Entropies & Information Theory

LECTURE II

Nilanjana Datta

University of Cambridge,U.K.

See lecture notes on: http://www.qi.damtp.cam.ac.uk/node/223

UNIVERSITY OF CAMBRIDGE

quantum system ⟷ Hilbert space $\mathcal{H}$ (finite-dimensional) (state space)

- States (of a physical system): density matrices

$$\rho \geq 0, \ \mathrm{Tr}\,\rho = 1$$

More generally: if a quantum system is in pure states:

$$|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_k\rangle \in \mathcal{H}, \quad \text{with probs. } p_1, p_2, \ldots, p_k$$

$$\mathcal{E} = \{p_i, |\psi_i\rangle\} \leftrightarrow \rho = \sum_{i=1}^{k} p_i |\psi_i\rangle\langle\psi_i|$$

in general $\langle\psi_i|\psi_j\rangle \neq \delta_{ij}$

- Spectral decomposition:

$$\rho = \sum_{i=1}^{d} \lambda_i |\varphi_i\rangle\langle\varphi_i|;$$

eigenvalues (λ_i), eigenvectors ($|\varphi_i\rangle$)

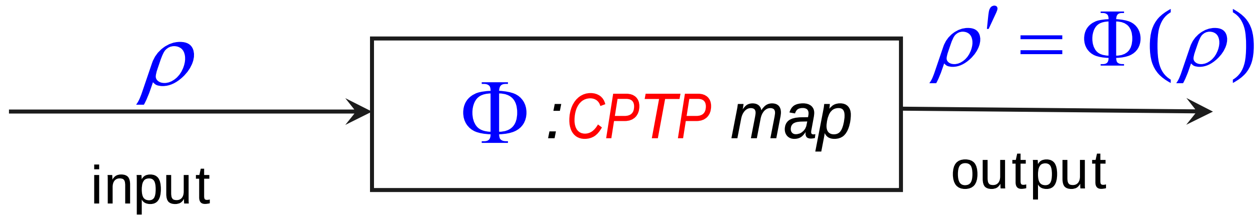
$$\lambda_i \geq 0, \ \sum_{i=1}^{d} \lambda_i = 1$$

$\{\lambda_i\}_{i=1}^{d}$: probability distribution

Quantum Operations or Quantum Channels

- Any allowed physical process that a quantum system can undergo is described by a :

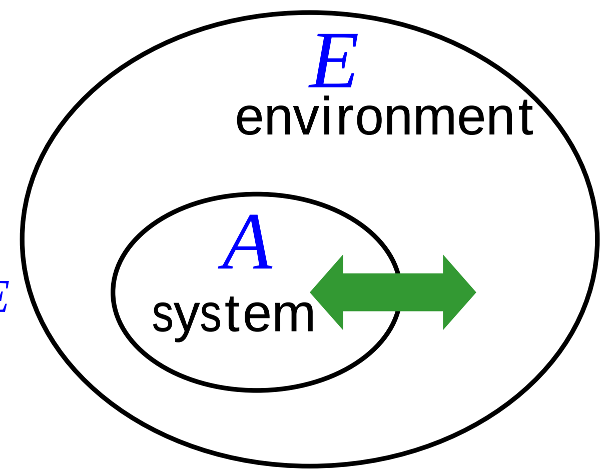
linear completely-positive, trace preserving (CPTP) map



- *Trace-preserving (TP):* $\mathrm{Tr}\,\rho' = \mathrm{Tr}\rho = 1$
- *Positive:* $\rho' = \Phi(\rho) \geq 0$
- *Completely positive (CP):*

$$\Phi : \mathcal{D}(\mathcal{H}_A) \to \mathcal{D}(\mathcal{H}_B)$$

ρ_{AE}



$(\Phi \otimes id_E)(\rho_{AE})$ = *an allowed state of the composite system* $\in \mathcal{D}(\mathcal{H}_B \otimes \mathcal{H}_E)$

$$(\Phi \otimes id_E)(\rho_{AE}) \geq 0$$

Generalized measurements – POVM:

