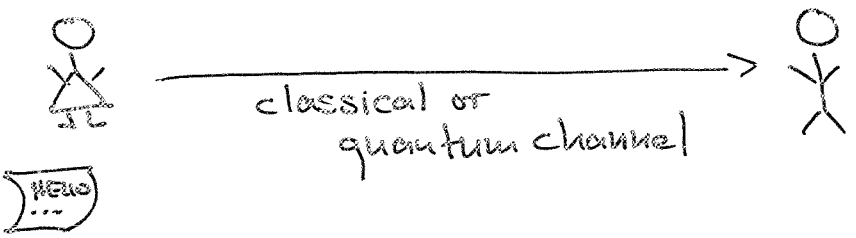
SSUSP67: Quantum Communication

Stephanie Wehner, steph@locc.la

What we'll be discussing today:

Challenge: How can Alice send information to Bob?



Outline:

1. How many qubits does Alice need to send to convey n classical bits?
2. Quantifying information formally
   - Entropies
   - ... and their operational interpretations
   - Smoothing (what the _ is that??)
   - Some useful properties of entropies and example applications
   - Relations between different forms of entropies

(3) The decoupling theorem (... if you were to learn just one theorem ...)

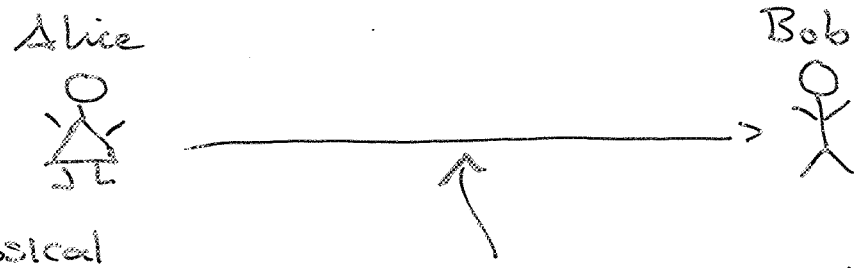
- What does it say intuitively?
- Origins of the decoupling theorem:
  Sending information over a noisy channel
- Formal statement
  + averages
  + probabilistic
  + ... and a converse!

(4) Examples!! ... but you'll have to work a bit yourselves ... ◎◎

see notes on tutorial session

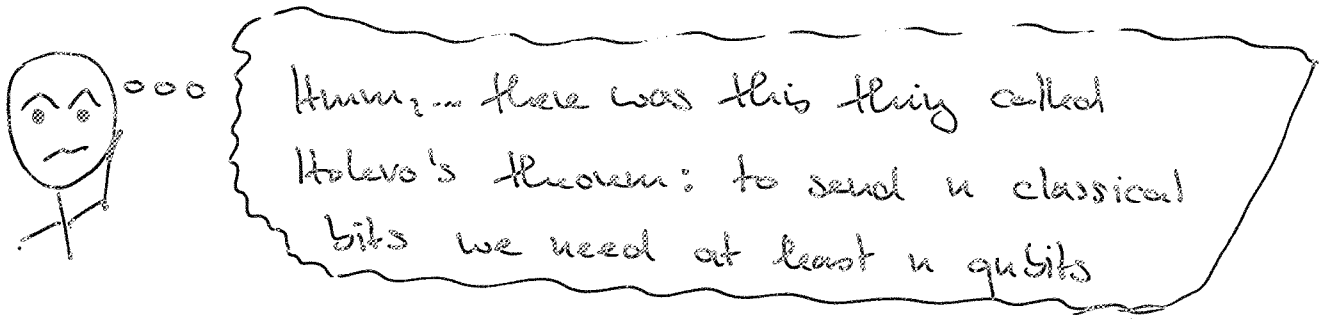
① Simple example: sending classical information over a noiseless quantum channel

Problem:



classical bits $x_1, \ldots, x_n \in \{0,1\}$

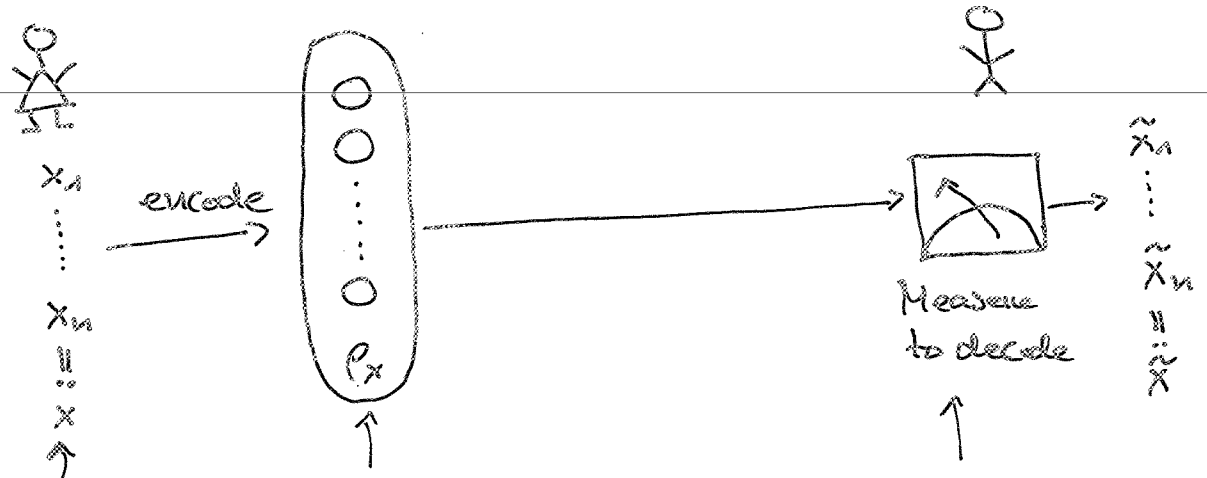
How many qubits does Alice need to send to Bob to convey n classical bits?



Quantum information demystified:

You already have all the tools to answer this question yourself!

Let's examine this problem in more detail....



n-classical bits (Let's make it hard: chosen uniformly at random $p_x = \frac{1}{2^n}$)

k qubits

POVM with measurement operators $\{M_x\}_{x \in \{0,1\}^n}$

$\uparrow$ string output by Bob

Sending information reliably: $\tilde{x} = x$!

Case 1: $k \geq n$ (Rate $R = \frac{n}{k} \leq 1$)

Can send reliably: • encode each classical bit in one qubit

$0 \to |0\rangle \quad 1 \to |1\rangle$

• Bob measures each qubit in $\{|0\rangle, |1\rangle\}$ basis to recover $\tilde{x} = x$.

Case 2: $k < n$ (Rate $R = \frac{n}{k} > 1$)

Cannot send reliably!

