

Quantum information and computing

Lecture 13: Quantum information

Jani-Petri Martikainen

`Jani-Petri.Martikainen@helsinki.fi`

`http://www.helsinki.fi/~jamartik`

Department of Physical Sciences

University of Helsinki

Entropy and information

- **Entropy:** How much uncertainty there is in the state of a physical system
- (classical) Suppose we learn the value of a random variable X . **Shannon entropy** quantifies how much information (on average) we gain **after** we learn the value of X .
- Alternatively, entropy of X measures the amount of uncertainty about X **before** we learn its value.

- $$H(X) = H(p_1 \dots p_n) = - \sum_x p_x \log p_x \quad (1)$$

(log is in the base two by convention)

Entropy and information

- Why? Answer: This can be used to quantify the resources needed to store information.
- Shannon asked, what are the minimal resources required to store the information produced by the source (for example radio antenna), in such a way that at later time the information can be reconstructed.
- He found that $H(X)$ bits are required per source symbol. (Shannon's noiseless coding theorem)
- Suppose that a source produces one of four symbols 1, 2, 3, and 4. Without compression two bits of storage space are consumed for each use of the source.
- If, however, 1 is produced with probability $1/2$, 2 with prob. $1/4$, and the other with prob. $1/8$...we can use the bias to compress the source.

Entropy and information

- For example, we could use less bits for the commonly occurring 1.
- For example encoding..1 = 0, 2 = 10, 3 = 110, and 4 = 111...then the average length of the compressed string is $1/2 \cdot 1 + 1/4 \cdot 2 + 1/8 \cdot 3 + 1/8 \cdot 3 = 7/4$ bits!
- matches (incidentally) the entropy of the source since $H(X) = 7/4$ also
- Any further compression will lead to an information loss
- Fundamental measures of information arise as answers to questions about physical resources required to solve some information processing problem!

Entropy and information

- Binary entropy: $H_{bin}(p) = -p \log p - (1 - p) \log(1 - p)$
- This attains its maximum at $p = 1/2$
- How the entropy behaves as we mix two or more prob. distributions?
- Alice has two biased coins. Prob. of heads on the US coin p_U while on the euro heads occurs with probability p_E .
- Alice flips the US coin with prob. q and euro with prob. $1 - q$. Then she tells the heads or tails information to Bob. How much information does Bob gain (on average)?
- Intuitively...at least as much as the average of the information with US coin flip and euro coin flip.

Entropy and information

- ...this means concavity of binary entropy

$$H(qp_U + (1 - q)p_E) \geq qH(p_U) + (1 - q)H(p_E) \quad (2)$$

- Sometimes the inequality is strict because of additional information about coin identity. ($p_U = 1/3$ and $p_E = 5/6$...if heads comes up then Bob has a pretty good indicator that the coin was a euro.)
- **Concavity:** real valued f is concave if $f(px + (1 - p)y) \geq pf(x) + (1 - p)f(y)$
- **Relative entropy:** Entropy like measure of a closeness between probability distributions

$$H(p(x)||q(x)) = \sum_x p(x) \log \frac{p(x)}{q(x)} = -H(X) - \sum_x p(x) \log q(x)$$

Entropy and information

- Why is relative entropy regarded as being like a distance measure? ...Answer: The relative entropy is non-negative $H(p(x)||q(x)) \geq 0$ with equality only if $p(x) = q(x)$ for all x .
- Example, $p(x)$ is a distribution over d outcomes and $q(x) = 1/d$ is the uniform distribution. Then...

$$H(p(x)||q(x)) = \log d - H(X) \geq 0 \quad (3)$$

where the equality applies only if $p(x)$ is a uniform distribution.