A quantum measurement is described by a **POVM**

$E = \{E_i\};$ (finite set) $E_i \geq 0, \ \sum_i E_i = I$

If the system is in a state ρ before the measurement,

Then, probability of getting the i^{th} outcome is:

$$p_i = \mathrm{Tr}(E_i \rho)$$

Purification

Any mixed state

$$\rho_A \in \mathcal{H}_A$$

A pure state

$$|\psi_{AR}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_R;$$

$$\rho_A = \mathrm{Tr}_R |\Psi_{AR}\rangle\langle\Psi_{AR}|;$$

purifying reference system

Von Neumann entropy

of a state ρ :

$$S(\rho) := -\mathrm{Tr}\,(\rho \log \rho)$$

$\log \equiv \log_2$

Spectral decomposition: $\rho = \sum_{i=1}^{d} \lambda_i |\varphi_i\rangle\langle\varphi_i|;$

$$S(\rho) := -\mathrm{Tr}\,(\rho \log \rho) = -\sum_{i=1}^{d} \lambda_i \log \lambda_i$$

$$= H(\{\lambda_i\})$$ *Shannon entropy*

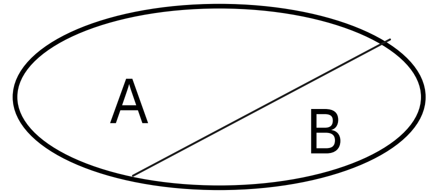
$S(\rho) = 0$ if and only if ρ is a pure state: $\rho = |\Psi\rangle\langle\Psi|$

$\therefore S(\rho) =$ a measure of the "mixedness" of the state ρ

UNIVERSITY OF CAMBRIDGE

Other Entropies

For a bipartite system in a state ρ_{AB} :



- **Joint entropy:**

$$S(\rho_{AB}) = -\mathrm{Tr}(\rho_{AB} \log \rho_{AB})$$

- **Conditional entropy:**

$$S(A|B)_\rho := S(\rho_{AB}) - S(\rho_B)$$

$$\rho_B = \mathrm{Tr}_A \rho_{AB}$$

reduced state

- **Quantum mutual information:**

$$I(A:B)_\rho := S(\rho_A) + S(\rho_B) - S(\rho_{AB});$$

UNIVERSITY OF CAMBRIDGE

Quantum Relative Entropy

- A fundamental quantity in Quantum Mechanics & Quantum Information Theory is the Quantum Relative Entropy

  of ρ w.r.t. σ, $\rho \geq 0,\ \mathrm{Tr}\,\rho = 1,\ \sigma \geq 0$:

$$\boxed{D(\rho \| \sigma) := \mathrm{Tr}\,\rho \log \rho - \mathrm{Tr}\,\rho \log \sigma}$$

$\log \equiv \log_2$

well-defined if $\mathrm{supp}\,\rho \subseteq \mathrm{supp}\,\sigma$

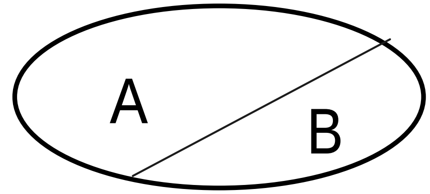
- It acts as a parent quantity for the *von Neumann entropy*:

$$\boxed{S(\rho) := -\mathrm{Tr}\,\rho \log \rho = -D(\rho \| I)} \qquad (\sigma = I)$$

UNIVERSITY OF CAMBRIDGE

- It also acts as a parent quantity for other entropies:

e.g. for a bipartite state ρ_{AB} :



- *Conditional entropy*

$$S(A|B) := S(\rho_{AB}) - S(\rho_B) = -D(\rho_{AB} \| I_A \otimes \rho_B)$$

- *Mutual information*

$$\rho_B = \mathrm{Tr}_A\, \rho_{AB}$$

$$I(A:B) := S(\rho_A) + S(\rho_B) - S(\rho_{AB}) = D(\rho_{AB} \| \rho_A \otimes \rho_B)$$

UNIVERSITY OF CAMBRIDGE

Some Properties of $D(\rho \| \sigma)$

- *"distance"*

  ~~*symmetric*~~
  ~~*triangle inequality*~~

$$D(\rho \| \sigma) \geq 0 \qquad \rho, \sigma \text{ states}$$

$$= 0 \text{ if \& only if } \rho = \sigma$$

……….(1)

- Monotonicity under a quantum operation (CPTP map)

$$D(\Lambda(\rho) \| \Lambda(\sigma)) \leq D(\rho \| \sigma)$$

……….(2)

Many properties of other entropies can be proved using (1) & (2)

UNIVERSITY OF CAMBRIDGE

Properties of quantum entropies

- $S(\rho) \geq 0; \quad S(\rho) \leq \log d;$ *where* $d = \dim \mathcal{H}$
- *Subadditivity:* $S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B)$
- *Concavity:* $S\left(\sum_i p_i \rho_i\right) \geq \sum_i p_i S(\rho_i)$
- *Invariance under unitaries:* $S(U \rho U^\dagger) = S(\rho)$
- $H(X) \leq H(X,Y)$ but $S(\rho_{AB}) \leq S(\rho_A)$ is possible!!

