

# Quantum Information Theory (Preskill ch. 5)

## Main Topics of QIT:

- (1) Transmission of classical info over quantum channels
- (2) Information/disturbance tradeoff in QM
- (3) Quantifying entanglement
- (3) Transmission of quantum info over quantum channels

Our Program: (1) & (4) ~ 3 Lectures

Key Concept – Incompressible information content

Classical Measure: Shannon Entropy

Quantum Measure: von Neumann Entropy

## Review of Classical Information Theory

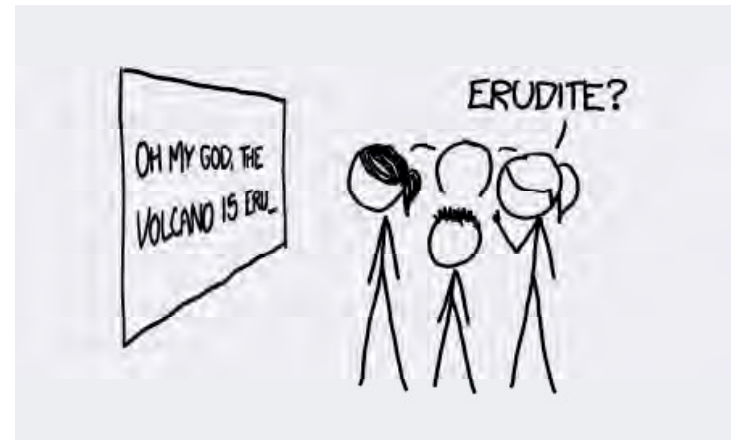
( Shannon for Dummies, Preskill 5.1 )

Shannon, 1948: Core findings of classical info theory

- (1) How much data can be compressed (redundancy)
- (2) Reliable communication rate over noisy channel (Redundancy needed to protect against errors)

## Shannon Entropy and Data Compression

(Shannon's noiseless coding theorem)



# Quantum Information Theory (Preskill ch. 5)

## Main Topics of QIT:

- (1) Transmission of classical info over quantum channels
- (2) Information/disturbance tradeoff in QM
- (3) Quantifying entanglement
- (3) Transmission of quantum info over quantum channels

Our Program: (1) & (4) ~ 3 Lectures

Key Concept – Incompressible information content

Classical Measure: Shannon Entropy

Quantum Measure: von Neumann Entropy

## Review of Classical Information Theory

(Shannon for Dummies, Preskill 5.1)

Shannon, 1948: Core findings of classical info theory

- (1) How much data can be compressed (redundancy)
- (2) Reliable communication rate over noisy channel (Redundancy needed to protect against errors)

## Shannon Entropy and Data Compression

(Shannon's noiseless coding theorem)

Message = String of letters chosen from  $\{a_1, a_2, \dots, a_k\}$

A priori probability of occurrence:  $p(a_x), \sum p(a_x) = 1$

Basic Question: given message w/  $n \gg 1$  letters

Can we compress to length  $< n$  ?

# Quantum Information Theory (Preskill ch. 5)

## Review of Classical Information Theory

(Shannon for Dummies, Preskill 5.1)

Shannon, 1948: Core findings of classical info theory

- (1) How much data can be compressed (**redundancy**)
- (2) Reliable communication rate over noisy channel (**Redundancy** needed to protect against errors)

## Shannon Entropy and Data Compression

(Shannon's noiseless coding theorem)

Message = String of letters chosen from  $\{a_1, a_2, \dots, a_d\}$

A priori probability of occurrence:  $p(a_x), \sum p(a_x) = 1$

Basic Question: given message w/  $n \gg 1$  letters

Can we compress to length  $< n$  ?

