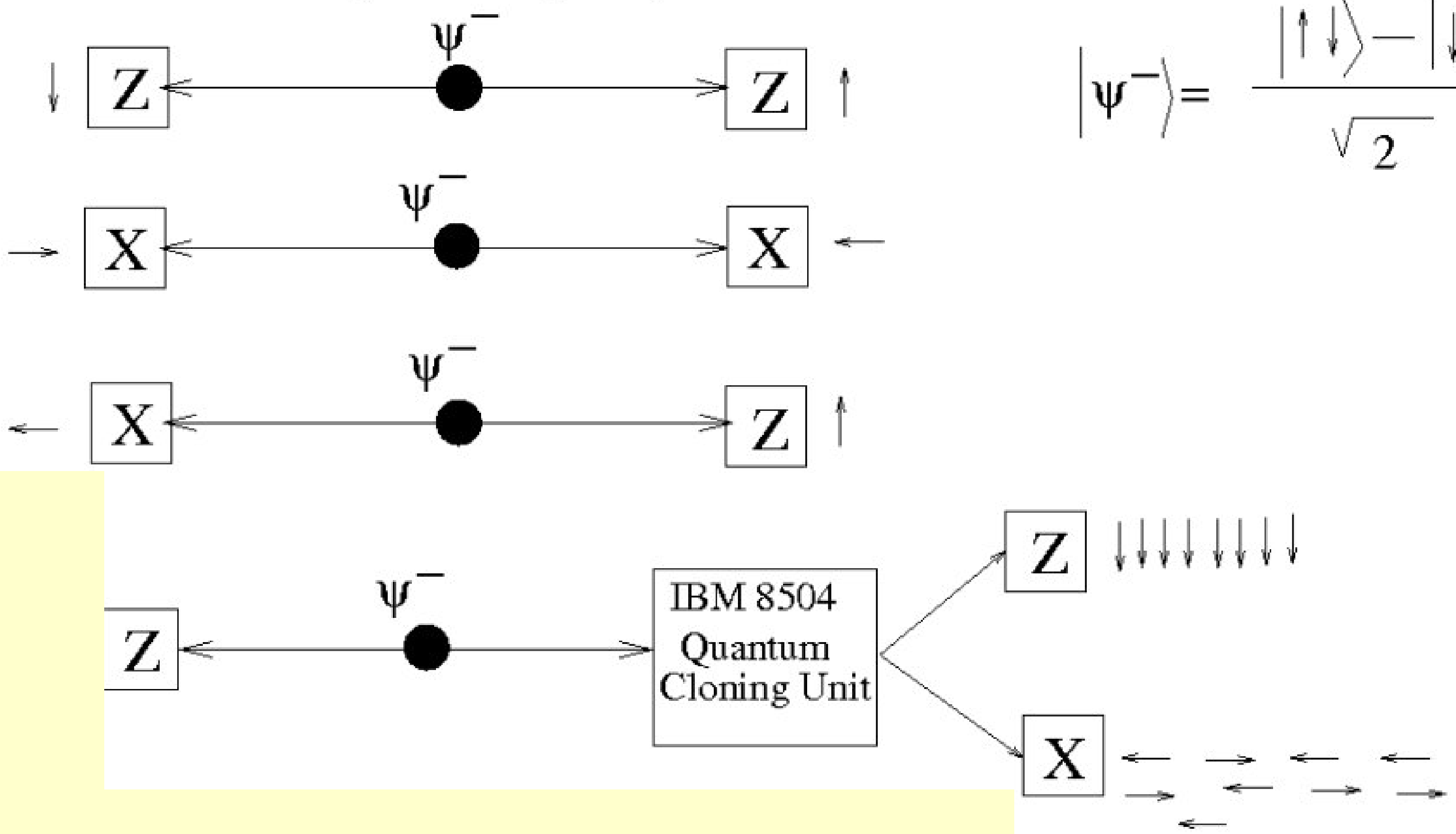
Brassard

Mermin

In fact, if the blank paper ends up with ANY information about the original state at all, then the original state has been disturbed!