- Suppose now that X and Y are two random variables. How is the information content of X related to that of Y

Entropy and information

- Concepts **conditional entropy** and **mutual information** help to clarify this.
- Joint entropy of X and Y is defined as
$$H(X, Y) = - \sum_{x,y} p(x, y) \log p(x, y)$$
(with the obvious extension to more variables)
- If we know the value of Y , we have acquired $H(Y)$ bits of information about the pair (X, Y) . The remaining uncertainty is associated with our uncertainty about X ...
- Entropy of X conditional on knowing Y is defined as

$$H(X|Y) = H(X, Y) - H(Y) \quad (4)$$

- This is a measure of how uncertain we are about the value of X given that we know the value of Y .

Entropy and information

- **Mutual information content of X and Y** measures how much information X and Y have in common.
- Suppose you sum information contents of X and Y . Common information is counted twice, while information not shared is counted only once
- ...subtract away the joint information $H(X, Y)$ and you get the mutual information

$$H(X : Y) = H(X) + H(Y) - H(X, Y) \quad (5)$$

- **Note:** $H(X : Y) = H(X) - H(X|Y)$ relating the conditional entropy and mutual information.

Basic properties of Shannon entropy

1. $H(X, Y) = H(Y, X)$, $H(X : Y) = H(Y : X)$
2. $H(Y|X) \geq 0$ and thus $H(X : Y) \leq H(Y)$ with equality if and only if $Y = f(X)$
3. $H(X) \leq H(X, Y)$, with equality if and only if $Y = f(X)$
4. **Subadditivity:** $H(X, Y) \leq H(X) + H(Y)$ with equality if X and Y are independent.
5. $H(Y|X) \leq H(Y)$ and thus $H(X : Y) \geq 0$ (with equality if X and Y are independent.)
6. **Strong subadditivity:**
 $H(X, Y, Z) + H(Y) \leq H(X, Y) + H(Y, Z)$ (with equality if $Z \rightarrow Y \rightarrow X$ forms a Markov chain)
7. **Conditioning reduces entropy:** $H(X|Y, Z) \leq H(X|Y)$

Data processing inequality

- Information we receive is imperfect due to noise.
- **Data processing inequality:** information of an output of a source can only **decrease with time**
- Suppose that $X \rightarrow Y \rightarrow Z$ is a Markov chain. Then

$$H(X) \geq H(X : Y) \geq H(X : Z) \quad (6)$$

- Moreover, the first inequality is saturated only if , given Y , it is possible to reconstruct X .
- Tells that if X is subjected to noise ,producing Y , then further actions on out part cannot be used to increase the amount of mutual information between the output of the process and the original information X .

Entropy and information

- Shannon entropy measures the uncertainty associated with a classical probability distribution.
- ...generalize to quantum states
- **von Neumann entropy** of ρ

$$S(\rho) = -\text{Tr}(\rho \log \rho) = -\sum_x \lambda_x \log \lambda_x \quad (7)$$

(use base two logarithm as usual)

- For example, completely mixed state $\rho = I/d$ gives the entropy $S(\rho) = \log d$
- Useful to define a quantum version of the relative entropy (between ρ and σ):

$$S(\rho||\sigma) = \text{Tr}(\rho \log \rho) - \text{Tr}(\rho \log \sigma) \quad (8)$$

Basic properties of von Neumann entropy

1. Non-negative and zero only for pure states
2. In a d -dimensional Hilbert space entropy is at most $\log d$
3. If a composite system AB is in a pure state, then $S(A) = S(B)$
4. Suppose p_i are probabilities and the states ρ_i have support on orthogonal subspaces. Then

$$S\left(\sum_i p_i \rho_i\right) = H(p_i) + \sum_i p_i S(\rho_i) \quad (9)$$

5. **Joint entropy theorem:** Suppose p_i are probabilities and $|i\rangle$ are orthogonal states for A , and ρ_i is any set of density operators for another system B Then

$$S\left(\sum_i p_i |i\rangle\langle i| \otimes \rho_i\right) = H(p_i) + \sum_i p_i S(\rho_i) \quad (10)$$