Look at Binary case:

$$0 \quad p(0) = 1-p$$

$$1 \quad p(1) = p$$

Typical Occurrence

$$n(1-p)$$

$$np$$

Binomial coefficient

Number of distinct typical strings  $\sim \binom{n}{np}$

$$\text{Log}_{\text{base 2}} \binom{n}{np} = \text{Log} \frac{n!}{(np)![n(1-p)]!} \quad \left\{ \begin{array}{l} \text{use Stirlings formula} \\ \text{Log } n! = n \text{Log } n - n + \mathcal{O}(\text{Log } n) \end{array} \right.$$

$$\approx n \text{Log } n - n - [np \text{Log}(np) - (np) + n(1-p) \text{Log}[n(1-p)] - n(1-p)]$$

$$\equiv n H(p),$$

$$H(p) = -p \text{Log } p - (1-p) \text{Log}(1-p) = \sum_{x=0,1} p(x) \text{Log } p(x)$$

entropy function

# of bits needed to specify all typical strings, for a given  $n$

# Quantum Information Theory (Preskill ch. 5)

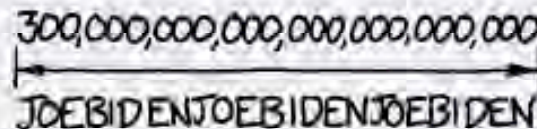
## Why the Log?

**Using only pronounceable letter combinations, how long would names have to be to give each star in the universe a unique one word name?**

–SEAMUS JOHNSON

---

There are about 300,000,000,000,000,000,000 stars in the universe. If you make a word pronounceable by alternating vowels and consonants (there are better ways to make pronounceable words, but this will do for an approximation). then every pair of letters you add lets you name 105 times as many stars (21 consonants times 5 vowels). 105 possibilities per two characters is about the same information density that numbers have, which suggests the name will end up being about as long as the total number of stars written out:



300,000,000,000,000,000,000  
JOEBIDENJOEBIDENJOEBIDEN

I like doing math that involves measuring the lengths of numbers written out on the page—which is really just a way of loosely estimating  $\log_{10} x$ . It works, but it feels so *wrong*.

# Quantum Information Theory (Preskill ch. 5)

Look at Binary case:

$$0 \quad p(0) = 1-p$$

$$1 \quad p(1) = p$$

Typical Occurrence

$$n(1-p)$$

$$np$$

Binomial coefficient

Number of distinct typical strings  $\sim \binom{n}{np}$

$$\text{Log}_{\text{base 2}} \binom{n}{np} = \text{Log} \frac{n!}{(np)![n(1-p)]!}$$

use Stirlings formula  $\left\{ \begin{array}{l} \text{use Stirlings formula} \\ \text{Log } n! = n \text{Log } n - n + \mathcal{O}(\text{Log } n) \end{array} \right.$

$$\approx n \text{Log } n - n - [np \text{Log}(np) - (np) + n(1-p) \text{Log}[n(1-p)] - n(1-p)]$$

$$\approx n H(p),$$

$$H(p) = -p \text{Log } p - (1-p) \text{Log}(1-p) = \sum_{x=0,1} p(x) \text{Log } p(x)$$

entropy function

# of bits needed to specify all typical strings, for a given  $n$

Basic idea of Data Compression:

- \* Assign integer code letter to each typical string
- \* This block code has  $2^{nH(p)}$  letters
- \* Each code letter specified by  $nH(p)$  bits

$$0 \leq p \leq 1 \rightarrow 0 \leq H(p) \leq 1$$

$$H(p) = 1 \text{ only for } p = 1/2$$

Block code compresses message for  $p \neq 1/2$

Generalization:

$k$  letters, prob.  $p(x)$

Ensemble  $\mathcal{X} = \{x, p(x)\}$  of letters

$n$  - letter string  $\rightarrow x$  occurs  $\sim np(x)$  times

$$\# \text{ of typical strings} \sim \frac{n!}{\prod_x [np(x)]!} \sim 2^{-nH(\mathcal{X})}$$

$$H(\mathcal{X}) = -\sum_x p(x) \text{Log } p(x)$$

Shannon entropy !

# Quantum Information Theory (Preskill ch. 5)

## Basic idea of Data Compression:

- \* Assign integer code letter to each typical string
- \* This block code has  $2^{nH(p)}$  letters
- \* Each code letter specified by  $nH(p)$  bits

$$\left. \begin{array}{l} 0 \leq p \leq 1 \rightarrow 0 \leq H(p) \leq 1 \\ H(p) = 1 \text{ only for } p = 1/2 \end{array} \right\} \text{Block code compresses} \\ \text{message for } p \neq 1/2$$

## Generalization:

$k$  letters, prob.  $p(x)$   
 Ensemble  $\mathcal{X} = \{x, p(x)\}$  of letters

$n$  - letter string  $\rightarrow x$  occurs  $\sim np(x)$  times

$$\# \text{ of typical strings} \sim \frac{n!}{\prod_x [np(x)]!} \sim 2^{-nH(\mathcal{X})}$$

$$H(\mathcal{X}) = -\sum_x p(x) \log p(x)$$

Shannon  
 entropy !