Measure of success: probability that Bob outputs the right string (on average)

$$P_{succ}(n) := \max_{\{\rho_x, M_x\}_x} \frac{1}{2^n} \sum_{x \in \{0,1\}^n} \underbrace{\mathrm{Tr}(M_x \rho_x)}_{\text{probability that Bob outputs "x" if the string was "x"}}$$

($p_x = \frac{1}{2^n}$)

n: # of bits

$\{\rho_x, M_x\}_x$: encodings / decoding

Cannot send reliably $\Leftrightarrow$ P_{succ} is very small.

Let's show that $P_{succ}(n)$ goes to 0 for large n whenever $k<0$:

Fix any encodings $\{\rho_x\}_x$ and decoding measurements $\{M_x\}_x$

$$P_{succ}(n) = \frac{1}{2^n} \sum_{x\in\{0,1\}^n} Tr(M_x \rho_x) \quad \text{(1) Definition}$$

$$\leq \frac{1}{2^n} \sum_{x\in\{0,1\}^n} Tr(M_x) \quad \text{(2)} \quad \rho_x \leq \mathbb{I}_k$$

and hence for

$$\rho_x = \sum_j \lambda_j |j\rangle\langle j|$$

$$Tr(M_x\rho_x) = \sum_j \underbrace{\lambda_j}_{\substack{\geq 0 \\ \leq 1}} \underbrace{\langle j|M_x|j\rangle}_{\geq 0}$$

$$\leq \sum_j \langle j|M_x|j\rangle = Tr(M_x)$$

$$\leq \frac{1}{2^n} Tr(\mathbb{I}_k) \quad \text{(3)} \quad \sum_x M_x \leq \mathbb{I}_k$$

since $\{M_x\}_x$ is a POVM on k qusits

$$= \frac{1}{2^n} 2^k = \underbrace{2^{k-n}}_{\text{SMALL!}}$$

$R = \frac{n}{k}$

$n = Rk$

$$= 2^{k-Rk} = 2^{-k(R-1)} = 2^{-k\gamma} \qquad \gamma = R-1 > 0 !$$

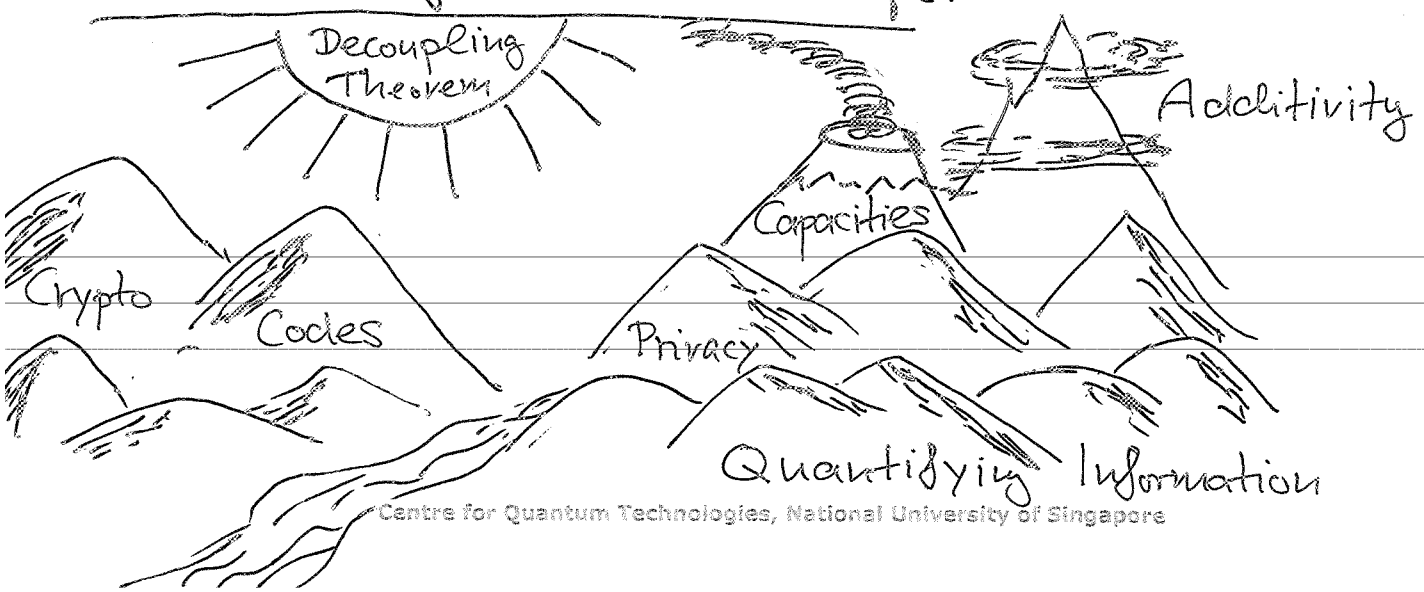
What we've shown: $P_{succ}(n=Rk) \leq 2^{-k(R-1)}$

What it means: If n is too large ($n>k$), or equivalently, the rate R at which Alice tries to send information is too high ($R>1$), then there is no encoding or decoding scheme that can send that many bits reliably

⇒ Need at least $k \geq n$ qubits to send n classical bits.

..... used only basic facts!

Quantum information landscape:



(2) Quantifying information formally

[1] Entropy

von Neumann to Shannon:

"
.... You should call it entropy for two reasons. First of all, your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more importantly, no one really knows what entropy really is so in a debate you will always have the advantage."

Concept: von Neumann entropy

$$\rho = \sum_j \lambda_j |j\rangle\langle j|$$

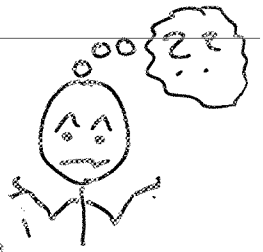
eigenvalues, eigenbasis

$$H(\rho) = -\mathrm{tr}(\rho \log \rho)$$

$$= -\sum_j \lambda_j \log \lambda_j$$

Same as Shannon entropy in the classical case.

Bad news: Other entropies become more and more important....



.... we will focus on other ones here!