Basic properties of von Neumann entropy

- Proof for $S(\sum_i p_i \rho_i) \leq \sum_i p_i S(\rho_i) + H(p_i)$
- Begin with a pure state $\rho_i = |\psi_i\rangle\langle\psi_i|$. Suppose ρ_i are states of A and introduce an auxiliary system B with an orthonormal basis $|i\rangle$ corresponding to the index i of the probabilities p_i .
- Define $|AB\rangle = \sum_i \sqrt{p_i} |\psi_i\rangle |i\rangle$. Since $|AB\rangle$ is a pure state we have

$$S(B) = S(A) = S\left(\sum_i p_i |\psi_i\rangle\langle\psi_i|\right) = S(\rho) \quad (11)$$

- Suppose we perform a projective measurement on the system B so that after the measurement $\rho^{B'} = \sum_i p_i |i\rangle\langle i|$. However, **projective measurements never decrease entropy** (see the book for the simple proof), so

$$S(\rho) = S(B) \leq S(B') = H(p_i)$$

Basic properties of von Neumann entropy

- Since $S(\rho_i) = 0$ for pure states the statement follows...this is easy to generalize to mixed states ρ_i as well.

Entropy and information

- For example, from the joint entropy theorem it follows that $S(\rho \otimes \sigma) = S(\rho) + S(\sigma)$
- By analogy we can also define quantum joint and conditional entropies as well as quantum mutual information.
- For a joint system AB the joint entropy is $S(A, B) = -\text{Tr}(\rho^{AB} \log(\rho^{AB}))$
- Conditional entropy and mutual information through

$$S(A|B) = S(A, B) - S(B) \quad (12)$$

$$\begin{aligned} S(A : B) &= S(A) + S(B) - S(A, B) \\ &= S(A) - S(A|B) = S(B) - S(B|A) \end{aligned} \quad (13)$$

Entropy and information

- **Some properties of Shannon entropy fail to hold in the quantum case!**
- for two random variables X and Y , the inequality $H(X) \leq H(X, Y)$ holds.
- Makes sense: Surely we cannot be more uncertain about X than we are about the joint state of X and Y .
- This intuition fails in the quantum case: consider a state $(|00\rangle + |11\rangle)/\sqrt{2}$. This is a pure state so $S(A, B) = 0$
- On the other hand the reduced density op. in A is $I/2$ which has an entropy equal to one....i.e.
 $S(B|A) = S(A, B) - S(A)$ is negative!

Entropy and information

- **Subadditivity:** suppose A and B have a joint state ρ^{AB} then the joint entropy satisfies

$$S(A, B) \leq S(A) + S(B) \quad (14)$$

and (Triangle inequality...analog of $H(X, Y) \geq H(X)$)

$$S(A, B) \geq |S(A) - S(B)| \quad (15)$$

- Entropy is a concave function of its inputs:

$$S\left(\sum_i p_i \rho_i\right) \geq \sum_i p_i S(\rho_i) \quad (16)$$

- **Strong subadditivity:** For a trio of quantum systems

$$S(A, B, C) + S(B) \leq S(A, B) + S(B, C) \quad (17)$$

Quantum information theory

- Classical information theory: problem of sending classical information (letters, speech, bits...) over communication channels.
- Quantum information theory: How does the picture change if we can build quantum mechanical communication channels?
- Can we transmit information more efficiently?
- What resources are required for various tasks?
- Can we use quantum mechanics for secure communication?

Accessible information

- Alice has a source producing symbols $X = 0 \dots n$ according to a probability distribution $p_0 \dots p_n$
- Bob aims to determine the value of X as well as he can.
- Alice prepares quantum state ρ_X chosen from some set $\rho_0 \dots \rho_n$ and gives this state for Bob who makes a measurement.
- Based in the measurement result Y Bob tries to make a guess for X .
- A good measure on how much information Bob has gained is the mutual information $H(X : Y)$.

