CO781 / QIC 890:

Theory of Quantum Communication

Topic 2, part 1

The asymptotic equipartition theorem,
Shannon entropy and classical data compression

von Neumann entropy, Quantum data compression,
entanglement concentration and dilution

References:

Nielsen and Chuang Section 12.2
Preskill Sections 10.1.1, 10.3, 10.4
Cover & Thomas

From reading material:

Def: Let ρ be a density matrix with spectral decomposition

$$\rho = \sum_{v=1}^{d} p(v) \, |e_v\rangle\langle e_v| \, .$$

Let V be a rv with sample space $\{1, 2, \ldots, d\}$

and distribution $p(v)$.

Let $T_{n,\delta}$ be the typical set for n iid draws of V.

For $v^n = v_1 v_2 \ldots v_n \in T_{n,\delta}$, let

$$|e_{v^n}\rangle = |e_{v_1}\rangle |e_{v_2}\rangle \ldots |e_{v_n}\rangle \, .$$

The δ-typical space of $\rho^{\otimes n}$ $= \text{span}\{ |e_{v^n}\rangle : v^n \in T_{n,\delta} \} =: S$.

Let $\boxed{\Pi_S = \sum_{v^n \in T_{n,\delta}} |e_{v^n}\rangle\langle e_{v^n}|}$ (projector onto S).

Def: H(V) = von Neumann entropy of ρ =: S(ρ)

from Thur class

by def of vN entropy

① $\dim S = |T_{n,\delta}| \leq 2^{n(H(V)+\delta)} = 2^{n(S(\rho)+\delta)}$

② $\mathrm{Tr}(\rho^{\otimes n} \Pi_S) = \sum_{v^n \in T_{n,\delta}} p(v^n) \geq 1-\varepsilon$ if $n \geq n_0 = \cdots$

$\sum_{\text{all } v'^n} |e_{v'^n}\rangle\langle e_{v'^n}| \, p(v'^n)$ $\qquad \sum_{v^n \in T_{n,\delta}} |e_{v^n}\rangle\langle e_{v^n}|$

$\rho^{\otimes n}$ "mostly" contained in its typical space

The "transmit the typical space" (TTS) protocol:

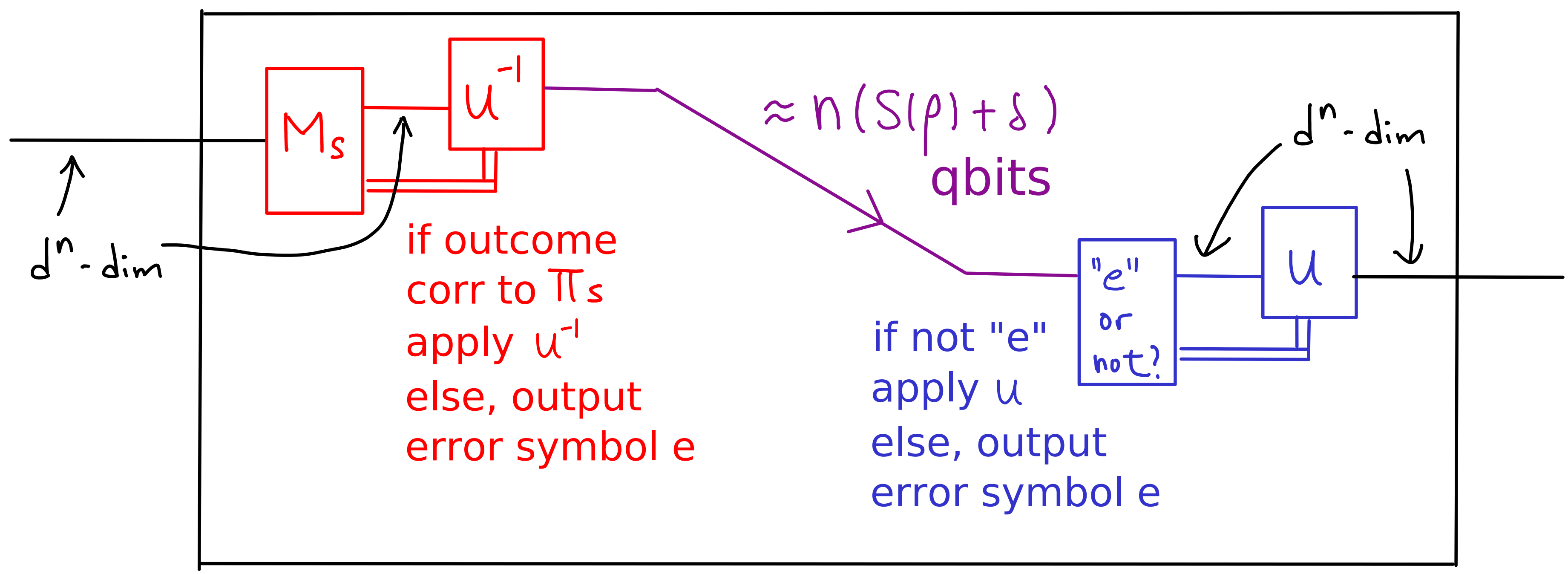
Fix ρ (dxd), for arbitrary $\delta > 0$, $\varepsilon > 0$, $n \geq n_0$,

with n_0, $T_{n,\delta}$, S, Π_S as defined before,

let M_S denote the binary meas with POVM $\{\Pi_S, I - \Pi_S\}$.