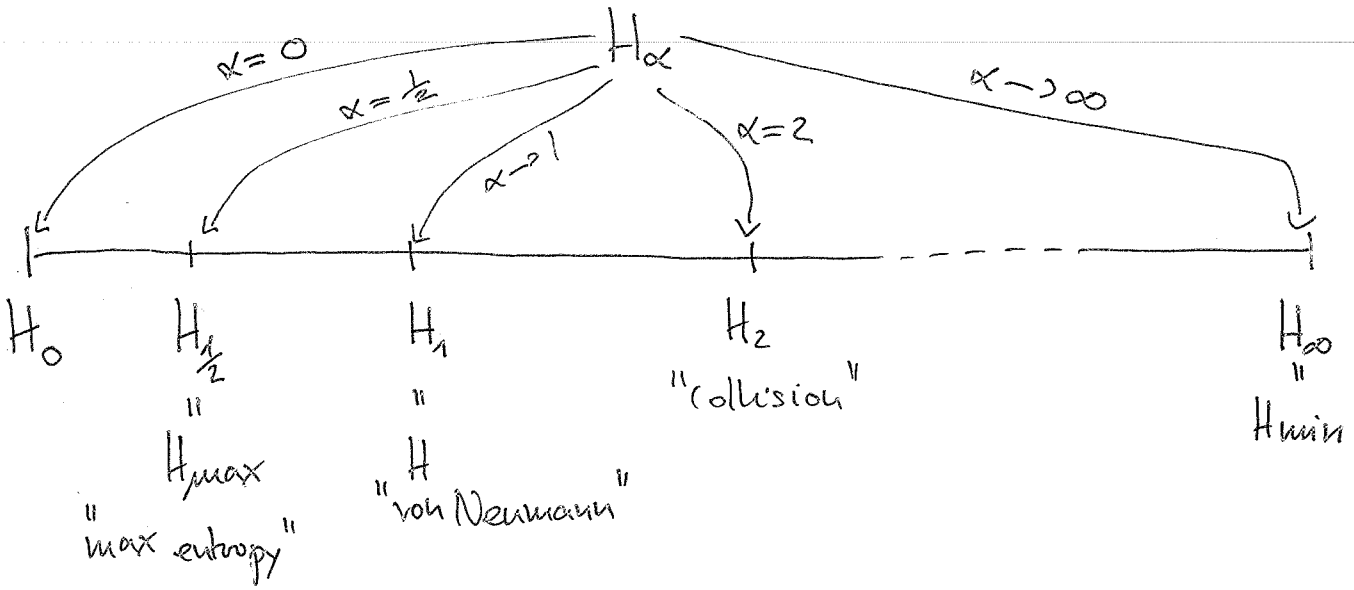
Good news: ... they have intuitive operational interpretations!

-- and interesting applications!

Rényi entropy: $0 \leq \alpha$

$$H_\alpha(\rho) = \frac{1}{1-\alpha} \log \mathrm{Tr}(\rho^\alpha)$$

Important special cases:



$$H_0 \geq H_{max} \geq H \geq H_{min}$$

Special cases more explicitly: $\rho = \sum_j \lambda_j |j\rangle\langle j|$

Min-entropy $H_{min}(\rho) = -\log \max_j \lambda_j$

Max-entropy $H_{max}(\rho) = \frac{1}{2} \log \mathrm{Tr}\sqrt{\rho}$

H_0 $H_0(\rho) = \log \mathrm{Tr}(\rho^0) = \log \mathrm{rank}(\rho)$

Properties they all have in common:

- ρ: d-dimensional state

$$0 \leq H_\alpha(\rho) \leq \log d$$

pure state: $\rho = |\cdot\rangle\langle\cdot|$ (only one symbol/eigenstate occurs)

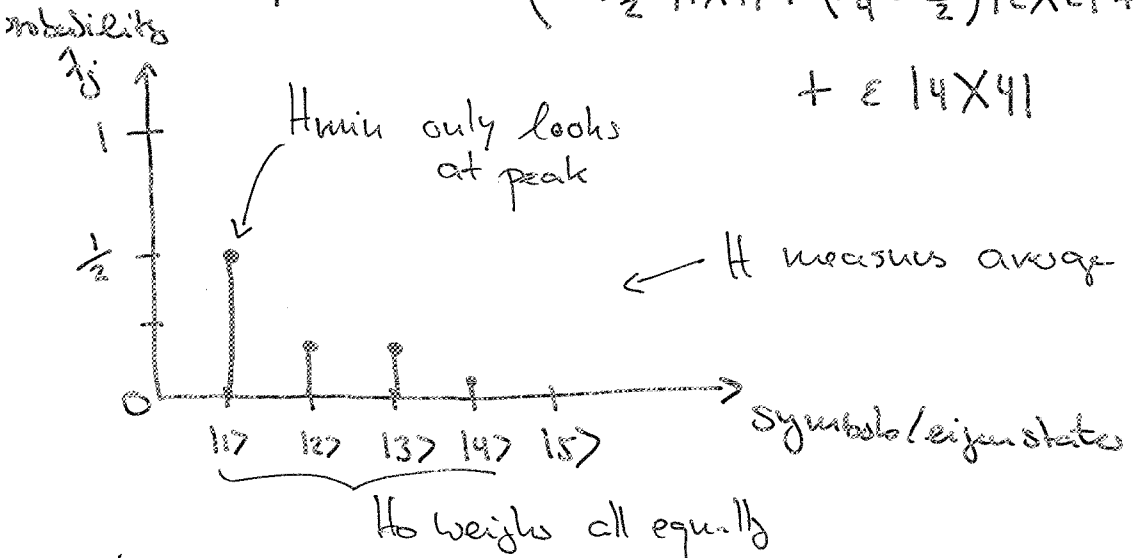
fully mixed: $\rho = \frac{\Pi}{d}$ (all symbols/eigenstates are equally likely)

- for pure and mixed states all entropies are the same: if ρ is pure or fully mixed $H_\alpha = H_\beta$ $\forall \alpha, \beta$

So, why care? What makes them different?

Example:

$$\rho = \frac{1}{2}|1\rangle\langle 1| + \left(\frac{1}{4}-\frac{\varepsilon}{2}\right)|2\rangle\langle 2| + \left(\frac{1}{4}-\frac{\varepsilon}{2}\right)|3\rangle\langle 3| + \varepsilon\,|4\rangle\langle 4|$$



$$H_{min}(\rho) = -\log \max_j \lambda_j = -\log \frac{1}{2} = \log 2 = 1$$

$$H_0(\rho) = \log \text{rank}(\rho) = \log 4 = 2$$

$$H_{max}(\rho) = \frac{1}{2}\log \text{Tr}\sqrt{\rho} = \frac{1}{2}\log\left(\frac{1}{\sqrt{2}} + 2\sqrt{\frac{1}{4}-\frac{\varepsilon}{2}} + \sqrt{\varepsilon}\right) \approx 1.55$$

$$H(\rho) = -\left(\frac{1}{2}\log\frac{1}{2} + 2\left(\frac{1}{4}-\frac{\varepsilon}{2}\right)\log\left(\frac{1}{4}-\frac{\varepsilon}{2}\right) + \varepsilon\log\varepsilon\right) \approx 1.5$$

Intuition

H_{min} — measures worst/best case behaviour
— used in one shot information theory and crypto (see tutorial!)

H_0 — risk avoidance: any positive probability weighted equally
— often used to measure storage size (see example later)

H_{max} — similar to H_0, used in one shot info. theory.

H — used in the asymptotic limit of many copies

So far we only considered a single system when talking about entropies.

Let's say we have two systems:

Suppose first that the joint state $\rho_{AB} = |\psi_{AB}\rangle\langle\psi_{AB}|$ is pure.

Useful trick: Write $|\psi_{AB}\rangle = \sum_j \sqrt{\lambda_j}\, |j_A\rangle |j_B\rangle$ in its Schmidt decomposition.