Accessible information

- By the data processing inequality we know that Bob can infer X from Y if and only if $H(X : Y) = H(X)$...however in general $H(X : Y) \leq H(X)$.
- Closeness of $H(X : Y)$ and $H(X)$ provides a measure of how well Bob can determine X .
- Bob should choose a measurement which maximizes $H(X : Y)$.
- We define **accessible information** to be a maximum of the mutual information $H(X : Y)$ over all possible measurement schemes.

Accessible information

- In classical theory accessible information is not too interesting, since in principle we can always distinguish two classical states.
- However, there is no way to distinguish reliably two non-orthogonal quantum states.
- If Alice prepares $|\psi\rangle$ with prob. p and state $|\phi\rangle$ not orthogonal to $|\psi\rangle$ with prob. $(1 - p)$ accessible information is always less than $H(p)$! Contrast with classical information theory.
- Caveat: context in which the concept of accessible information makes sense classically.
- Alice prepares 0 or 1 according to one of two prob. distributions $(p, 1 - p)$ or $(q, 1 - q)$. Given the state, Bob's task is to figure out which distribution was used.

Accessible information

- Clearly, Bob cannot always do this with perfect reliability!
- This classical thought experiment is analogous to accessible information for a quantum system prepared in one of a set of mixed states.
- What is important is that the pure quantum states enjoy markedly different distinguishability properties from their classical counterparts 0:s and 1:s
- Remember no-cloning theorem...
- Connection between no-cloning and accessible information?
- Suppose Alice prepares $|\psi\rangle$ and $|\phi\rangle$ with probabilities p and $1 - p$.

Accessible information

- Suppose Bob's accessible information about these states was $H(p)$, that is laws of QM allow Bob to obtain enough information to determine which state, $|\psi\rangle$ or $|\phi\rangle$, Alice had prepared.
- This would allow cloning!: He would perform a measurement telling which state Alice prepared, and once he had made the identification, could make copies of whichever state Alice had given him.
- No-cloning theorem can also be seen as a consequence that accessible information must be less than $H(p)$
- **No general method of calculating the accessible information is known!. However, variety of bounds can be proven.**

The Holevo bound

- **The Holevo bound:** Suppose Alice prepares a state ρ_X where $X = 0 \dots n$ with probabilities $p_0 \dots p_n$. Bob performs a measurement with a POVM elements $\{E_y\} = \{E_0 \dots E_m\}$ on that state, with measurement outcome Y . The Holevo bound states that for any such measurement Bob may do:

$$H(X : Y) \leq S(\rho) - \sum_x p_x S(\rho_x), \quad (18)$$

where $\rho = \sum_x p_x \rho_x$.

- Bound on accessible information.
- The quantity appearing on the right hand side is called **Holevo χ quantity**
- Proof: Three quantum systems P , Q , and M . Q is the one Alice gives to Bob. P and M are fictitious systems to ease the proof

The Holevo bound

- P ("preparation system") has an orthonormal basis $|x\rangle$ whose elements correspond to labels $0 \dots n$ on possible preparations for the quantum system Q .
- M is the "measuring device" and has a basis $|y\rangle$ whose elements correspond to the possible outcomes $0 \dots n$ of Bob's measurement.
- The initial state is assumed to be

$$\rho^{PQM} = \sum_x p_x |x\rangle\langle x| \otimes \rho_x \otimes |0\rangle\langle 0| \quad (19)$$

where we write the tensor product decomposition in the order PQM

- To describe measurement we introduce quantum operation \mathcal{E} acting on Q and M only, whose action is to perform a measurement with POVM elements $\{E_y\}$ on the system Q and store the result in system M .

The Holevo bound

$$\mathcal{E}(\sigma \otimes |0\rangle\langle 0|) = \sum_y \sqrt{E_y} \sigma \sqrt{E_y} \otimes |y\rangle\langle y| \quad (20)$$

where σ is any state of Q

- Use primes to denote states of PQM after measurement. We have $S(P : Q) = S(P : Q, M)$ since M is initially uncorrelated with P and Q . Also $S(P : Q, M) \geq S(P' : Q', M')$ since applying the quantum operation \mathcal{E} cannot increase the mutual information between P and QM .
- Finally $S(P' : Q', M') \geq S(P' : M')$ since discarding systems can't increase mutual information... These imply that

$$S(P' : M') \leq S(P : Q) \quad (21)$$

This result is easily understood to be the Holevo bound!