Einstein–Podolsky–Rosen (EPR)

$$|\psi^-\rangle = \frac{|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle}{\sqrt{2}}$$



Big Idea

(That sometimes is lost in all the formalism)

If you have shared an EPR pair, then you can send a quantum state by teleportation

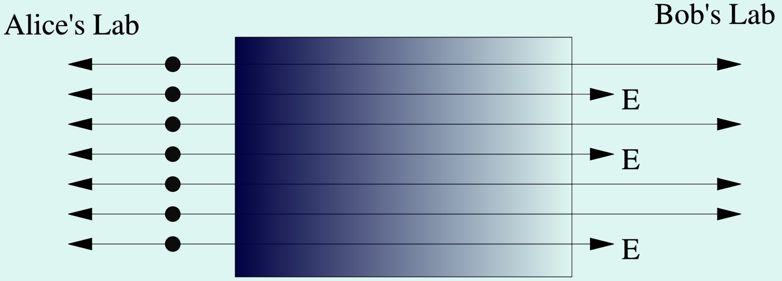
Furthermore, even though to actually teleport would require classical communication, if you can share an EPR pair through a channel then you can also send a state without the classical communication

Less Big Idea

If you can share an arbitrary quantum state through a channel, the you can share an EPR pair

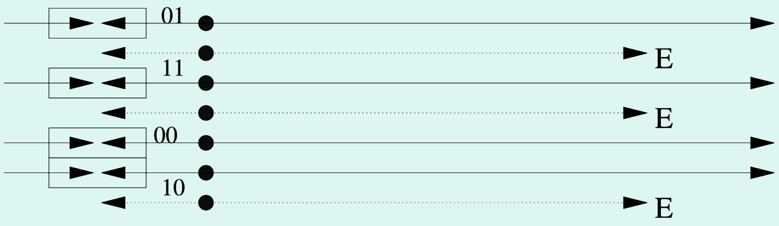
Obviously

Big idea plus less big idea gets capacity converse for free *qteleport.fig*



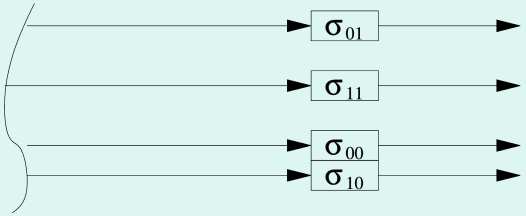
Bob Sends classical message to Alice

Hey Alice!
Forget about EPR
pairs 2,4 and 7.



Alice sends classical message to Bob

Hey Bob! Rotate your qubits
as follows: 01,11,00,10.



We have used redundancy without copying unknown qubits.

With Identity Independent Distributions (IID), everything works far better than one could even hope:

Shannon coding manages to get to a high probability that **every bit is correct**, when we might have been pretty happy with each bit being correct with high probability

One is exponentially better than the other:
If the probability of each bit being correct is p,
then the probability of all bits being correct is only p^n,
and we expect np of them to be wrong.

The cryptographer wants all the bits to be secure.
If np bits leak, what if they're the most important ones?

Cryptographers are just paranoid information theorists.

EMBRATUR
RIO
RIO

Unfortunately, in the cryptographic setting channels are not IID.

They're **adversarial**, which is the worst possible thing. We don't even get to know what the channel is!

Aside: There are lots of beautiful results about privacy in the IID case. Devetak.

Remarkably, QKD still achieves the strong form of security where every bit is safe!

How is strong security obtained in the adversarial case?

That was the hard part to prove.....

Mayers, Preskill-Shor.