Compute the reduced states: $\rho_A = \sum_j \lambda_j |j_A\rangle\langle j_A|$

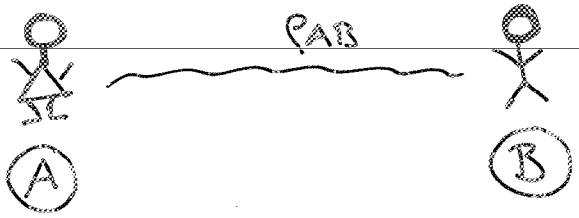
$\rho_B = \sum_j \lambda_j |j_B\rangle\langle j_B|$

↑ same eigenvalues!

⇒ Since all the entropies only depend on the eigenvalues, and <u>not</u> on the eigenbasis

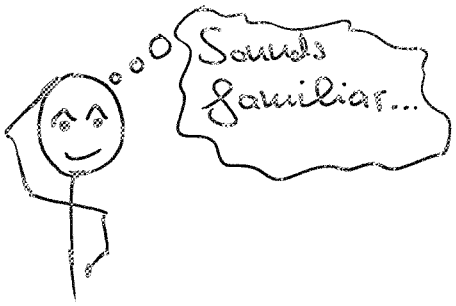
$$H_\alpha(\rho_A) = H_\alpha(\rho_B) \quad \forall \alpha$$

Useful: more entangled --> higher local entropy



Quantifying the amount of information that B has about A:

- Conditional entropies



von Neumann

$$H(A|B) = H(AB) - H(B)$$

intuition:

total amount of "ignorance"

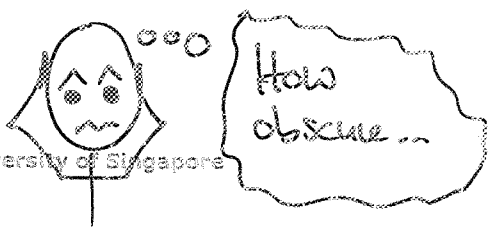
given amount of "knowledge"

Conditional min/max entropies:

General definition:

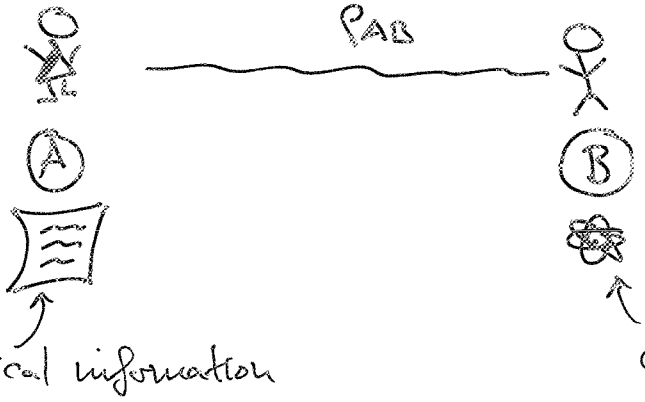
Conditional min-entropy:

$$H_{min}(A|B) := \max_{\sigma_B} \max_{\lambda \in \mathbb{R}} \left\{ 2^{-\lambda} (\mathbb{I} \otimes \sigma_B) \geq \rho_{AB} \right\}$$



Operational interpretations of the conditional min-entropy

Case 1: A is classical



Concept: c-q state:

probability distribution

classical symbol

quantum "encoding" of a

$$\rho_{AB} = \sum_{a \in \mathcal{A}} p_a \underbrace{|a\rangle\langle a|}_{A} \otimes \underbrace{\rho_a}_{B}$$

Such a state corresponds to the following scenario:

1. Alice picks a classical symbol $a \in \mathcal{A}$ with probability p_a
2. Alice creates the encoding ρ_a
3. Alice sends ρ_a to Bob

(Remember the transmission of classical things we discussed earlier!)

Guessing probability

probability that a was chosen

$$P_{guess}(A|B) = \max_{\{M_a\}_a} \sum_{a \in \mathcal{A}} p_a \, \mathrm{Tr}(M_a \rho_a^{(a)})$$

encoding of a

Maximization over POVMs

probability of correctly outputting a if the encoded symbol was really a.

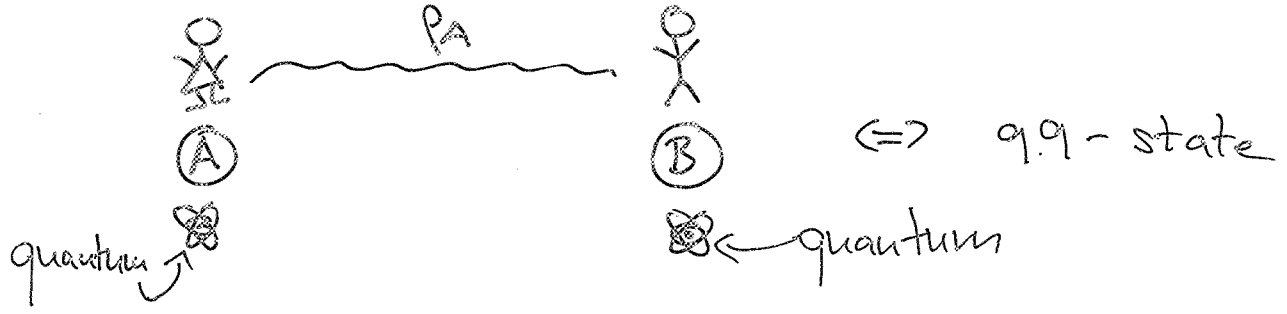
Min-entropy if A is classical:

Note: always positive!

$$H_{min}(A|B) = -\log P_{guess}(A|B)$$

➡ small H_{min} ⇐ easy to guess
large H_{min} ⇐ hard to guess

Case 2: A is quantum



⇐ 9.9 - state

$$H_{min}(A|B) = -\log d_A \max F(\Phi_{AA'}, \mathcal{I}_A \otimes \Lambda(\rho_{AB}))$$

$\Lambda : B(\mathcal{H}_B) \to B(\mathcal{H}_{A'})$

maximally entangled state

➡ small H_{min} ⇐ lots of entanglement,
large H_{min} ⇐ little entanglement

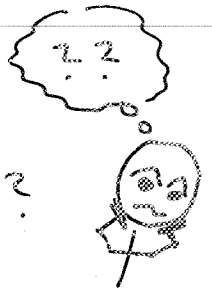
..... so conditional min-entropy was not so bad..

When is the conditional min-entropy used?

- one-shot information theory
- quantum cryptography

$H_{min}(X|E)$ quantifies how many perfectly random bits of key (unknown to Eve) can be extracted from X when an eavesdropper holds information E about X

Let's consider some more special cases:



$\rho_{AB} = \rho_A \otimes \rho_B$ What do we expect?

$$H_{min}(A|B)_\rho = \max_{\sigma_B, \lambda} \lambda : 2^{-\lambda}(\mathbb{I} \otimes \sigma_B) \geq \rho_A \otimes \rho_B$$

→ only depends on $2^{-\lambda}\mathbb{I} \geq \rho_A$

$$\Rightarrow H_{min}(A|B) = \overbrace{-\log \lambda_{max}(\rho_A)}^{2^{-\lambda}\mathbb{I} \geq \rho_A} = H_{min}(A)$$

largest eigenvalue ρ_A

$\Rightarrow$ SANE! 