We will see that  $H(\mathcal{X})$  quantifies how much info is conveyed, on average, by a letter drawn from the ensemble  $\mathcal{X}$  (alphabet)

Note: Boltzman Entropy

$$S = -k \sum_i p_i \log p_i$$

Here the sum is over the microstates consistent with the given microstate. Assuming all microstates are equally likely, the System will be in the macrostate with the largest  $S$ .

# Quantum Information Theory (Preskill ch. 5)

We will see that  $H(\mathcal{X})$  quantifies how much info is conveyed, on average, by a letter drawn from the ensemble  $\mathcal{X}$  (alphabet)

Note: Boltzman Entropy

$$S = -k \sum_i p_i \log p_i$$

Here the sum is over the microstates consistent with the given microstate. Assuming all microstates are equally likely, the System will be in the macrostate with the largest  $S$ .

## Shannons Noiseless Coding Theorem

Consider a random, typical message  $x_1 x_2 \dots x_n$   
with  $x_j \in \mathcal{X}$  in  $j$ 'th place

$$\begin{cases} p(x_1 \dots x_n) = p(x_1) p(x_2) \dots p(x_n) \\ p(x) = \text{a priori prob. of } x \text{ in place } j \end{cases}$$

Then  $-\frac{1}{n} \log p(x_1 \dots x_n) = -\frac{1}{n} \sum_{j=1}^n \log p(x_j)$  occurs  $n_x$  times on average

$$= -\frac{1}{n} \sum_x n_x \log p(x) \sim H(\mathcal{X})$$

Here, in the 2<sup>nd</sup> to last step we grouped together the  $n_x$  occurrences of  $x$ , used  $p(x_j) \sim p(x)$ , and converted the sum over positions in the string into a sum over letters in the alphabet. As a last step,  $\frac{n_x}{n} \sim p(x)$

**Now:** For any  $\epsilon, \delta > 0$  there exist an  $n$  large enough s. t.

$$H(\mathcal{X}) - \delta \leq -\frac{1}{n} \log p(x_1 \dots x_n) \leq H(\mathcal{X}) + \delta$$

$$\Rightarrow 2^{-n(H+\delta)} \geq p(x_1 \dots x_n) \geq 2^{-n(H-\delta)} \quad (1)$$



# Quantum Information Theory (Preskill ch. 5)

## Shannons Noiseless Coding Theorem

Consider a random, typical message  $x_1 x_2 \dots x_n$   
with  $x_j \in \mathcal{X}$  in  $j$ 'th place

→ 
$$\begin{cases} p(x_1 \dots x_n) = p(x_1) p(x_2) \dots p(x_n) \\ p(x) = \text{a priori prob. of } x \text{ in place } j \end{cases}$$

Then 
$$-\frac{1}{n} \log p(x_1 \dots x_n) = -\frac{1}{n} \sum_{j=1}^n \log p(x_j)$$

$$= -\frac{1}{n} \sum_x n_x \log p(x) \sim H(\mathcal{X})$$

occurs  $n_x$  times  
on average

Here, in the 2<sup>nd</sup> to last step we grouped together the  $n_x$  occurrences of  $x$ , used  $p(x_j) \sim p(x)$ , and converted the sum over positions in the string into a sum over letters in the alphabet. As a last step,  $\frac{n_x}{n} \sim p(x)$

Now: For any  $\epsilon, \delta > 0$  there exist an  $n$  large enough s. t.

$$H(\mathcal{X}) - \delta \leq -\frac{1}{n} \log p(x_1 \dots x_n) \leq H(\mathcal{X}) + \delta \quad (1)$$

→  $2^{-n(H+\delta)} \geq p(x_1, \dots, x_n) \geq 2^{-n(H-\delta)}$

But:  $p(x_1 \dots x_n)$  is just one of many typical strings with the same number of occurrences of each letter and thus identical a priori probabilities  $p(\text{typical})$ . Then for  $n$  large enough, we also have

$$1 - \epsilon \leq \underbrace{\sum p(\text{typical})}_{= N(\epsilon, \delta) \times p(\text{typical})} \leq 1 \quad (2)$$