Define isometric bijections : $\mathbb{C}^{2^{n(S(\rho)+\delta)}} \underset{U^{-1}}{\overset{U}{\rightleftarrows}} S$.

The "TTS" protocol $\mathcal{T}$:



In particular, for $\rho^{\otimes n} = \sum_{v^n} p(v^n) |e_{v^n}\rangle\langle e_{v^n}|$

$$T(\rho^{\otimes n}) = \sum_{v^n \in T_{n,\delta}} p(v^n) |e_{v^n}\rangle\langle e_{v^n}| + \underbrace{(1 - p(T_{n,\delta}))}_{\leq \varepsilon} |e\rangle\langle e|$$

$$\therefore \|\rho^{\otimes n} - T(\rho^{\otimes n})\|_t \leq \varepsilon ,$$

Note that both Alice and Bob need to know ρ in TTS.

The task to send $\rho^{\otimes n}$ using TTS is NOT interesting, since Bob should simply create the state without Alice's help.

But TTS turns out very useful ... for much harder tasks !

Quantum source, vN entropy, & quantum data compression

X : random variable, Ω : sample space, Pr(X=x) = q(x).

For each x, ρ_x : quantum state (in d-dim) labeled by x.

Consider the process:

1. sample X, obtain $x \in \Omega$ wp $q(x)$
2. prepare quantum state ρ_x

Resulting state: $\Lambda = \sum_x q(x) \underbrace{|x\rangle\langle x|_R}_{\text{classical random outcome}} \otimes \underbrace{\rho_{x\,A}}_{\text{quantum random outcome}}$

Terminology:

- receiving a specimen ρ_x in system A wp q(x) is called "one draw" of the ensemble $\mathcal{E} = \{q(x), \rho_x\}$
- average state of $\mathcal{E}$ is $\rho = \sum_x q(x)\, \rho_x = \mathrm{tr}_R \Lambda$.

evals $\neq q(x)$ in general

e.g., B92 $q(0) = q(1) = \frac{1}{2}$

$$\rho_0 = |0\rangle\langle 0|, \quad \rho_1 = |+\rangle\langle +|, \quad |+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$$

$$\Lambda = \frac{1}{2}|0\rangle\langle 0|_R \otimes |0\rangle\langle 0|_A + \frac{1}{2}|1\rangle\langle 1|_R \otimes |+\rangle\langle +|_A, \quad \rho = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

e.g., BB84 $q(0) = q(1) = q(2) = q(3) = \frac{1}{4}$

$$\rho_0 = |0\rangle\langle 0|, \quad \rho_1 = |+\rangle\langle +|,$$
$$\rho_2 = |1\rangle\langle 1|, \quad \rho_3 = |-\rangle\langle -|. \qquad |\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$$

$$\Lambda = \frac{1}{4}|0\rangle\langle 0|_R \otimes |0\rangle\langle 0|_A + \frac{1}{4}|1\rangle\langle 1|_R \otimes |+\rangle\langle +|_A$$
$$+ \frac{1}{4}|0\rangle\langle 0|_R \otimes |1\rangle\langle 1|_A + \frac{1}{4}|1\rangle\langle 1|_R \otimes |-\rangle\langle -|_A, \quad \rho = \frac{I}{2}$$

In both examples, Alice prepares both RA & transmits A to Bob. Eve sees A (a draw from the ensemble).

An iid quantum source: repeating the above process ...

Repeating n times means:

(1) Sample X n times iid

Obtain $x^n = x_1 x_2 \cdots x_n$ with prob $q(x^n) = q(x_1)\, q(x_2) \cdots q(x_n)$

(2) Prepare $\rho_{x^n} = \rho_{x_1} \otimes \rho_{x_2} \cdots \otimes \rho_{x_n}$

Resulting state: $\sum_{x^n} q(x^n) \underbrace{|x^n\rangle\langle x^n|}_{\text{in } R^n = R_1 R_2 \cdots R_n} \otimes \underbrace{\rho_{x^n}}_{\text{in } A^n = A_1 A_2 \cdots A_n} = \Lambda^{\otimes n}$

Quantum data compression:
Transmit A1 A2 ... An using nr qbits, and minimize r.

Difference from "qbit" : there is restriction to what states might have to be sent. There are many scenarios ...

Blind compression: Sender Alice doesn't know what's being compressed.