Conditional max-entropy

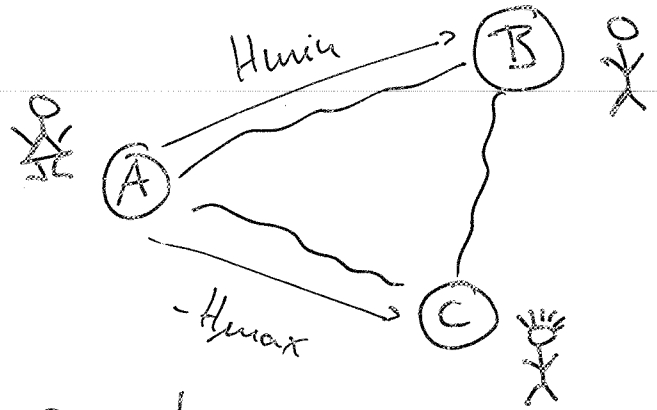
Max-entropy

$$H_{max}(A|B) = \log d_A \max_{\sigma_B} F\left(\rho_{AB}, \frac{\mathbb{I}_A}{d_A} \otimes \sigma_B\right)^2$$

intuition: distance from a state that is maximally mixed on A and uncorrelated from B

→ how random is A given access to B

Duality of entropies:



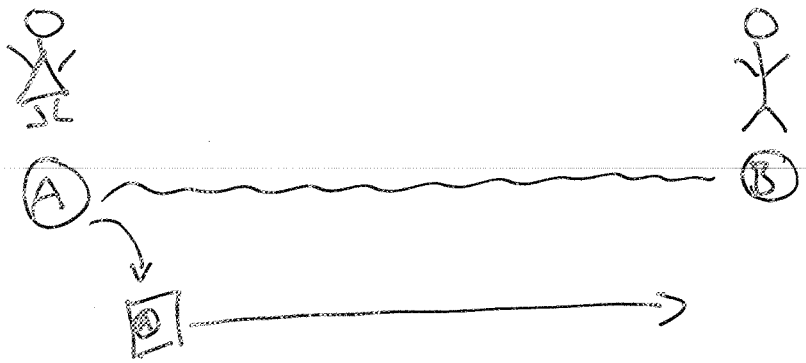
Purify ρ_{AB} to $\rho_{ABC} = |\psi_{ABC}\rangle\langle\psi_{ABC}|$

$$H_{min}(A|B) = -H_{max}(A|C)$$

fully entangled ←Hmin→ ρ_{AB} ←Hmax→ $\frac{\mathbb{I}_A}{d_A} \otimes \rho_B$

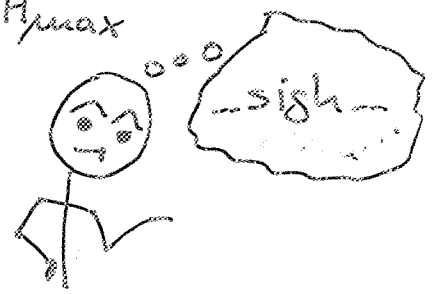
What H_{max} is good for:

- recall: cryptographic interpretation
- one-shot information theory: quantifies how many qubits Alice needs to send to Bob to transfer A, if Bob already has B.



• Smoothing (what the ... is that?)

Maybe you've seen H_{min}^{ε} H_{max}^{ε}



Learning smoothing the smooth way:

-- via a small detour.

Imagine two states that are very similar

Classical example:

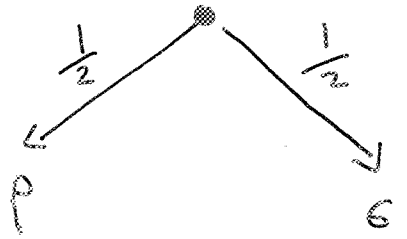
$\rho = |0\rangle\langle 0|$ $\quad \sigma = (1-\varepsilon)|0\rangle\langle 0| + \varepsilon |1\rangle\langle 1|$

Measure of "similarity": trace distance

$$D(\rho,\sigma) = \frac{1}{2}\|\rho-\sigma\|_1 = \max_{0\le P\le \mathbb{I}} \mathrm{Tr}\left(P(\rho-\sigma)\right)$$

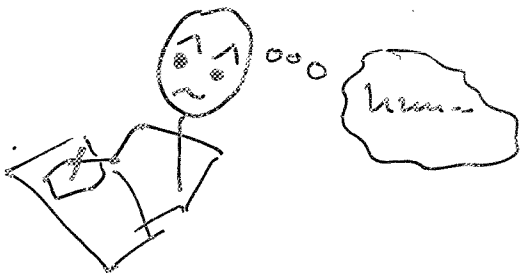
in example $D(\rho,\sigma) = \frac{1}{2}\left(|1-(1-\varepsilon)| + |\varepsilon|\right) = \varepsilon$

in general: $\|A\|_1 = \mathrm{Tr}\sqrt{A^\dagger A} = \max_{-\mathbb{I}\le P\le \mathbb{I}} \mathrm{Tr}(PA)$



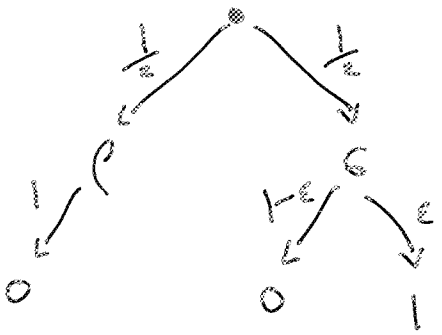
What's the probability of telling them apart?

Let's consider our classical example.
Can you think of a strategy?



If we see "0" → guess ρ
If ... "1" → guess σ

What's our probability of success?



$$p_{succ} = \frac{1}{2}\cdot 1 + \frac{1}{2}\cdot\varepsilon = \frac{1}{2} + \frac{\varepsilon}{2} = \frac{1}{2} + \frac{D(\rho\|\sigma)}{2}$$

(we are given ρ; we see a 1; we are given σ)

Coincidence??