The Holevo bound

- The quantity on the right hand side. Note that $\rho^{PQ} = \sum_x p_x |x\rangle\langle x| \otimes \rho_x$. From this it follows

$$S(P) = H(p_x) \quad S(Q) = S(\rho) \quad (22)$$

and

$$S(P, Q) = H(p_x) + \sum_x p_x S(\rho_x) \quad (23)$$

- Using the theorem 11.10 on the book $S(\rho) - \sum_x p_x S(\rho_x) \leq H(X)$
- ...we get

$$S(P : Q) = S(P) + S(Q) - S(P, Q) = S(\rho) - \sum_x p_x S(\rho_x)$$

- We still need the quantity on the LHS of the Holevo bound

The Holevo bound

- Note that,

$$\rho^{P'Q'M'} = \sum_{xy} p_x |x\rangle\langle x| \otimes \sqrt{E_y} \rho_x \sqrt{E_y} \otimes |y\rangle\langle y| \quad (24)$$

- If we now trace out the system Q' and use the observation that the joint distribution $p(x, y)$ satisfies $p(x, y) = p_x p(y|x) = p_x \text{Tr}(\rho_x E_y) = p_x \text{Tr}(\sqrt{E_y} \rho_x \sqrt{E_y})$ we get

$$\rho^{P'M'} = \sum_{xy} p(x, y) |x\rangle\langle x| \otimes |y\rangle\langle y| \quad (25)$$

-whence $S(P' : M') = H(X : Y)$ which is what we have on the LHS of the Holevo bound!

The Holevo bound: example

- Remember that $S(\rho) - \sum_x p_x S(\rho_x) \leq H(X)$ with equality only if ρ_x have orthogonal support. Suppose that ρ_x don't have an orthogonal support so the inequality is strict.
- Then Holevo bound implies $H(X : Y) < H(X)$ and it is impossible for Bob to determine X with perfect reliability from his measurement outcomes Y .
- Take a concrete example where Alice prepares a single qubit in one of the two quantum states according to the outcome of a coin toss.
- Heads implies $|0\rangle$ and tails $\cos \theta |0\rangle + \sin \theta |1\rangle \dots$ it follows that

$$\rho = \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \cos^2 \theta & \cos \theta \sin \theta \\ \cos \theta \sin \theta & \sin^2 \theta \end{bmatrix} \quad (26)$$

The Holevo bound: example

- Eigenvalues are $(1 \pm \cos \theta)/2$ and therefore the Holevo bound is given by the binary entropy $H((1 + \cos \theta)/2)$
- Holevo bound is maximized for $\theta = \pi/2$ corresponding to Alice choosing an orthogonal set.
- as $\theta \rightarrow 0$ Bob has less and less ability to distinguish states...finally he can do no better than tossing a coin himself!

Data compression

- What are the minimal physical requirements needed to store an information source?
- (Classical) **Shannon's noiseless channel coding theorem:** quantifies the extent to which we can compress information.
- Assume a source of random variables X_1, X_2, \dots different uses of the sources are independent and identically distributed i.e. **the source is i.i.d. information source**
- Real world sources are not exactly like this...take language for example.
- Suppose a source produces bits X_1, X_2, X_3 each being 0 with probability p and one with probability $1 - p$.