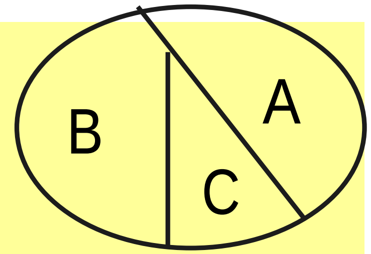
  $X, Y:$ classical r.v.s

  Conditional entropy $S(A|B)_\rho$ *can be negative!*
- *Araki-Lieb inequality:* $S(\rho_{AB}) \geq |S(\rho_A) - S(\rho_B)|$

UNIVERSITY OF CAMBRIDGE

Properties of quantum entropies contd.

- Strong subadditivity: ρ_{ABC} tripartite state



$$S(\rho_{ABC}) + S(\rho_B) \le S(\rho_{AB}) + S(\rho_{BC})$$

Lieb & Ruskai '73

Consequences of strong subadditivity:

- Conditioning reduces entropy $S(A|BC)_\rho \le S(A|B)_\rho$
- Discarding quantum systems never increases mutual information

$$I(A:B)_\rho \le I(A:BC)_\rho$$

- Quantum operations never increase mutual information

$$I(A:B')_\sigma \le I(A:B)_\rho; \quad \sigma_{AB'} = (\mathrm{id}_A \otimes \Lambda_{B\to B'})\rho_{AB}$$

UNIVERSITY OF CAMBRIDGE

- *Operational significance of the von Neumann entropy*

= optimal rate of data compression for a memoryless (i.i.d.) quantum information source

UNIVERSITY OF CAMBRIDGE

Quantum Data Compression

Quantum Info source → *signals*

signals (*pure states*) $|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_r\rangle \in \mathcal{H}$

with probabilities $p_1, p_2, \ldots, p_r$ $\qquad \langle \psi_i | \psi_j \rangle \neq \delta_{ij}$

- Then source characterized by: $\{\rho, \mathcal{H}\}$

$$\rho = \sum_{i=1}^{r} p_i |\psi_i\rangle\langle\psi_i|$$

density matrix

- Memoryless quantum information source

State of n copies of the source: $\rho_n = \rho^{\otimes n}$ → no correlation

Quantum data compression

- Evaluated in the asymptotic limit $n \to \infty$

  $n =$ number of copies/uses of the source

- emits signals $|\psi_1^{(n)}\rangle, |\psi_2^{(n)}\rangle, \ldots, |\psi_m^{(n)}\rangle \in \mathcal{H}^{\otimes n}$

- with probs. $p_1^{(n)}, p_2^{(n)}, \ldots, p_m^{(n)}$

  $\langle \psi_i^{(n)} | \psi_j^{(n)} \rangle \neq \delta_{ij}$ in general

- State : $\rho_n = \sum_{i=1}^{m} p_i^{(n)} |\psi_i^{(n)}\rangle\langle\psi_i^{(n)}|$

Compression-Decompression Scheme

- Encoding: $\mathcal{E}_n : |\psi_i^{(n)}\rangle\langle\psi_i^{(n)}| \to \sigma_i^{(n)} \in \mathcal{D}(\mathcal{H}_c^n)$

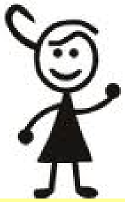
  signal → compressed state; $\mathcal{H}_c^n$: compressed Hilbert space

- Decoding: $\mathcal{D}_n : \sigma_i^{(n)} \to \omega_i^{(n)} \in \mathcal{D}(\mathcal{H}^{\otimes n})$

  $\omega_i^{(n)}$: recovered signal

UNIVERSITY OF CAMBRIDGE

Quantum Data Compression



n copies of a quantum information source ρ_n

signal $|\psi_i^{(n)}\rangle\langle\psi_i^{(n)}| \in \mathcal{B}(\mathcal{H}^{\otimes n})$ with prob. $p_i^{(n)}$