# of typical strings

Taking the ratio  $\frac{1}{(1)} \times (2)$  gives us the final result

$$(1 - \epsilon) 2^{n(H - \delta)} \leq N(\epsilon, \delta) \leq 2^{n(H + \delta)}$$



# Quantum Information Theory (Preskill ch. 5)

But:  $p(x_1 \dots x_n)$  is just one of many typical strings with the same number of occurrences of each letter and thus identical a priori probabilities  $p(\text{typical})$ . Then for  $n$  large enough, we also have

$$1 - \epsilon \leq \underbrace{\sum p(\text{typical})}_{\substack{\uparrow = N(\epsilon, \delta) \times p(\text{typical}) \\ \uparrow \text{ \# of typical strings}}} \leq 1 \quad (2)$$

Taking the ratio  $\frac{1}{(1)} \times (2)$  gives us the final result

$$(1 - \epsilon) 2^{n(H - \delta)} \leq N(\epsilon, \delta) \leq 2^{n(H + \delta)}$$

## Conclusion

(Shannons Noiseless Coding theorem)

- \* We can encode all typical strings w/blocks of  $n(H + \delta)$  bits
- \* Atypical strings occur w/prob.  $< \epsilon$ , where  $\epsilon, \delta \rightarrow 0$  for  $n \rightarrow \infty$
- \* An optimal code thus compresses each letter to  $H(X)$  bits

# Quantum Information Theory (Preskill ch. 5)

But:  $p(x_1 \dots x_n)$  is just one of many typical strings with the same number of occurrences of each letter and thus identical a priori probabilities  $p(\text{typical})$ . Then for  $n$  large enough, we also have

$$1 - \epsilon \leq \underbrace{\sum p(\text{typical})}_{\substack{\uparrow = N(\epsilon, \delta) \times p(\text{typical}) \\ \uparrow \text{ \# of typical strings}}} \leq 1 \quad (2)$$

Taking the ratio  $\frac{1}{(1)}$   $\times$  (2) gives us the final result

$$(1 - \epsilon) 2^{n(H - \delta)} \leq N(\epsilon, \delta) \leq 2^{n(H + \delta)}$$

## Conclusion

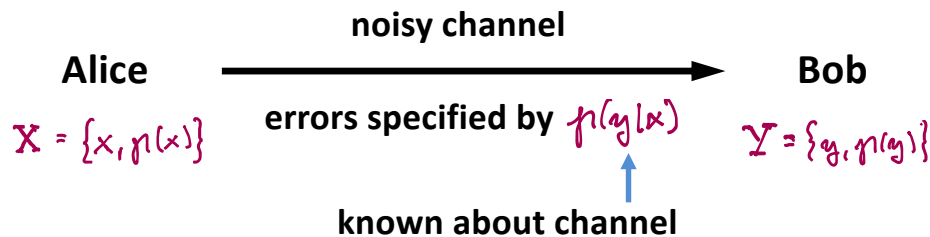
( Shannons Noiseless Coding theorem )

- \* We can encode all typical strings w/blocks of  $n(H + \delta)$  bits
- \* Atypical strings occur w/prob.  $< \epsilon$ , where  $\epsilon, \delta \rightarrow 0$  for  $n \rightarrow \infty$
- \* An optimal code thus compresses each letter to  $H(X)$  bits

# Quantum Information Theory (Preskill ch. 5)

## Joint and Conditional Entropy, Mutual Information

Consider the following scenario:



Bayes Rule:  $p(x|y) = \frac{p(y|x)p(x)}{p(y)}$

known about Alices alphabet

$p(y) = \sum_x p(y|x)p(x)$

Bob uses this to estimate the prob. that Alice sent  $x$  given he received  $y$ . The “width” of the distribution  $p(x|y)$  is thus a measure of Bob’s information gain per letter.