Helstrom: $\boxed{p_{succ} = \frac{1}{2} + \frac{D(\rho\|\sigma)}{2} \quad \text{always!}}$

ρ and σ are ε-close $\Leftrightarrow$ $p_{succ} \leq \frac{1}{2} + \frac{\varepsilon}{2}$
$(\Leftrightarrow D(\rho|\sigma) \leq \varepsilon)$

OR:

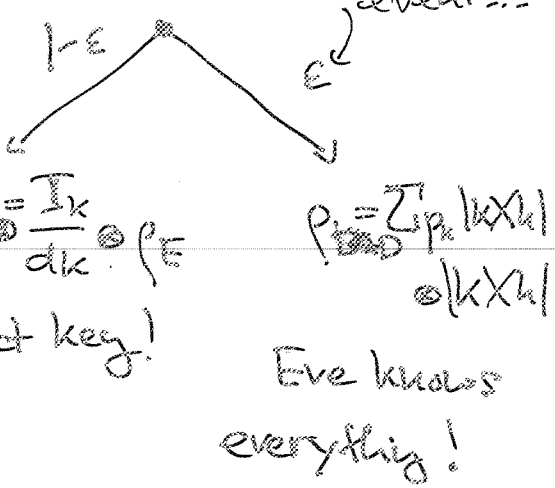
States that are ε-close cannot easily be distinguished using ANY method

Question: How does the entropy of σ relate to ρ when $\sigma \approx_\varepsilon \rho$?

Another example:

Alice — Key K

Eve — E



Density operator $\sigma_{KE} = (1-\varepsilon)\rho_{GOOD} + \varepsilon \rho_{BAD}$

Very close to an ideal key:

$$D\left(\sigma_{KE} : \frac{I_K}{d_K} \otimes \rho_E\right) = \frac{1}{2} \left\| (1-\varepsilon)\rho_{GOOD} - \rho_{GOOD} + \varepsilon \rho_{BAD} \right\|_1$$

$$= \frac{1}{2} \left\| \varepsilon \rho_{BAD} - \varepsilon \rho_{GOOD} \right\|_1 = \varepsilon \left(\frac{1}{2} \left\| \rho_{BAD} - \rho_{GOOD} \right\|_1 \right)$$

≤ 1

$\leq \varepsilon \Rightarrow$ very close!

Idea of smoothing: neglect events that occur with small probability

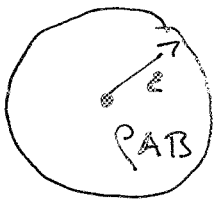
ε-smooth min-entropy

$$H_{min}^{\varepsilon}(A|B)_{\rho} = \max_{\sigma_{AB} \in B^{\varepsilon}(\rho_{AB})} H_{min}(A|B)_{\sigma}$$

ρ: original state

σ: state in ε-Ball

$B^{\varepsilon}(\rho_{AB})$: ε-Ball around ρ_{AB}



Consider our example:

$$H_{min}(K|E)_{\rho_{GOOD}} = H_{min}(K) = \log d_K$$

$$\rho_{GOOD} = \frac{\mathbb{I}_K}{d_K} \otimes \rho_E$$

$\Rightarrow$ maximal entropy.

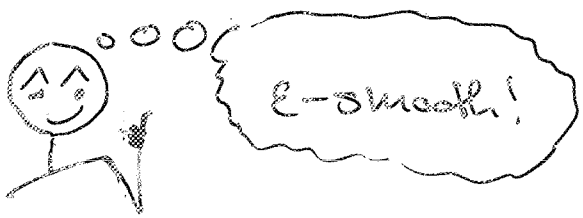
How about $H_{min}(K|E)_{\sigma}$? $\left(\sigma_{KE} = (1-\varepsilon)\rho_{GOOD} + \varepsilon \rho_{BAD}\right)$

$$H_{min}^{\varepsilon}(K|E)_{\sigma} \geq H_{min}(K|E)_{\rho_{GOOD}} = \log d_K$$

distance ε!

$$H_{min}^{\varepsilon}(K|E)_{\sigma} \geq \log d_K$$

Intuitive interpretation: σ_{KE} has min-entropy $H_{min}(K|E) \geq \log d_K$, except with an error probability ε



ε-smooth max-entropy

$$H_{max}^{\varepsilon}(K|E)_{\sigma} = \min_{\rho \in \mathcal{B}^{\varepsilon}(\sigma)} H_{max}(K|E)$$

Why min instead of max??

Remember: $H_{max}(K|E) = -H_{min}(K|P)$

purifying system

When are these entropies used?

- one-shot information theory
- crypto

Interpretations:

$H_{min}^{\varepsilon}(A|B)$ → degree of decoupling possible between A & B (see later)

$H_{max}^{\varepsilon}(A|B)$ → state merging

Caution: several distance measures used in smoothing

- trace distance
- fidelity
- purified distance

May need to convert

Properties — very many .. scattered ..

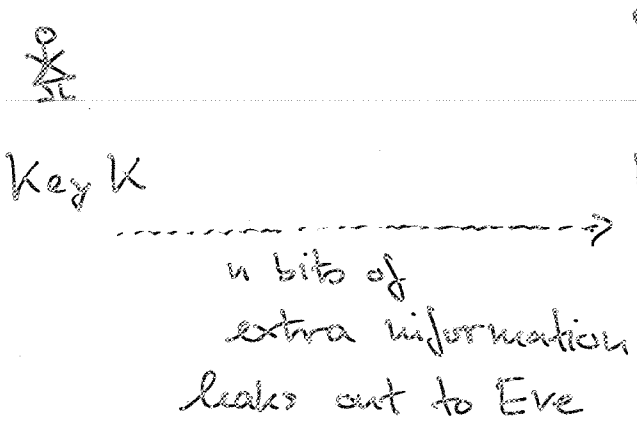
Focus on H_{min}:

- $H_{min}(AB|C) \geq H_{min}(A|C)$ (Monotonicity)
- $H_{min}(A) \geq H_{min}(A|B)$

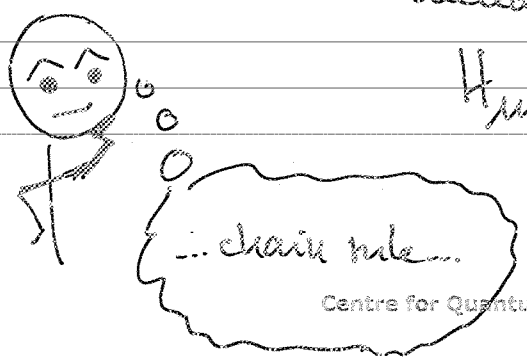
Chain rule:

$$H^{\varepsilon}_{min}(A|BC) \geq H^{\varepsilon}_{min}(AB|C) - H_0(B)$$

Example use:



Question: How much can n extra bits decrease the min-entropy $H_{min}(K|E)$?



$$H_{min}(K|EX) \geq H_{min}(KX|E) - H_0(X)$$

extra bits; chain rule

$$\text{monotonicity} \Rightarrow \geq H_{min}(K|E) - H_0(X)$$

$$H_0(X) \leq \log d_X \Rightarrow \geq H_{min}(K|E) - \log d_X$$

$$= H_{min}(K|E) - n$$

n: size of d_X is n bits

$\Rightarrow$ Leaking n extra bits can reduce the min-entropy by at most n.

- Relations between entropies

H_{min} vs. H

H_{max} vs. H

"iid limit" $\rho_{AB}^{\otimes n}$ $n \to \infty$

Asymptotic equipartition theorem (AEP):

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} H_{min}^{\varepsilon}(A|B) = H(A|B)$$

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} H_{max}^{\varepsilon}(A|B) = H(A|B)$$

$\Rightarrow$ coincide in the iid case!