1. Referee Richard prepares $\Lambda^{\otimes n}$ & gives A1A2...An to Alice.

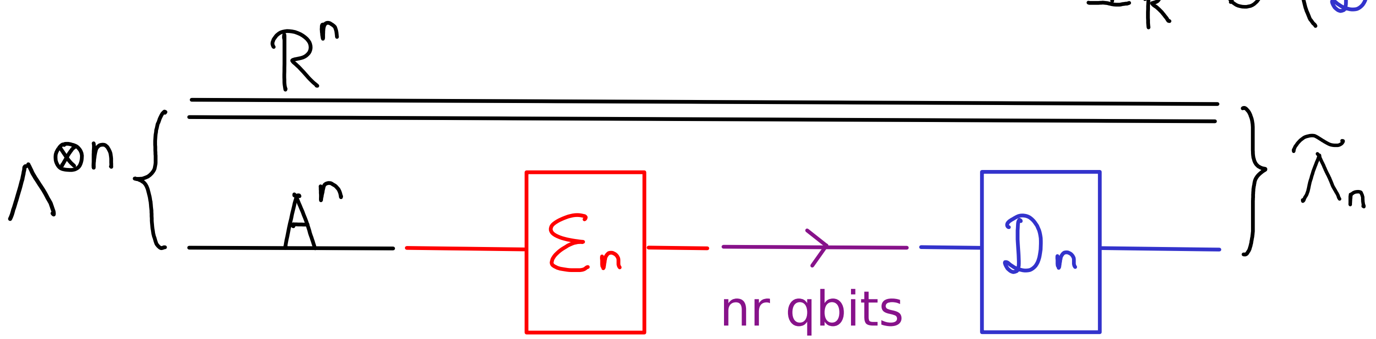
$$\Lambda^{\otimes n} = \sum_{x^n} q(x^n)\, |x^n\rangle\langle x^n|_{R_1R_2\cdots R_n} \otimes \rho_{x^n\, A_1A_2\cdots A_n}$$

i.e., draws x^n wp $q(x^n)$, records on R1R2...Rn, prepares ρ_{x^n} on A1A2...An

2. Alice encodes A1...An in nr qubits, transmits them to Bob

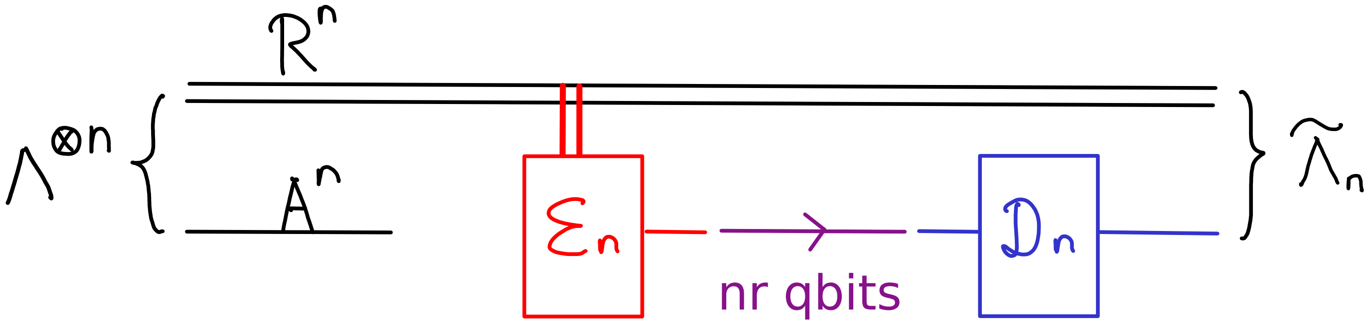
3. Bob decodes to an output, which on average over x^n should be close to ρ_{x^n}

$$I_{R^n} \otimes (\mathcal{D}_n \circ \mathcal{E}_n)_{A^n}(\Lambda^{\otimes n})$$



Correctness: $\| \Lambda^{\otimes n} - \tilde{\Lambda}_n \|_1$ small.

Visible compression: Sender Alice is also referee Richard



Correctness: $\| \Lambda^{\otimes n} - \tilde{\Lambda}_n \|_1$ small.

Remarks on the above correctness condition:

-Preserves correlation between A1A2 ... An and R1R2 ... Rn.

-Simulation of "noiseless" comm of A1A2...An only on $\Lambda^{\otimes n}$.

-Error is global: on the entire ρ_{x^n} but weighted by $q(x^n)$.

Definition: r is called an achievable rate, if for all n large enough, a protocol above exists with error vanishing with n.

Schumacher compression:

pure state ensemble

Theorem: Let $\Lambda = \sum_x q(x)\, |x\rangle\langle x|_R \otimes |\psi_x\rangle\langle\psi_x|_A$

$$\rho = \sum_x q(x)\, |\psi_x\rangle\langle\psi_x|$$

Consider blind Q data compression task.

$\forall \epsilon > 0 \quad \forall r > S(\rho)$

$\exists n_0$ s.t. $\forall n \geq n_0 \quad \exists \mathcal{E}_n, \mathcal{D}_n$

s.t. output dim of $\mathcal{E}_n \leq 2^{nr}$

$$\| \Lambda^{\otimes n} - \mathcal{I}_{R^n} \otimes (\mathcal{D}_n \circ \mathcal{E}_n)_{A^n} (\Lambda^{\otimes n}) \|_1 \leq 2\sqrt{2\epsilon} + \epsilon$$

vanishing w/ ϵ

indep of dim, n

PS Alice doesn't know x^n cannot send it.