Data compression

- Key idea behind Shannon's theorem is to divide the possible sequences of values $x_1, x_2 \dots$ for random variables $X_1, X_2, X_3 \dots$ up into two types...
- Common **typical sequences** and unusual **atypical sequences**
- How? When n gets large we expect fraction p of the symbols to be zeros...
- Sequences for which this is true are typical. Combining this with the assumption of a source independence

$$p(x_1 \dots x_n) = p(x_1) \dots p(x_n) \approx p^{np} (1 - p)^{(1-p)n} \quad (27)$$

for typical sequences

- Taking the logarithms..., $-\log p(x_1 \dots x_n) = nH(X)$ where $H(X) = -p \log(p) - (1 - p) \log(1 - p)$ is the **entropy of the source distribution**

Data compression

- Thus $p(x_1 \dots x_n) \approx 2^{-nH(X)}$ from which we see that there can be at most $2^{nH(X)}$ typical sequences
- To compress the output $x_1 \dots x_n$ of the source, we check if the sequence is typical. If it's not, we give up. (Happens rarely if n is large.)
- Otherwise, for typical sequences it only requires $nH(X)$ bits to uniquely identify a particular typical sequence...compress the source to the corresponding string of $nH(X)$ bits...decompress later.
- Small error probability can be removed with refinements.
- Compression requires large n ...also this restriction can be removed. Also schemes for mapping into compressed sequences have been worked out...

Data compression

- Data compression depends on the output distribution of the source...use universal compression algorithms.
- **Shannon's noiseless channel coding theorem:** Suppose $\{X_i\}$ is an i.i.d. information source with entropy rate $H(X)$ and R is the compression scheme rate. suppose $R > H(X)$. Then there exists a reliable compression scheme of rate R for the source. Conversely if $R < H(X)$ then any compression will not be reliable.
- This classical theorem can be extended to quantum information.
- Treat quantum states as if they are information and ask information theoretical questions about those states.

Data compression

- Quantum information source? Our definition is based on the idea that entanglement is what we are trying to compress and decompress.
- Quantum source described by a Hilbert space H and a density matrix ρ on that Hilbert space.
- We imagine that ρ is merely a part of the larger system which is in a pure state. (mixed nature of ρ due to entanglement with the remainder of the system)
- Compression scheme \mathcal{C}^n of rate R takes states in $H^{\otimes n}$ to states in a 2^{nR} -dimensional state space, the compressed space.
- Compressed space represents nR qubits.
- Decompression \mathcal{D}^n takes us back to the original space.

Data compression

- Quantum version of typical sequences: Suppose the density operator associated with the source has orthonormal decomposition

$$\rho = \sum_x p(x) |x\rangle\langle x| \quad (28)$$

- eigenvalues $p(x)$ obey same rules a probability distribution. Also $H(p(x)) = S(\rho) \dots \epsilon$ -typical sequence

$$\left| \frac{1}{n} \log \left(\frac{1}{p(x_1)p(x_2) \dots p(x_n)} \right) - S(\rho) \right| \leq \epsilon \quad (29)$$

- **Schumacher's noiseless channel coding theorem:** Let $\{H, \rho\}$ be a i.i.d. quantum source. If $R > S(\rho)$ then there exists a reliable compression scheme of rate R for the source. If $R < S(\rho)$ then any compression scheme is not reliable.

Classical information over noisy q. channel

- How much information can be transmitted through a noisy channel?
- For example, 1000 uses of the channel can be used to transmit 500 bits of information using some error correcting code.
- Code has a rate $500/1000 = 1/2$... **What is the maximum rate?** This number is known as the **capacity** of the channel.
- For classical channels use **Shannon's noisy channel coding theorem**: Capacity $C(\mathcal{N}) = \max_{p(x)} H(X : Y)$, where the maximum is taken over all input distributions $p(x)$ for X , for one use of the channel and Y is the corresponding induced random variable at the output channel.

Classical information over noisy q. channel

- Example: Take a symmetric bit flip channel which flips bits with probability p . Input distribution $p(0) = q$, $p(1) = 1 - q$. We have for mutual information

$$\begin{aligned} H(X : Y) &= H(Y) - H(Y|X) \\ &= H(Y) - \sum_x p(x) H(Y|X = x) \end{aligned}$$

- But for each x , $H(Y|X = x) = H(p)$ so $H(X : Y) = H(Y) - H(p)$.