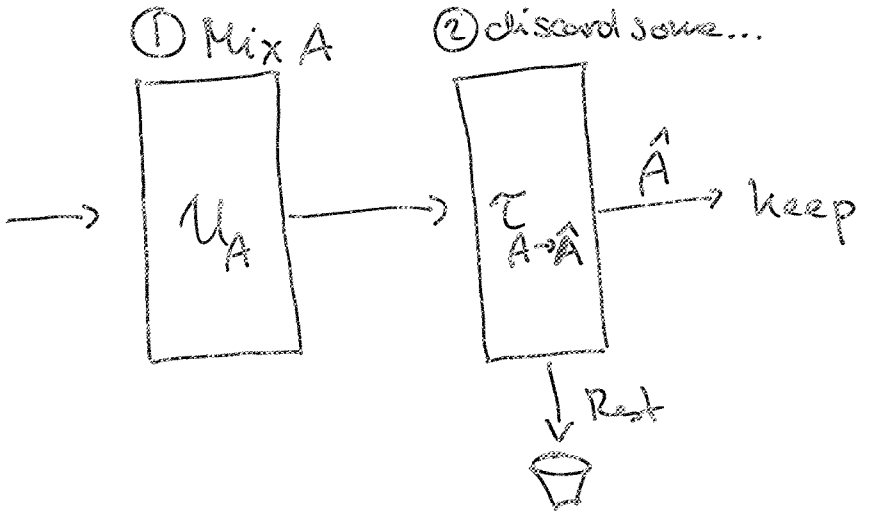
(3) The decoupling theorem (... if you only learn one theorem ...)

Idea:

Initially, A and R are correlated.

Question: can Alice extract some part of her system $\hat{A}$ that is uncorrelated from R? $\rho_{AR} \stackrel{?}{=} \rho_{\hat{A}} \otimes \rho_R$?

Idea: two steps



What $\mathcal{U}_A$ and maps $\tau_{A\to\hat{A}}$ could do the trick?

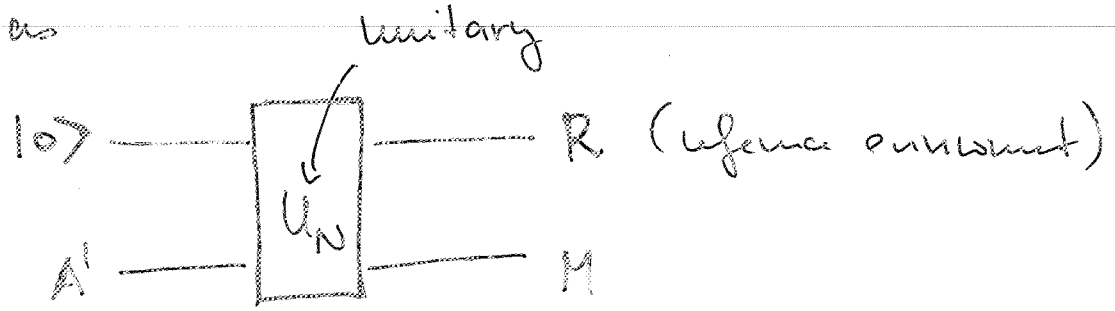
Decoupling first introduced for analysis the transmission of quantum information.

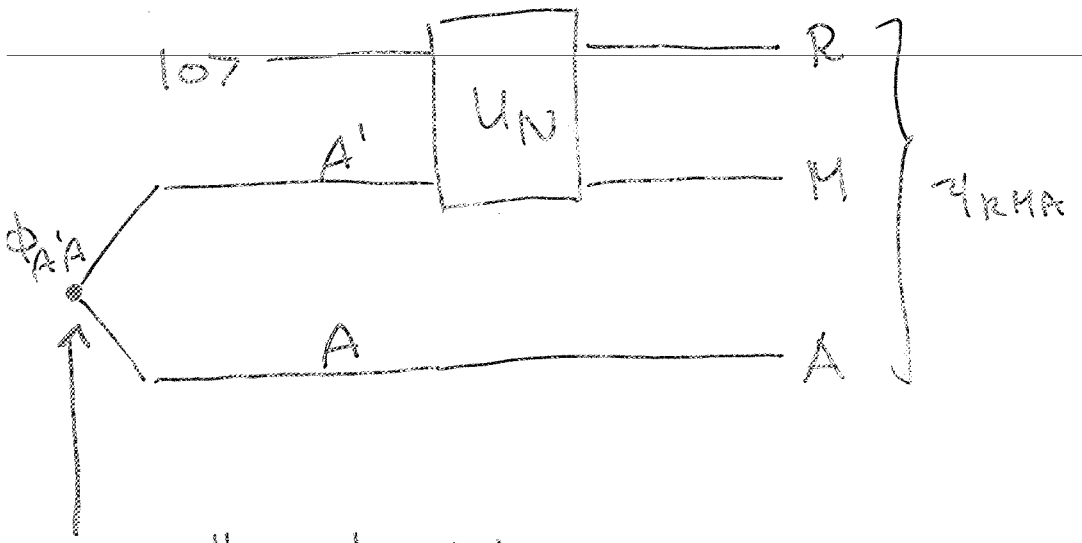
U_A : chosen uniformly at random, Haar measure

$\tau_{A \to \hat{A}}$: partial trace

To get some intuition on how it's useful in this area, let's first consider this history and its applications:

Any noisy channel $N_{A' \to M}$ can be expressed as





Question: Can we find an encoding scheme for A' and decoding scheme on M to recover the entanglement?

Consider ψ_{RMA}. This state is pure.

Let's say we could decouple A from R by some procedure P_A (assume for simplicity unitary) / maximally mixed

ie $(I_{RM} \otimes P_A)\,\psi_{RMA} \longmapsto \tilde{\psi}_{RMA}$

such that $\tilde{\psi}_{AR} \approx_\varepsilon \pi_A \otimes \rho_R$

- $\pi_A \otimes \rho_R$ purified $\Phi_{A'A} \otimes \rho_R$

- compare $\tilde{\psi}_{MAR}$ with $\Phi_{A'A} \otimes \rho_R$

Uhlmann's theorem implies $\exists\, D_{M \to A'J}$ (J ← junk)

$$D_{M \to A'J}(\tilde{\psi}_{MAR}) = \Phi_{A'A} \otimes \rho_{RJ}$$

$$(\mathbb{I}_{A'} \otimes P_A)\,|\phi_{A'A}\rangle\langle\phi_{A'A}| = (P_{A'}^T \otimes \mathbb{I}_A)\,|\phi_{AA}\rangle\langle\phi_{A'A}|$$

→ can think of the decoupling procedure as an encoding procedure on A'

→ once we are decoupled, Uhlmann's theorem gives decoding procedure

The first decoupling theorems were limited to the partial trace and somewhat crude in that the relation between the strength of decoupling and the information that R holds about A was not explicit.