Important tool: Back communication

Not needed in for IID classical capacity

Channel Tomography: Figure out what the channel *is*, on the average

Randomization: Change the order around so the adversary loses some power

Privacy amplification

Reasons your QKD might fail

Quantum Mechanics is wrong

Your random numbers are bad

Proofs of security are wrong

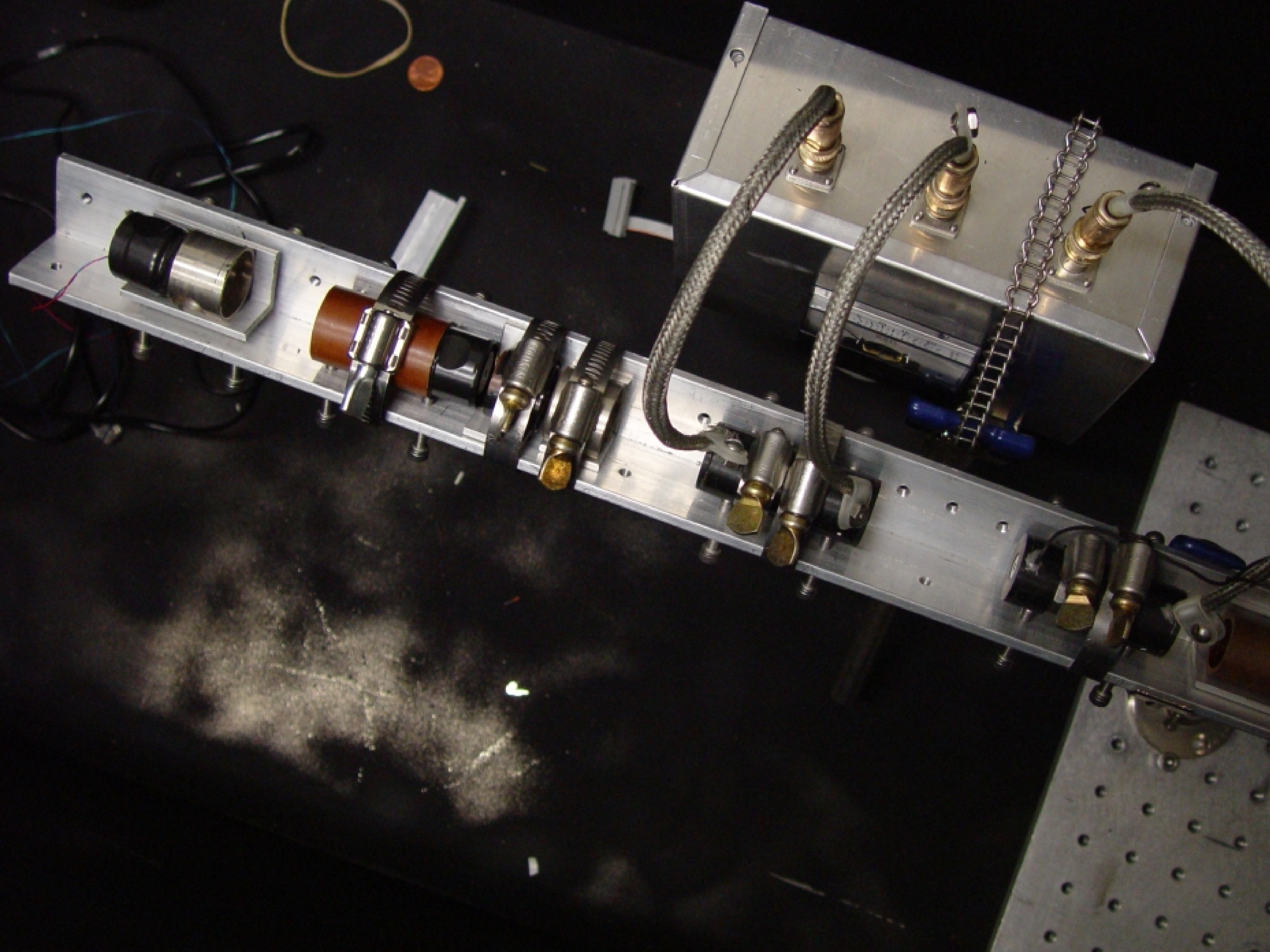
You're using the proofs wrong

Your system isn't described by the physics you think it is

Your lab is insecure

Your people are insecure

It got jammed



Authentication is Key

QKD requires the quantum channel and a public,
but authenticated, classical channel

QKD really should be thought of as “key expansion”
because you have to start with some key for authentication

Perhaps this is less of a problem point to point between
ships, but then why do we need QKD at all?