- This is maximized by choosing $q = 1/2$, so $H(Y) = 1$ and

$$C(\mathcal{N}) = 1 - H(p) \tag{30}$$

(If $p = 1/2$ no information can be transmitted)

Classical information over noisy q. channel

- Suppose Alice encodes her message M as a quantum state and sends it through a noisy quantum channel to Bob. What is the capacity of the channel?
- Problem is not completely solved, but some things are known.
- Capacity for a channel \mathcal{E} is known if Alice encodes her message using product states of form $\rho_1 \otimes \rho_2 \dots$ where each ρ_i are potential inputs for one of the channel.
- Note! This still allows for Bob to decode using measurements entangled across multiple uses of the channel.
- Holevo-Schumacher-Westmoreland (HSW) theorem gives the product state capacity $C^{(1)}(\mathcal{E})$

Classical information over noisy q. channel

- Holevo-Schumacher-Westmoreland (HSW) theorem: Let \mathcal{E} be a trace preserving quantum operation. Define

$$\chi(\mathcal{E}) = \max_{p_j, \rho_j} \left[S \left(\mathcal{E} \left(\sum_j p_j \rho_j \right) \right) - \sum_j p_j S(\mathcal{E}(\rho_j)) \right] \quad (31)$$

where the maximum is over all ensembles $\{p_j, \rho_j\}$ of possible input states ρ_j to the channel. Then

$\chi(\mathcal{E}) = C^{(1)}(\mathcal{E})$ is the product state capacity for the channel \mathcal{E}

- It is **believed** by many that allowing entangled signals does not increase the capacity.
- Example: Any quantum channel \mathcal{E} can be used to transmit classical information (unless \mathcal{E} is constant).

Classical information over noisy q. channel

- If the channel is not constant then there exists pure states $|\psi\rangle$ and $|\phi\rangle$ such that $\mathcal{E}(|\psi\rangle\langle\psi|) \neq \mathcal{E}(|\phi\rangle\langle\phi|)$
- Substituting an ensemble made up of these states with equal probabilities $1/2$ into the definition we find

$$C^{(1)}(\mathcal{E}) \geq S\left(\frac{\mathcal{E}(|\psi\rangle\langle\psi|) + \mathcal{E}(|\phi\rangle\langle\phi|)}{2}\right) - \frac{1}{2}\mathcal{E}(|\psi\rangle\langle\psi|) - \frac{1}{2}\mathcal{E}(|\phi\rangle\langle\phi|) > 0$$

where the second inequality follows from the strict concavity of the entropy.

- Take a depolarizing channel with parameter p ... then

$$\mathcal{E}(|\psi_j\rangle\langle\psi_j|) = p|\psi_j\rangle\langle\psi_j| + (1-p)\frac{I}{2} \quad (32)$$

Classical information over noisy q. channel

- This quantum state has eigenvalues $(1 + p)/2$ and $(1 - p)/2$...from which it follows that

$$S(\mathcal{E}(|\psi_j\rangle\langle\psi_j|)) = H\left(\frac{1+p}{2}\right) \quad (33)$$

which is independent of the input state $|\psi_j\rangle$.

- Thus maximum in the capacity is reached by maximizing the entropy $S(\sum_j \mathcal{E}(|\psi_j\rangle\langle\psi_j|))$, which may be done by choosing $|\psi_j\rangle$ to form an orthonormal basis for the state space of a single qubit.
- This gives the product state capacity as

$$C(\mathcal{E}) = 1 - H\left(\frac{1+p}{2}\right) \quad (34)$$

Quantum information over noisy q. channels

- How much quantum information may be transmitted through a noisy quantum channel?
- Less well understood than the problem of classical information over quantum channels.
- For some discussion see the book...