"The" decoupling theorem

Formally:

maps A to $\hat{A}$ identity on R

Haar measure

$$\int_{\mathcal{U}(A)} \left\| \tau_{A\to\hat{A}}\left(\mathcal{U}_A \rho_{AR} \mathcal{U}_A^\dagger\right) - \tau_{\hat{A}} \otimes \rho_R \right\|_1 d\mathcal{U}$$

$$\leq 2^{-\frac{1}{2} H_{\min}^{\varepsilon}(A|R)_\rho - \frac{1}{2} H_{\min}^{\varepsilon}(A'|\hat{A})_\tau} + 12\varepsilon$$

original state ρ_{AR}

Caution:

$$\tau_{A'\hat{A}} = \pi_{A'} \otimes \tau_{A\to\hat{A}}\left(|\phi\rangle\langle\phi|_{A'A}\right)$$

maximally entangled

Caution: As stated here we smooth in terms of the purified distance

$$P(\rho,\sigma) = \sqrt{1 - F(\rho,\sigma)^2}$$

Very general... but can't tell what it means...

Example: Partial trace

m qubits total

$A = \widehat{A_1 A_2}$ (m' qubits)

$\tau_{A \to \hat{A}}(\rho_A) = \mathrm{Tr}_{m-m'}(\rho_{A=A_1A_2})$

$A_1 = \hat{A}$

Step 1: Compute $\tau_{A'\hat{A}}$

Note $|\Phi_{A'A}\rangle\langle\Phi_{A'A}| = |\Phi_{EPR}\rangle\langle\Phi_{EPR}|^{\otimes m}$

A'_{q_1} A_{q_1}

A'_{q_2} A_{q_2}

A'_{q_m} A_{q_m}

$$\tau_{A'\hat{A}} = (\mathbb{I}_{A'} \otimes \mathrm{Tr}_{m-m'})|\Phi_{EPR}\rangle\langle\Phi_{EPR}|^{\otimes m}$$

$$= \underbrace{|\Phi_{EPR}\rangle\langle\Phi_{EPR}|^{\otimes m'}}_{\text{first } m' \text{ qubits of } A' \text{ and } A : F} \otimes \frac{\mathbb{I}_{A'_{q_{rest}}}}{2^{m-m'}}$$

Step 2: Compute $H_{min}^{\varepsilon}(A'|\hat{A}) = \underbrace{-m'}_{H_{min}(F|\hat{A})} + \underbrace{(m-m')}_{H_{min}\left(\frac{\mathbb{I}_{A'_{rest}}}{2^{m-m'}}\right)}$

$= m - 2m'$

Step 3: Compute $\tau_{\hat{A}}$

$$\tau_{\hat{A}} = Tr_{A'}(\tau_{A'\hat{A}}) = Tr_{A'}\left(\Phi_{EPR}^{\otimes m'} \otimes \frac{\mathbb{I}_{A'^{km}}}{d_{A'^{km}}}\right)$$

$$= \frac{\mathbb{I}_{\hat{A}}}{d_{\hat{A}}}$$

Useful trick:

In general, $\tau_{\hat{A}} = Tr_{A'}(\tau_{A'\hat{A}}) =$

order doesn't matter

$$= Tr_{A'}\left(\mathbb{I}_{A'} \otimes \tau_{A\to\hat{A}}(|\Phi_{A'A}\rangle\langle\Phi_{A'A}|)\right)$$

$$= \tau_{A\to\hat{A}}\left(Tr_{A'}|\Phi_{A'A}\rangle\langle\Phi_{A'A}|\right)$$

$\Phi_{A'A}$ max. entangled

$$= \tau_{A\to\hat{A}}\left(\frac{\mathbb{I}_A}{d_A}\right)$$

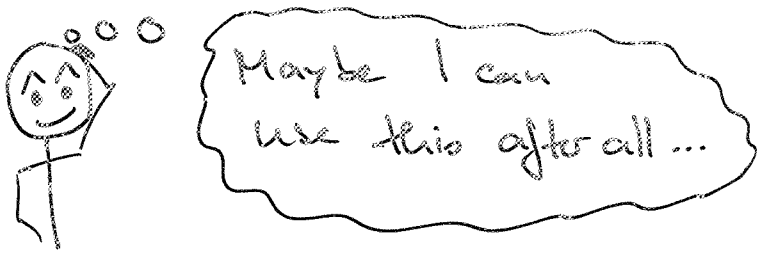
in our example we get:

$$\int_{U(A)} \| \mathrm{Tr}_{A_2}(U_A \rho_{AR} U_A^\dagger) - \frac{\mathbb{I}_{\hat{A}}}{d_{\hat{A}}} \otimes \rho_R \|_1 \, dU$$

$$\leq 2^{-\frac{1}{2} H_{\min}^{\varepsilon}(A|R)_\rho - \frac{1}{2}(n-2m')} + 12\varepsilon =: \varepsilon_D$$

the less entanglement, the better the approximation

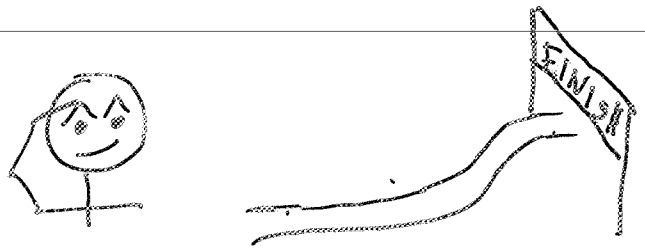
the more we trace out, the better the approximation



Probabilistic statement:

$$\Pr_{U_A}\left[\| \tau_{A \to \hat{A}}(U_A \rho_{AR} U_A^\dagger) - \tau_{\hat{A}} \otimes \rho_R \|_1 > \varepsilon_D + \delta \right]$$

$$\leq 2 e^{-d_A \frac{\delta^2}{16}}$$



"How optimal" is the decoupling theorem?

Converse: For any $\varepsilon', \varepsilon'', \varepsilon''' > 0$

if $H_{min}^{\varepsilon' + 2\varepsilon'' + \varepsilon''' + \sqrt{\varepsilon}}(A|R)_\rho + H_{max}^{\varepsilon'''}(A'\hat{A})_\tau - H_{min}^{\varepsilon'''}(\hat{A})_\tau$

$$< -\log \frac{2}{\varepsilon'^2}$$

Then there exists no state ω_B at all such that

$$\int_{U(A)} \| \tau_{A \to \hat{A}}(\mathcal{U}_A(\rho_{AR})\mathcal{U}_A^\dagger) - \omega_B \otimes \rho_R \|_1 \leq \frac{\varepsilon}{2}$$

Converse $\quad \Longleftrightarrow \quad$ Decoupling

$\ldots + H_{max}(A'\hat{A})_\tau - H_{min}(\hat{A})_\tau$ $\qquad$ $H_{min}(A'|\hat{A})_\tau$

$\geq H_{min}(A'\hat{A})_\tau - H_0(\hat{A})_\tau$