$\mathcal{E}_n$ → $\sigma_i^{(n)}$ → $\mathcal{D}_n$ → $\omega_i^{(n)}$

compressed state in $\mathcal{H}_c^n$

$\dim \mathcal{H}_c^n < \dim \mathcal{H}^{\otimes n}$

recovered state $\in \mathcal{B}(\mathcal{H}^{\otimes n})$

- Require: ensemble average fidelity $\bar{F}_n \to 1$ as $n \to \infty$ ....(a)

$$\bar{F}_n = \sum p_i^{(n)} \langle \psi_i^{(n)} | \mathcal{D}_n \circ \mathcal{E}_n \left(|\psi_i^{(n)}\rangle\langle\psi_i^{(n)}| \right) | \psi_i^{(n)} \rangle$$

- Optimal rate of data compression: **Data compression limit**

- Minimum value of $R_\infty := \lim_{n\to\infty} \frac{\log(\dim \mathcal{H}_c^n)}{n}$ such that (a) holds

UNIVERSITY OF CAMBRIDGE

Memoryless quantum information source

state of n *copies of the source* $$\rho_n = \sum_{i=1}^{m} p_i^{(n)} \left|\psi_i^{(n)}\right\rangle\left\langle\psi_i^{(n)}\right| = \rho^{\otimes n};$$

$\left|\psi_i^{(n)}\right\rangle$:*signal emitted with prob.* $p_i^{(n)}$; $\left\langle\psi_i^{(n)}\middle|\psi_j^{(n)}\right\rangle \neq \delta_{ij}$

$$\rho \in \mathcal{D}(\mathcal{H}), \dim\mathcal{H} = d \qquad \therefore \rho_n = \rho^{\otimes n} \in \mathcal{D}(\mathcal{H}^{\otimes n})$$

Spectral decompositions:

$$\rho = \sum_{j=1}^{d} q_j \left|\varphi_j\right\rangle\left\langle\varphi_j\right|; \qquad \rho_n = \sum_{k=1}^{d^n} \lambda_k^{(n)} \left|\Psi_k^{(n)}\right\rangle\left\langle\Psi_k^{(n)}\right|$$

$$\therefore \rho_n = \rho^{\otimes n} \Rightarrow \quad \left|\Psi_k^{(n)}\right\rangle = \left|\varphi_{k_1}\right\rangle \otimes \left|\varphi_{k_2}\right\rangle \otimes \ldots \left|\varphi_{k_n}\right\rangle$$

$$\lambda_k^{(n)} = q_{k_1} q_{k_2} \ldots\ldots q_{k_n}$$

Identification of the label k *as a sequence of classical indices*

$$k = (k_1, k_2, \ldots\ldots, k_n)$$

UNIVERSITY OF CAMBRIDGE

Memoryless quantum information source

state of n *copies of the source*

$$\rho_n = \sum_{i=1}^{m} p_i^{(n)} \left|\psi_i^{(n)}\right\rangle\left\langle\psi_i^{(n)}\right| = \rho^{\otimes n};$$

$\left|\psi_i^{(n)}\right\rangle$:*signal emitted with prob.* $p_i^{(n)}$; $\quad \left\langle\psi_i^{(n)}\middle|\psi_j^{(n)}\right\rangle \neq \delta_{ij}$

$$\rho \in \mathcal{D}(\mathcal{H}), \dim\mathcal{H} = d \;\therefore\; \rho_n = \rho^{\otimes n} \in \mathcal{H}^{\otimes n}$$

Spectral decompositions:

$$\rho = \sum_{j=1}^{d} q_j \left|\varphi_j\right\rangle\left\langle\varphi_j\right|; \qquad \rho_n = \sum_{\underline{k}} \lambda_{\underline{k}}^{(n)} \left|\Psi_{\underline{k}}^{(n)}\right\rangle\left\langle\Psi_{\underline{k}}^{(n)}\right|$$

$$\therefore \rho_n = \rho^{\otimes n} \Rightarrow \begin{cases} \left|\Psi_{\underline{k}}^{(n)}\right\rangle = \left|\varphi_{k_1}\right\rangle \otimes \left|\varphi_{k_2}\right\rangle \otimes \ldots \left|\varphi_{k_n}\right\rangle \\ \lambda_{\underline{k}}^{(n)} = q_{k_1} q_{k_2} \ldots\ldots q_{k_n} \end{cases}$$