Idea: Entanglement distillation

- Idea: Alice and Bob convert some large number of copies of a known pure state $|\psi\rangle$ into as many copies of the Bell state $(|00\rangle + |11\rangle)/\sqrt{2}$ as possible using local operations and classical communication.
- We want them to succeed, not exactly, but with high fidelity.
- Entanglement dilution would be the reverse process.
- Why?
- Suppose we take entanglement as a “resource”...we need a way to quantify the amount of entanglement.

Idea: Entanglement distillation

- Take the Bell state $(|00\rangle + |11\rangle)/\sqrt{2}$ as our standard unit of entanglement.
- A potential approach to defining the amount of entanglement present in $|\psi\rangle$ is to imagine a large number (n) of Bell states and are asked to produce as many (high-fidelity) copies of $|\psi\rangle$ as possible using just local operations and classical communication.
- If the number of copies of $|\psi\rangle$ is m we define the ratio n/m to be the **entanglement of formation** of the state $|\psi\rangle$.
- Alternatively, go from m copies of $|\psi\rangle$ to n Bell states... n/m is the **distillable entanglement**.
- Understanding entanglement in systems with three or more components is still in its infancy....

Idea: Quantum cryptography

- Quantum cryptography or **quantum key distribution** for secure communication.
- In private key cryptography Alice has an **encoding key** while Bob has a matching **decoding key** and these are used to encrypt and decrypt the message.
- One time pad: Alice and Bob have the same key which is of same length as the message (random for example). Alice sums the message and the key while Bob just subtracts.
- Provably secure as long as the keys remain secret.
- Public key cryptography relies on unproven mathematical assumptions about the difficulty of solving certain problems (factoring).

Idea: Quantum cryptography

- Major difficulty in private key cryptosystems is secure distribution of keys!
- Quantum cryptography can solve this key distribution problem as long as we have a channel with a sufficiently low error rate.
- Eve (eavesdropper) cannot gain any information from the qubits transmitted from Alice to Bob without disturbing their state.
- For this reason Alice and Bob can detect the eavesdropper and discard those bits of the key which might have leaked.
- **Information gain implies disturbance:** In any attempt to distinguish between two non-orthogonal quantum states, information gain is only possible at the expense of introducing disturbance to the signal.

Idea: Quantum cryptography

- ...Let $|\psi\rangle$ and $|\phi\rangle$ be non-orthogonal quantum states Eve is trying to obtain information about. She unitarily interacts the state with an ancilla prepared in a standard state $|u\rangle$. She gets...

$$|\psi\rangle|u\rangle \rightarrow |\psi\rangle|v\rangle \quad (35)$$

$$|\phi\rangle|u\rangle \rightarrow |\phi\rangle|v'\rangle \quad (36)$$

- Eve would like $|v\rangle$ and $|v'\rangle$ to be different so that she can acquire information about the state identity. However, inner product is preserved under unitary transformations so...

$$\langle v|v'\rangle\langle\psi|\phi\rangle = \langle u|u\rangle\langle\psi|\phi\rangle \quad (37)$$

Idea: Quantum cryptography

- ...and

$$\langle v|v'\rangle = \langle u|u\rangle = 1 \quad (38)$$

which implies that $|v\rangle$ and $|v'\rangle$ must be identical.

- By transmitting non-orthogonal states between Alice and Bob they can establish an upper bound on any noise or eavesdropping occurring in the channel.
- These checks are interspersed randomly among the data qubits so that the same error bound applies to the data bits as well.
- See BB84 protocol etc. for the detailed procedure.
- Quantum key distribution has been experimentally demonstrated using (for example) commercial fiber-optic components.

Some comments on the final exam

- You should know the basic notation: common single- and two-qubit gates like Pauli-matrices, Hadamard gate, CNOT. (Also their notation in quantum circuits.)
- Basic concepts of QM, for example density matrix, reduced density matrix...
- You should be familiar with basic ideas and concepts involved in quantum algorithms. However, I will not ask for extensive calculations.
- General understanding of physical requirements for quantum computing.