Most things aren’t additive

- Q^1 is not additive for the very noisy depolarizing channel (Shor-Smolin ‘96)
- P^1 isn’t additive for BB84 channel (Smith-Renes-Smolin, ‘08)
- χ is nonadditive for high-dimensional random channel (Hastings ‘09)
- Q^1 and P^1 can both be extremely nonadditive (Smith-Smolin 08, 09)

But sometimes they are

- χ is additive for depolarizing, erasure, and entanglement breaking channels.
- Q^1 and P^1 are additive for degradable channels*, Q^1 is for PPT channels.

* Just like a degraded broadcast channel when you take the less noisy user to be the channel output and the more noisy user to be the environment

See King, Shor, Devetak-Shor, Horodecki, …

A different kind of (non)additivity

Already saw that Q^1 wasn't additive, but what about $Q(\mathcal{N}) = \lim_{n\to\infty}(1/n)Q^1(\mathcal{N}^{\otimes n})$?

Since $Q(\mathcal{N}\otimes\mathcal{N}) = 2\,Q(\mathcal{N})$, this is actually a question about how different channels interact: Can $Q(\mathcal{N}\otimes\mathcal{M}) > Q(\mathcal{N}) + Q(\mathcal{M})$?

Yes

A different kind of (non)additivity

The only channels with zero classical capacity have no correlation between input and output. However, quantum information is more delicate, so there are nontrivial quantum channels with $Q(\mathcal{N}) = 0$. A good example is the 50% quantum erasure channel ($\rho \rightarrow \frac{1}{2}\rho + \frac{1}{2} |e\rangle\langle e|$).

There's a more complicated kind of channel with $Q(\mathcal{M}) = 0$, called a private PPT channel. These have $P(\mathcal{M}) > 0$.

You can show that for any such PPT channel, $Q(\mathcal{N} \otimes \mathcal{M}) \geq \frac{1}{2} P(\mathcal{M}) > 0$, so in the end, we get

$Q(\mathcal{N}) = 0$ and $Q(\mathcal{M}) = 0$, but $Q(\mathcal{N} \otimes \mathcal{M}) > 0$.

This is for two qubit channels, but with larger channels you can make the additivity violation very large (1/8 log d). Get similar nonadditivity for the private classical capacity.

Additivity Questions

Information \ Quantity	Capacity	Correlation Measure
Classical	Classical Capacity **?**	Holevo Information max I(X;B) **No** (Hastings ‘09)
Private	Private Capacity **No** (Li-Winter-Zou-Guo ’09 Smith-Smolin-08/09)	Private Information max I(X;B)-I(X;E) **No** (Smith-Renes-Smolin ‘08)
Quantum	Quantum Capacity **No** (Smith-Yard ‘08)	Coherent Information max S(B)-S(E) **No** (Div-Shor-Smolin ‘98)
Entanglement assisted	Entanglement assisted classical capacity **Yes** (Bennett-Shor-Smolin-Thapliyal ‘99)	Quantum Mutual Information **Yes** (Bennett-Shor-Smolin-Thapliyal ‘99)

Additivity: definition and motivation

- A function on channels is called additive if $f(\mathcal{N} \otimes \mathcal{M}) = f(\mathcal{N}) + f(\mathcal{M})$
- Recall that $Q(\mathcal{N}) = \lim_{n \to \infty} (1/n) Q^1(\mathcal{N}^{\otimes n})$. If we could show that Q^1 was additive, we'd have $Q(\mathcal{N}) = Q^1(\mathcal{N})$.
- Similarly, $C(N) = \lim_{n \to \infty} (1/n)\, \chi(\mathcal{N}^{\otimes n})$ and $P(\mathcal{N}) = \lim_{n \to \infty} (1/n) P^1(\mathcal{N}^{\otimes n})$, so if χ and P^1 were additive, we'd have single-letter capacities for classical and private communication.

Lots of great stuff from quantum information:

Superactivation

Now with gaussian channels

Reverse Shannon theorem

Entropy-Power inequality

Quantum computation (the flip side of QKD)

Entanglement assisted communication