Identification of the label k *as a sequence of classical indices*

$$\underline{k} \equiv k = (k_1, k_2, \ldots\ldots, k_n)$$

UNIVERSITY OF CAMBRIDGE

- sum over all possible sequences

$$\rho_n \equiv \rho^{\otimes n} = \sum_{\underline{k}} \lambda_{\underline{k}}^{(n)} \left|\Psi_{\underline{k}}^{(n)}\right\rangle\left\langle\Psi_{\underline{k}}^{(n)}\right|$$

$$\underline{k} \equiv (k_1, k_2, \ldots\ldots, k_n):$$
$$k_i \in \{1, 2, .., d\}; \quad d = \dim \mathcal{H}$$

$$\lambda_{\underline{k}}^{(n)} = q_{k_1} q_{k_2} \ldots\ldots q_{k_n}$$

von Neumann entropy $$S(\rho_n) = S(\rho^{\otimes n}) = nS(\rho) = nH(\{q_k\})$$

Probability: $$p(\underline{k}) \equiv \lambda_{\underline{k}}^{(n)} = q_{k_1} q_{k_2} \ldots\ldots q_{k_n}$$

$\forall \varepsilon > 0$, a sequence $\underline{k} \equiv (k_1, k_2, \ldots\ldots, k_n)$ is $\varepsilon -$typical if:

$$2^{-n(H(\{q_k\})+\varepsilon)} \leq p(\underline{k}) \leq 2^{-n(H(\{q_k\})-\varepsilon)},$$

$$2^{-n(S(\rho)+\varepsilon)} \leq p(\underline{k}) \leq 2^{-n(S(\rho)-\varepsilon)},$$

$T_\varepsilon^{(n)} := \varepsilon -$ typical set

eigenvalues $\lambda_{\underline{k}}^{(n)}$ ⟷ sequences $\underline{k}$

eigenvectors $\left|\Psi_{\underline{k}}^{(n)}\right\rangle$ → $\mathcal{T}_\varepsilon^{(n)} := \varepsilon -$ typical subspace

UNIVERSITY OF CAMBRIDGE

$\varepsilon -$ typical subspace $\mathcal{T}_\varepsilon^{(n)} \subset \mathcal{H}^{\otimes n}$

- Subspace spanned by those eigenvectors

$$\left|\Psi_{\underline{k}}^{(n)}\right\rangle = \left|\varphi_{k_1}\right\rangle \otimes \left|\varphi_{k_2}\right\rangle \otimes \ldots \left|\varphi_{k_n}\right\rangle \quad \text{for which} \quad \underline{k} \in T_\varepsilon^{(n)}$$

- Let $P_\varepsilon^{(n)}$: orthogonal projection on to the typical subspace

Typical Sequence Theorem → **Typical Subspace Theorem**

Fix $\varepsilon > 0,$ then $\forall \delta > 0,$ and n large enough:

Typical Sequence Theorem	Typical Subspace Theorem
$P\left(T_\varepsilon^{(n)}\right) > 1-\delta$	$\mathrm{Tr}\left(P_\varepsilon^{(n)}\rho_n\right) > 1-\delta$
$(1-\delta)2^{n(H(\{q_k\})-\varepsilon)} \le \left|T_\varepsilon^{(n)}\right| \le 2^{n(H(\{q_k\})+\varepsilon)}$	$(1-\delta)2^{n(S(\rho)-\varepsilon)} \le \dim \mathcal{T}_\varepsilon^{(n)} \le 2^{n(S(\rho)+\varepsilon)}$

UNIVERSITY OF CAMBRIDGE

Idea behind the compression scheme

$\left|\psi_i^{(n)}\right\rangle$: signal emitted with prob. $p_i^{(n)}$; $\quad \left\langle \psi_i^{(n)} \middle| \psi_j^{(n)} \right\rangle \neq \delta_{ij}$

$$\left|\psi_i^{(n)}\right\rangle = P_\varepsilon^{(n)}\left|\psi_i^{(n)}\right\rangle + \left(I - P_\varepsilon^{(n)}\right)\left|\psi_i^{(n)}\right\rangle$$

$\in \mathcal{T}_\varepsilon^{(n)}$ keep this part unchanged

$\notin \mathcal{T}_\varepsilon^{(n)}$ map this onto a fixed pure state $\left|\phi_0^{(n)}\right\rangle \in \mathcal{T}_\varepsilon^{(n)}$