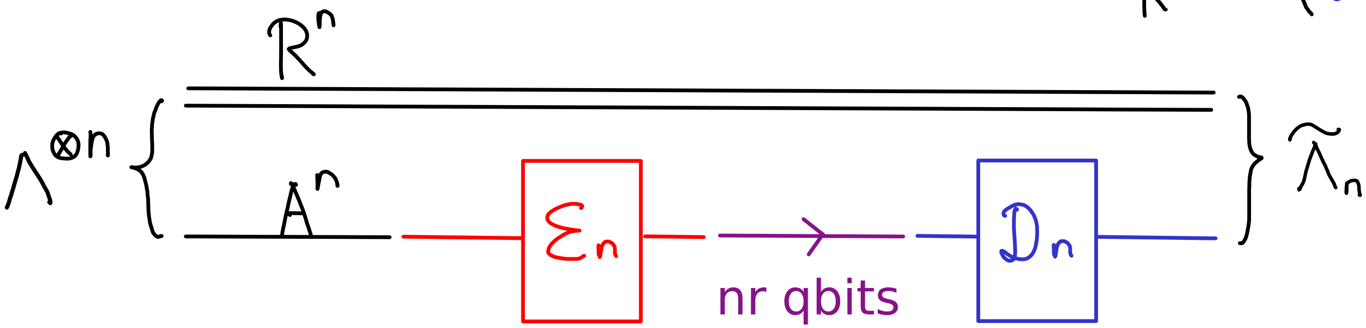
Sending x^n via classical data compression works in the visible setting, but can be suboptimal since

$$r \doteq H(X) \geq S(\rho).$$

Proof: will show that TTS works !

but remember the task is NOT sending $\rho^{\otimes n}$

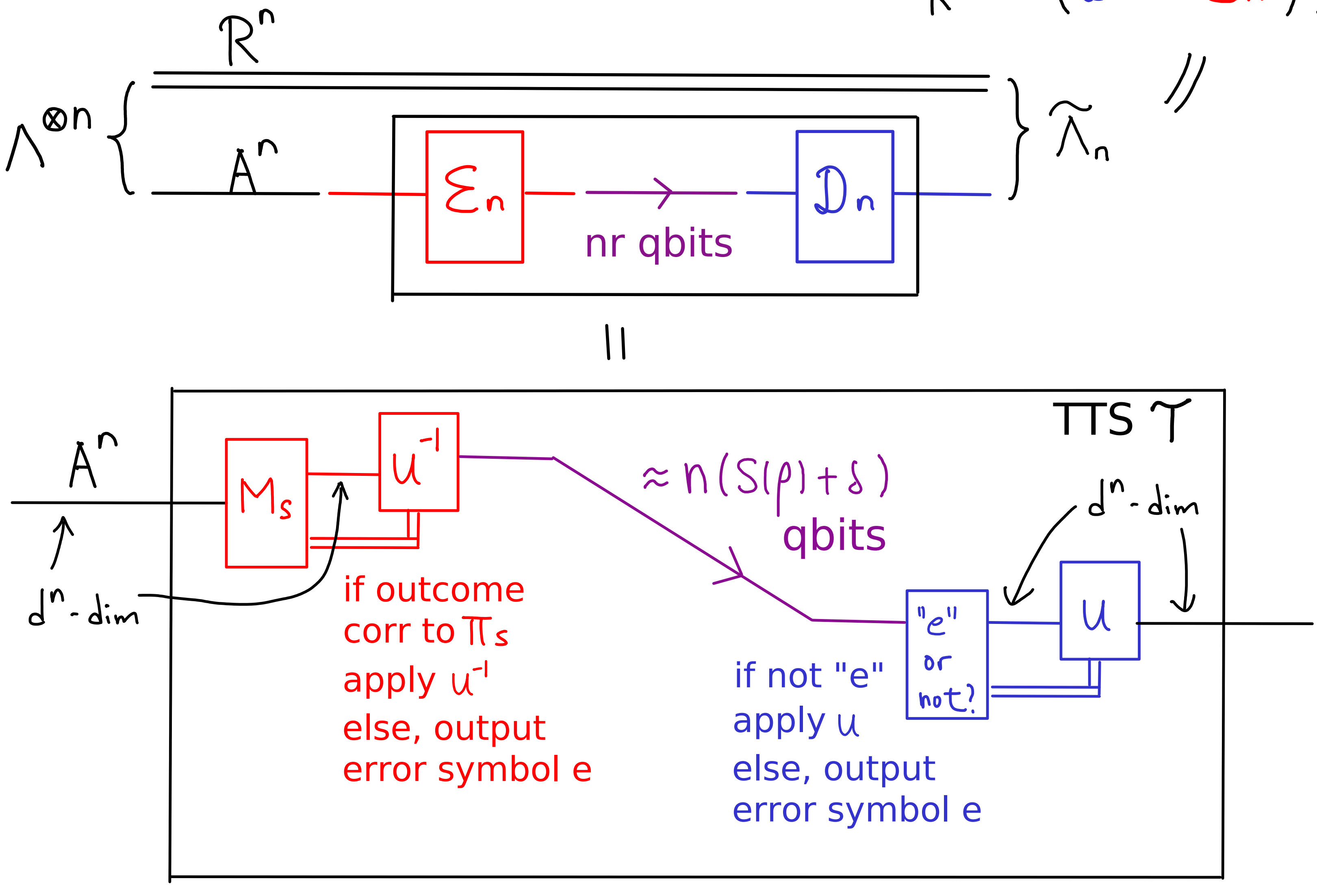
$$I_{R^n} \otimes (\mathcal{D}_n \circ \mathcal{E}_n)_{A^n} (\Lambda^{\otimes n}) =$$