Think about this in terms of joint events

$$\{X, Y\} = \{(x, y), p(x, y)\}$$

→ Joint entropy

$$H(X, Y) = - \sum_{x, y} p(x, y) \log p(x, y)$$

This is a measure of information content per letter in the combined strings

- \* Assume Bob measures the value of a letter  $y$  in the message
- \* He gets  $H(Y)$  bits of info about the letter pair  $x, y$
- \* Bob’s remaining uncertainty about the letter  $X_i$  is then tied to his lack of knowledge about  $X_i$  given that he knows  $y$ .

# Quantum Information Theory (Preskill ch. 5)

Think about this in terms of joint events

$$\{X, Y\} = \{(x, y), p(x, y)\}$$

## Joint entropy

$$H(X, Y) = - \sum_{x, y} p(x, y) \log p(x, y)$$

This is a measure of information content per letter in the combined strings

- \* Assume Bob measures the value of a letter  $y$  in the message
- \* He gets  $H(Y)$  bits of info about the letter pair  $x, y$
- \* Bob's remaining uncertainty about the letter  $X_i$  is then tied to his lack of knowledge about  $X_i$  given that he knows  $y$ .

The entropy of  $X$  conditioned on  $Y$  is therefore

$$H(X, Y) = H(Y) + H(X|Y)$$

$$\Rightarrow H(X|Y) = H(X, Y) - H(Y)$$

## The Conditional Entropy $H(X|Y)$

is the number of bits of info per letter in Alice's message that Bob is missing due to channel errors

– measure of information loss due to errors –

Equivalently, it is the # of extra bits Alice must send to ensure Bob gets the complete message in the presence of channel errors.

# Quantum Information Theory (Preskill ch. 5)

The entropy of  $X$  conditioned on  $Y$  is therefore as

$$H(X, Y) \equiv H(Y) + H(X|Y)$$
$$\Rightarrow H(X|Y) \equiv H(X, Y) - H(Y)$$

## The Conditional Entropy $H(X|Y)$

is the number of bits of info per letter in Alice's message that Bob is missing due to channel errors

– measure of information loss due to errors –

Equivalently, it is the # of extra bits Alice must send to ensure Bob gets the complete message in the presence of channel errors.

Note: From the above it follows that the conditional entropy is given by

$$H(X|Y) = H(X, Y) - H(Y)$$
$$= - \sum_{x,y} p(x,y) \log p(x,y) + \sum_y p(y) \log p(y)$$

We can similarly quantify the # of bits of info about  $X$  that Bob has gained by measuring  $Y$ .

This is the Mutual Information:

$$I(X; Y) \equiv H(X) + H(Y) - H(X, Y)$$
$$\equiv H(X) - H(X|Y) = H(Y) - H(Y|X)$$

Note: When we added the info content of  $X$  to the info content of  $Y$  we overcounted the total info because some info is common to  $X$  and  $Y$ , and must be subtracted to get the proper measure for the Mutual Information

# Quantum Information Theory (Preskill ch. 5)

**Note:** From the above it follows that the conditional entropy is given by

$$\begin{aligned} H(X|Y) &= H(X, Y) - H(Y) \\ &= - \sum_{x,y} p(x,y) \log p(x,y) + \sum_y p(y) \log p(y) \end{aligned}$$

We can similarly quantify the # of bits of info about  $X$  that Bob has gained by measuring  $Y$ .

This is the Mutual Information:

$$\begin{aligned} I(X; Y) &\equiv H(X) + H(Y) - H(X, Y) \\ &\equiv H(X) - H(X|Y) = H(Y) - H(Y|X) \end{aligned}$$

**Note:** When we added the info content of  $X$  to the info content of  $Y$  we overcounted the total info because some info is common to  $X$  and  $Y$ , and must be subtracted to get the proper measure for the Mutual Information

**Note:** We can also quantify the # of bits of info about  $X$  that Bob has gained by measuring  $Y$ .

This is the Mutual Information:

$$\begin{aligned} I(X; Y) &\equiv H(X) + H(Y) - H(X, Y) \\ &\equiv H(X) - H(X|Y) = H(Y) - H(Y|X) \end{aligned}$$

**Note:** When we added the info content of  $X$  to the info content of  $Y$  we overcounted the total info because some info is common to  $X$  and  $Y$ , and must be subtracted to get the proper measure for the Mutual Information

Next Lecture – Final Topic in QIS:

Shannon's Noisy Channel Coding Theorem