Compression scheme $\quad \mathcal{E}_n\left(\left|\psi_i^{(n)}\right\rangle\left\langle\psi_i^{(n)}\right|\right) = \tilde{\rho}_i^{(n)}$

$$\tilde{\rho}_i^{(n)} = \alpha_i^2 \left|\tilde{\psi}_i^{(n)}\right\rangle\left\langle\tilde{\psi}_i^{(n)}\right| + \beta_i^2 \left|\phi_0^{(n)}\right\rangle\left\langle\phi_0^{(n)}\right| \in \mathcal{D}\left(\mathcal{T}_\varepsilon^{(n)}\right)$$

$$\left|\tilde{\psi}_i^{(n)}\right\rangle \propto P_\varepsilon^{(n)}\left|\psi_i^{(n)}\right\rangle; \alpha_i^2 = \left\|P_\varepsilon^{(n)}\left|\psi_i^{(n)}\right\rangle\right\|^2; \beta_i^2 = \left\|\left(I - P_\varepsilon^{(n)}\right)\left|\psi_i^{(n)}\right\rangle\right\|^2$$

Decompression scheme $\quad \mathcal{D}_n\left(\tilde{\rho}_i^{(n)}\right) = \tilde{\rho}_i^{(n)} \oplus 0$

UNIVERSITY OF CAMBRIDGE

$$\tilde{\rho}_i^{(n)} = \alpha_i^2 \left|\tilde{\psi}_i^{(n)}\right\rangle\left\langle\tilde{\psi}_i^{(n)}\right| + \beta_i^2 \left|\phi_0^{(n)}\right\rangle\left\langle\phi_0^{(n)}\right| \in \mathcal{D}\left(\mathcal{T}_\varepsilon^{(n)}\right)$$

Ensemble average fidelity

$$\bar{F}_n = \sum_i p_i^{(n)} \left\langle \psi_i^{(n)} \middle| \tilde{\rho}_i^{(n)} \middle| \psi_i^{(n)} \right\rangle \geq 2\sum_i p_i^{(n)} \alpha_i^2 - 1$$

Schumacher proved (1995): for a memoryless source

$$\{\rho, \mathcal{H}\}$$

Data compression limit = $S(\rho)$: *von Neumann entropy of the source*

UNIVERSITY OF CAMBRIDGE

Schumacher's Theorem : Quantum Data Compression

Suppose $\{\rho, \mathcal{H}\}$ is an *memoryless, quantum information source*

$\rho_n = \rho^{\otimes n}$; $S(\rho)$: von Neumann entropy

- Suppose $R > S(\rho)$: then there exists a reliable compression scheme of rate R for the source.

- If $R < S(\rho)$ then any compression scheme of rate R will not be reliable.

Proof follows from the Typical Subspace theorem

UNIVERSITY OF CAMBRIDGE

Schumacher's Theorem : Quantum Data Compression

- Suppose $R > S(\rho)$: then there exists a reliable compression scheme of rate R for the source.

- **Proof:**

Compressed Hilbert space $\mathcal{H}_c^n; \quad \dim \mathcal{H}_c^n = 2^{nR}$ $\qquad R > S(\rho)$

- *Choose* $\varepsilon > 0$, *such that* $R > S(\rho) + \varepsilon$

Fix $\delta > 0$, choose n large enough such that:

$$\mathrm{Tr}\left(P_\varepsilon^{(n)} \rho_n\right) > 1 - \delta; \qquad \dim \mathcal{T}_\varepsilon^{(n)} \leq 2^{n(S(\rho)+\varepsilon)} < 2^{nR} = \dim \mathcal{H}_c^n$$

$$\Rightarrow \quad \mathcal{T}_\varepsilon^{(n)} \subset \mathcal{H}_n^c$$

UNIVERSITY OF CAMBRIDGE

Ensemble average fidelity

$$\bar{F}_n = \sum_i p_i^{(n)} \left\langle \psi_i^{(n)} \middle| \tilde{\rho}_i^{(n)} \middle| \psi_i^{(n)} \right\rangle \geq 2\sum_i p_i^{(n)} \alpha_i^2 - 1$$

$$> 1 - 2\delta$$

(by the Typical Subspace Theorem)

$$\Rightarrow \quad \bar{F}_n \rightarrow 1 \quad \text{as} \quad n \rightarrow \infty$$

■

UNIVERSITY OF CAMBRIDGE

Schumacher's Theorem : Quantum Data Compression

Suppose $\{\rho, \mathcal{H}\}$ is an *memoryless, quantum information source*

$$\rho_n = \rho^{\otimes n}; \quad S(\rho):$$ von Neumann entropy

- Suppose $R > S(\rho)$: then there exists a reliable compression scheme of rate R for the source.

- If $R < S(\rho)$ then any compression scheme of rate R will not be reliable.

(See Cambridge lecture notes)