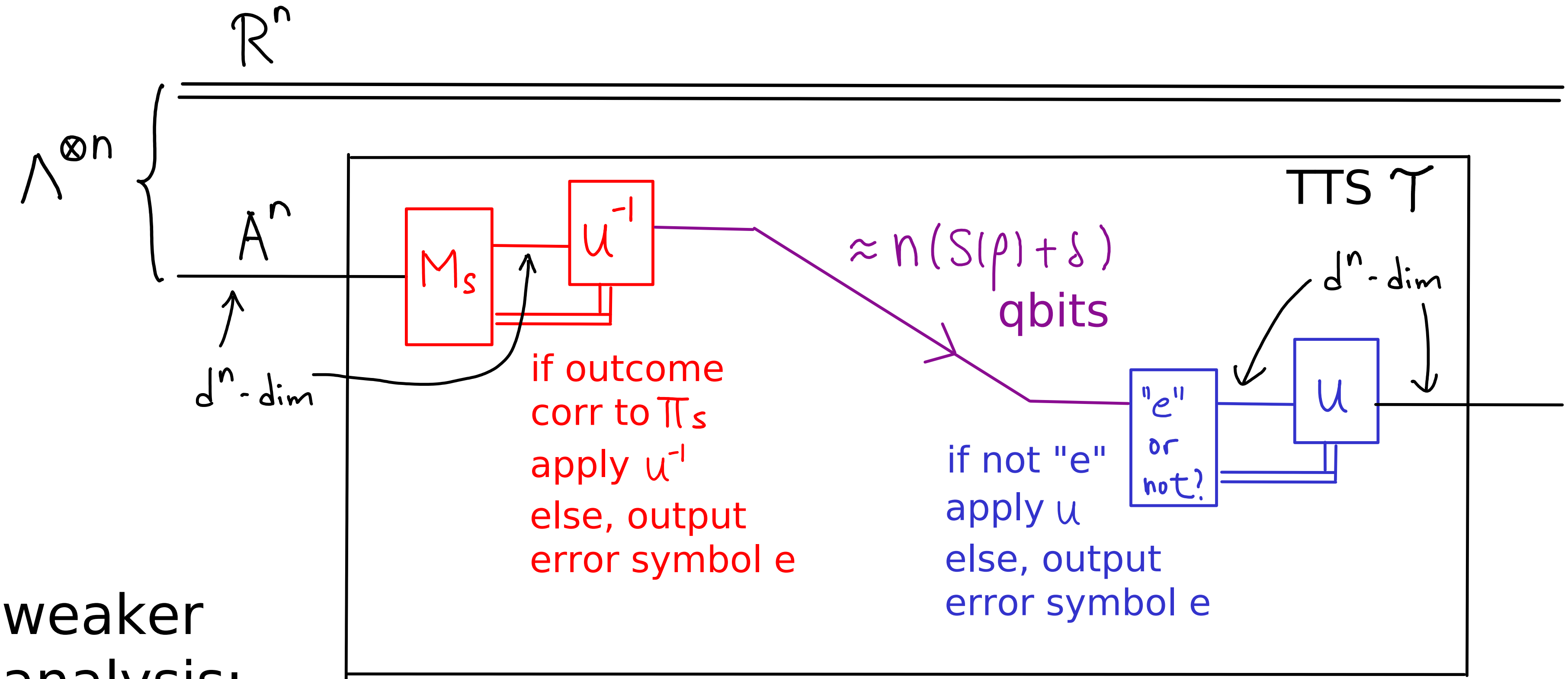
Proof: will show that TTS works !

but remember the task is NOT sending $\rho^{\otimes n}$

$$I_{R^n} \otimes \left(\mathcal{D}_n \circ \mathcal{E}_n \right)_{A^n} \left(\Lambda^{\otimes n} \right)$$



Proof: will show that TTS works !



weaker analysis:

$|\psi_{x^n}\rangle$ is transformed to $\Pi_s|\psi_{x^n}\rangle\langle\psi_{x^n}|\Pi_s + \mathrm{tr}(I-\Pi_s)\cdot|e\rangle\langle e|$

Fidelity $= \left[\langle\psi_{x^n}|\left(\Pi_s|\psi_{x^n}\rangle\langle\psi_{x^n}|\Pi_s + \mathrm{tr}(I-\Pi_s)\cdot|e\rangle\langle e|\right)|\psi_{x^n}\rangle\right]^{\frac{1}{2}}$

$= \langle\psi_{x^n}|\Pi_s|\psi_{x^n}\rangle$

(note unlike classical case, projection onto typical space can distort the state)

Average fidelity: $\displaystyle\sum_{x^n} q(x^n) \langle \psi_{x^n} | \Pi_S | \psi_{x^n} \rangle$

all

$$= \sum_{x^n} q(x^n) \,\mathrm{Tr}\left(|\psi_{x^n}\rangle\langle\psi_{x^n}| \, \Pi_S \right)$$

$$= \mathrm{Tr}\left(\sum_{x^n} q(x^n) |\psi_{x^n}\rangle\langle\psi_{x^n}| \, \Pi_S \right)$$

$$= \mathrm{Tr}\left(\rho^{\otimes n} \, \Pi_S \right)$$

$$\geq 1 - \epsilon$$

not the state being transmitted
but pops out in this error measure

Detailed analysis:

$$\widetilde{\Lambda}_n = I_{R^n} \otimes (D_n \circ E_n)_{S^n} (\Lambda^{\otimes n})$$

$$\text{all } x^n \longrightarrow \; = \sum_{x^n} p(x^n)\, |x^n\rangle\langle x^n|_R \otimes \Pi_S |\psi_{x^n}\rangle\langle\psi_{x^n}| \Pi_S$$

$$+ \sum_{x^n} p(x^n)\, |x^n\rangle\langle x^n|_R \otimes \mathrm{ERR}\left[\operatorname{tr}(I-\Pi_S)\, |\psi_{x^n}\rangle\langle\psi_{x^n}|\right]$$

$(*)$ $\operatorname{tr}(\text{2nd term}) \leq \varepsilon$

$$\| \widetilde{\Lambda}_n - \Lambda^{\otimes n} \|_1$$

$$\overset{\Delta\,\text{Ineq}}{\leq} \| \text{1st term of } \widetilde{\Lambda}_n - \Lambda^{\otimes n} \|_1 + \| \text{2nd term of } \widetilde{\Lambda}_n \|_1$$

$$\overset{\Delta I,\, (*)}{\leq} \sum_{x^n} p(x^n)\, \| \Pi_S |\psi_{x^n}\rangle\langle\psi_{x^n}| \Pi_S - |\psi_{x^n}\rangle\langle\psi_{x^n}| \|_1 + \varepsilon$$

ΔI, ⊛

$$\leq \sum_{x^n} q(x^n) \, \| \Pi_S |\psi_{x^n}\rangle\langle\psi_{x^n}| \Pi_S - |\psi_{x^n}\rangle\langle\psi_{x^n}| \|_1 + \varepsilon$$

see notes

$$\leq \sum_{x^n} q(x^n) \, 2\sqrt{1 - \langle\psi_{x^n}|\Pi_S|\psi_{x^n}\rangle^2} + \varepsilon$$

$1 - z^2 = (1-z)(1+z)$

$\leq 2(1-z)$

$$\leq \sum_{x^n} q(x^n) \, 2\sqrt{2}\sqrt{1 - \langle\psi_{x^n}|\Pi_S|\psi_{x^n}\rangle} + \varepsilon$$

$$\leq 2\sqrt{2}\sqrt{\sum_{x^n} q(x^n)\left(1 - \langle\psi_{x^n}|\Pi_S|\psi_{x^n}\rangle\right)} + \varepsilon$$

Concavity of $\sqrt{\ }$

$$\leq 2\sqrt{2}\sqrt{1 - \mathrm{tr}(\rho^{\otimes n}\Pi_S)} + \varepsilon$$

$$\leq 2\sqrt{2}\sqrt{\varepsilon} + \varepsilon.$$

Notes on a minor detail:

Want: $\left\| \Pi_s |\psi_{x^n}\rangle\langle\psi_{x^n}| \Pi_s - |\psi_{x^n}\rangle\langle\psi_{x^n}| \right\|_1 \leqslant 2\sqrt{1 - \langle\psi_{x^n}|\Pi_s|\psi_{x^n}\rangle^2}$

NC 9.2.3, (9.97)-(9.99): for unit vectors $|a\rangle, |b\rangle$

$$\frac{1}{2} \left\| |a\rangle\langle a| - |b\rangle\langle b| \right\|_1 = \sqrt{1 - |\langle a|b\rangle|^2}$$

but $\Pi_s |\psi_{x^n}\rangle$ is not a unit vector ... so we tweak the proof in NC.

Proof: $|\psi_{x^n}\rangle = \Pi_s |\psi_{x^n}\rangle + (I - \Pi_s) |\psi_{x^n}\rangle$ (two ortho terms)

$= \alpha |e_0\rangle + \beta |e_1\rangle$, $\alpha \geqslant 0$, $\beta \geqslant 0$, $\alpha^2 + \beta^2 = 1$,

$|e_0\rangle, |e_1\rangle$ ortho unit vecs

$$|\psi_{x^n}\rangle\langle\psi_{x^n}| - \Pi_s |\psi_{x^n}\rangle\langle\psi_{x^n}| \Pi_s = \begin{bmatrix} \alpha^2 & \alpha\beta \\ \alpha\beta & \beta^2 \end{bmatrix} - \begin{bmatrix} \alpha^2 & 0 \\ 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & \alpha\beta \\ \alpha\beta & \beta^2 \end{bmatrix} = \beta \begin{bmatrix} 0 & \alpha \\ \alpha & \beta \end{bmatrix}$$

$$\begin{bmatrix} 0 & \alpha \\ \alpha & \beta \end{bmatrix} = \alpha \sigma_x + \frac{\beta}{2}(I - \sigma_z)$$

Eigenvalues of $\mu \sigma_x + \nu \sigma_z = \pm\sqrt{\mu^2 + \nu^2}$

Eigenvalues of $\begin{bmatrix} 0 & \alpha \\ \alpha & \beta \end{bmatrix} = \frac{\beta}{2} + \sqrt{\alpha^2 + \frac{1}{4}\beta^2}, \quad \frac{\beta}{2} - \sqrt{\alpha^2 + \frac{1}{4}\beta^2}$

positive negative

$$\left\| \begin{bmatrix} 0 & \alpha \\ \alpha & \beta \end{bmatrix} \right\|_1 = \frac{\beta}{2} + \sqrt{\alpha^2 + \frac{1}{4}\beta^2} - \left(\frac{\beta}{2} - \sqrt{\alpha^2 + \frac{1}{4}\beta^2} \right)$$

$$= 2\sqrt{\alpha^2 + \frac{1}{4}\beta^2} \leq 2 \qquad (\because \alpha^2 + \beta^2 = 1)$$

$$\left\| \Pi_s |\psi_{x^n}\rangle\langle\psi_{x^n}| \Pi_s - |\psi_{x^n}\rangle\langle\psi_{x^n}| \right\|_1$$

$$= \left\| \beta \begin{bmatrix} 0 & \alpha \\ \alpha & \beta \end{bmatrix} \right\|_1 \leq 2\beta = 2\sqrt{1 - \alpha^2}, \qquad \alpha = \langle\psi_{x^n}|\Pi_s|\psi_{x^n}\rangle$$

Remarks:

1. If the states $|\psi_x\rangle$ are orthogonal, $S(\rho) = H(X)$ and Schumacher compression coincides with the classical data compression protocol last time.

2. Schumacher compression is rate optimal (see next pages). In fact, even visible compression of pure state requires the same rate, so, the sender's knowledge does not reduce the rate in this set up.

Other scenarios and bounds on the rate:

See p2-3 from arXiv:1911.09126 for a summary.

Following: excerpt from a recent talk.

Quantum data compression scenarios:

1. pure or mixed, quantum vs classical ensembles
2. blind vs visible (latter: Alice knows x1, ..., xn)
3. assistance: entanglement, shared coins, none
4. global vs local error
5. asymptotic vs one-shot

Quantum data compression optimal rates:

Lower bound in unassisted scenario

Ensemble:
$X = x$ wp p_x, state ρ_x

WARNING: WE USE ρ_x earlier this lecture

Let $\rho = \sum_x p_x |x\rangle\langle x|_X \otimes \rho_{x\,C}$

$$\chi = I(X:C)_\rho = S\left(\sum_x p_x \rho_x\right) - \sum_x p_x S(\rho_x)$$

Holevo information

Theorem: $r \geq \chi$

M Horodecki 98,
Barnum, Caves, Fuchs, Jozsa, Schumacher 00

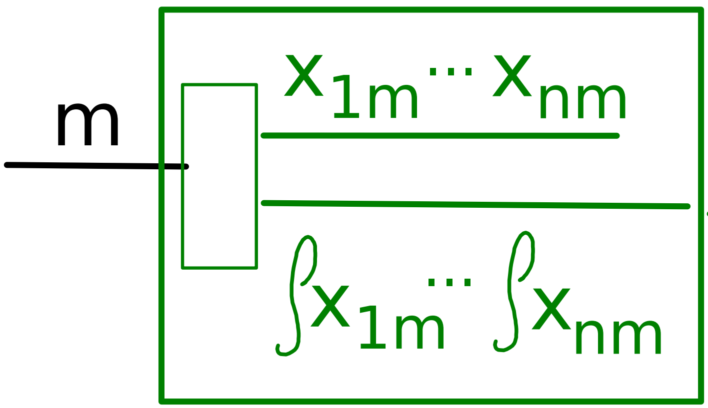
Quantum data compression optimal rates:

Lower bound in unassisted scenario

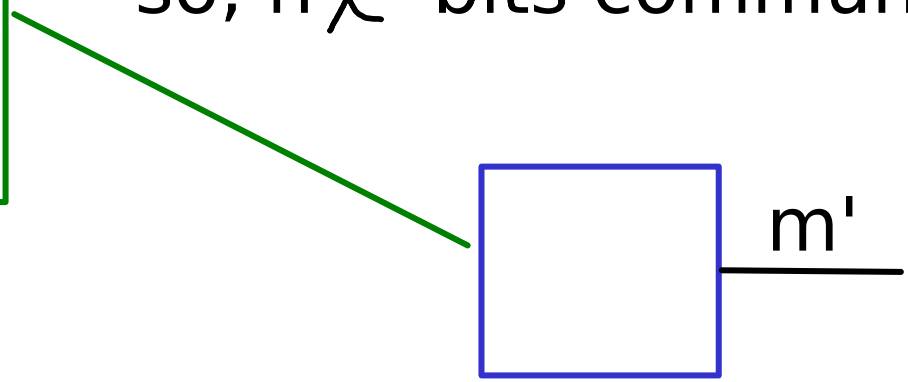
Theorem: $r \geqslant \chi$

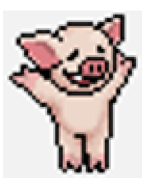
Proof idea (visible, implies same for blind):

HSW theorem (capacity of q states to convey bits):



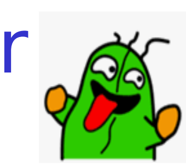
if $m \in \{1, \ldots, 2^{n\chi}\}$
then $m=m'$ with high prob
so, $n\chi$ bits communicated





encoder
Alice

decoder
Bob



where each $x_{ij} \sim X$ iid

Quantum data compression optimal rates:

Lower bound in unassisted scenario

Theorem: $r \geq \chi$

Proof idea (visible, implies same for blind):

HSW theorem (capacity of q states to convey bits):

if $m \in \{1, \ldots, 2^{n\chi}\}$
then m=m' with high prob
so, $n\chi$ bits communicated

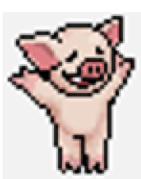
m

$x_{1m} \cdots x_{nm}$

$\rho_{x_{1m}} \cdots \rho_{x_{nm}}$

compression

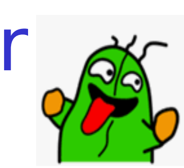
m'



encoder
Alice

where each $x_{ij} \sim X$ iid

decoder
Bob



Quantum data compression optimal rates:

Lower bound in unassisted scenario

Theorem: $r \geqslant \chi$

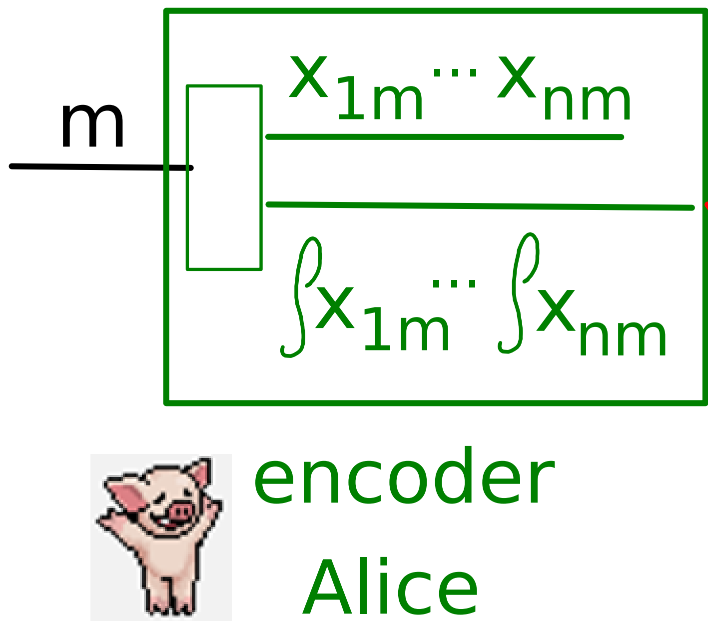
Proof idea (visible, implies same for blind):

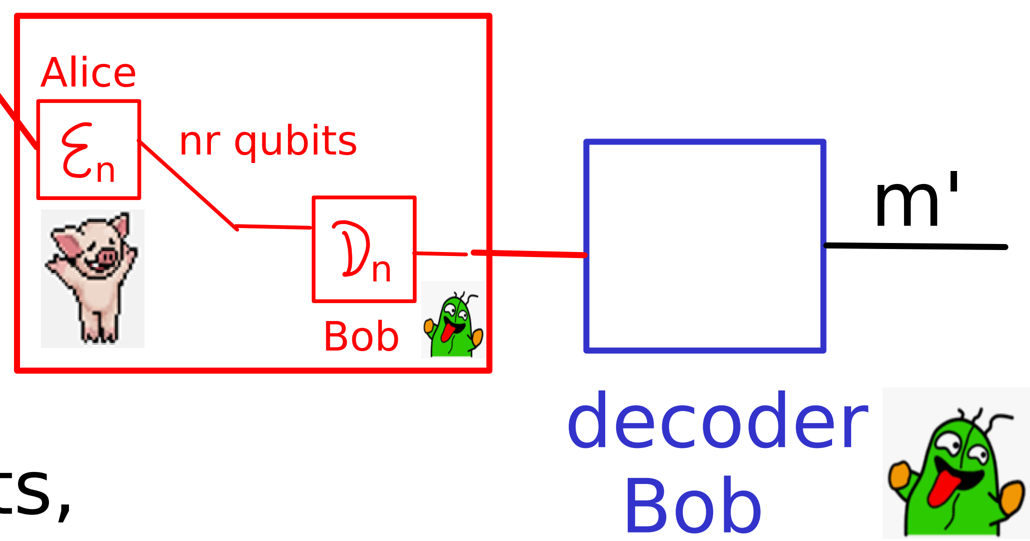
HSW theorem (capacity of q states to convey bits):

if $m \in \{1, \ldots, 2^{n\chi}\}$

then $m=m'$ with high prob

so, $n\chi$ bits communicated





Holevo 73: to send k bits,
need k qubits, so, $nr \geqslant n\chi$

Quantum data compression optimal rates:

Lower bound in entanglement-assisted scenario

χ BITS / copy (quantum reverse Shannon thm)

visible case, implies same for blind compression

M Horodecki 00,
Barnum, Caves, Fuchs, Jozsa, Schumacher 00
Bennett, Shor, Smolin, Thapliyal 01
Bennett, Devetak, Harrow, Shor, Winter "14" + Berta et al

Is the Holevo info achievable for compression?