Compression of Quantum Mixed State Sources

Emina Soljanin

Bell Labs, visiting EPFL

UMLV&EPFL Entropy Workshop

Lausanne, September 2008

A Classical Source of Information

- Discrete: produces sequences of letters.
- Letters belong to a finite alphabet $\mathcal{X}$.
- Memoryless: each letter is produced independently.
- Probability of letter a is P_x.
- Example: coin tossing with $\mathcal{X} = \{H, T\}$.
- Shannon Entropy: $-\sum_x P_x \log P_x$

A Quantum Source of Information

- Quantum letters are represented as unit-length vectors in $\mathcal{H}_d$.
- A qubit is a vector in $\mathcal{H}_2$.
- Example: Alphabet $\mathcal{X} = \{0, 1, 2, 3\}$ mapped onto 4 qubits

$$|\psi_0\rangle = \alpha_0|e_0\rangle + \beta_0|e_1\rangle \quad |\psi_1\rangle = \alpha_1|e_0\rangle + \beta_1|e_1\rangle$$
$$|\psi_2\rangle = \alpha_2|e_0\rangle + \beta_2|e_1\rangle \quad |\psi_3\rangle = \alpha_3|e_0\rangle + \beta_3|e_1\rangle.$$

where $|e_0\rangle$ and $|e_1\rangle$ are the basis vectors of 2D space $\mathcal{H}_2$:

$$|e_0\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad |e_1\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

- We will deal with (sequences of) qubits, WOLG.

The Density Matrix and Entropy

- Source density matrix:

$$\rho = \sum_{a \in \mathcal{X}} P_x \underbrace{|\psi_x\rangle\langle\psi_x|}_{\rho_x}.$$

The Density Matrix and Entropy

- Source density matrix:

$$\rho = \sum_{a \in \mathcal{X}} P_x \underbrace{|\psi_x\rangle\langle\psi_x|}_{\rho_x}.$$

- Von Neumann entropy of the source:

$$S(\rho) = -\operatorname{Tr} \rho \log \rho$$
$$= -\sum_i \lambda_i \log \lambda_i,$$

where λ_i are the eigenvalues of ρ.

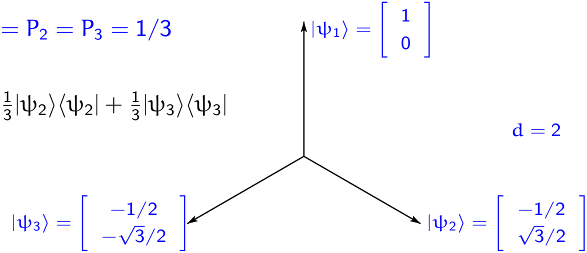
THE MB EXAMPLE

$$\mathcal{X} = \{1, 2, 3\} \quad P_1 = P_2 = P_3 = 1/3$$

$$\rho = \tfrac{1}{3}|\psi_1\rangle\langle\psi_1| + \tfrac{1}{3}|\psi_2\rangle\langle\psi_2| + \tfrac{1}{3}|\psi_3\rangle\langle\psi_3|$$

$$= \tfrac{1}{2}I$$

$$S(\rho) = 1$$



$d = 2$

Vector Sequences

- Source vector-sequence (state)

$$|\Psi_{\boldsymbol{x}}\rangle = |\psi_{x_1}\rangle \otimes |\psi_{x_2}\rangle \otimes \cdots \otimes |\psi_{x_n}\rangle, \qquad x_i \in \mathcal{X},$$

  appears with probability $P_{\boldsymbol{x}} = P_{x_1} \cdot P_{x_2} \cdot \ldots \cdot P_{x_n}$.
- Typical states $|\Psi_{\boldsymbol{x}}\rangle \in \mathcal{H}^{2^n}$ correspond to typical sequences $\boldsymbol{x}$.
- There are approximately $2^{nH(P)}$ typical states.

Lossless Quantum Data Compression

- Source vector-sequence $|\Psi_{\boldsymbol{x}}\rangle$ is in $\mathcal{H}^{2^n}$, $(\boldsymbol{x} \in \mathcal{X}^n)$
- Vector $|\Psi_{\boldsymbol{x}}\rangle$ is compressed and then reproduced as $|\widehat{\Psi_{\boldsymbol{x}}}\rangle$.
- Fidelity between $|\Psi_{\boldsymbol{x}}\rangle$ and $|\widehat{\Psi_{\boldsymbol{x}}}\rangle$:

$$F(|\Psi_{\boldsymbol{x}}\rangle, |\widehat{\Psi_{\boldsymbol{x}}}\rangle) = |\langle \Psi_{\boldsymbol{x}} | \widehat{\Psi_{\boldsymbol{x}}} \rangle|^2$$

Lossless Quantum Data Compression

- Source vector-sequence $|\Psi_{\boldsymbol{x}}\rangle$ is in $\mathcal{H}^{2^n}$, $(\boldsymbol{x} \in \mathcal{X}^n)$
- Vector $|\Psi_{\boldsymbol{x}}\rangle$ is compressed and then reproduced as $|\widehat{\Psi_{\boldsymbol{x}}}\rangle$.
- Fidelity between $|\Psi_{\boldsymbol{x}}\rangle$ and $|\widehat{\Psi_{\boldsymbol{x}}}\rangle$:

$$F(|\Psi_{\boldsymbol{x}}\rangle, |\widehat{\Psi_{\boldsymbol{x}}}\rangle) = |\langle \Psi_{\boldsymbol{x}} | \widehat{\Psi_{\boldsymbol{x}}}\rangle|^2$$

- For asymptotically lossless compression, the average fidelity

$$\bar{F} = \sum_{\boldsymbol{x} \in \mathcal{X}^n} P(\boldsymbol{x}) F(|\Psi_{\boldsymbol{x}}\rangle, |\widehat{\Psi_{\boldsymbol{x}}}\rangle)$$

should approach 1 as $n \to \infty$.

Typical States and Visible Compression

- Visible: the encoder Alice knows sequence $\boldsymbol{x}$.
- She can compress with perfect fidelity the typical states.
- Instead of n qubits, she can transmit $nH(P)$ bits.
- The decoder Bob prepares $|\Psi_{\boldsymbol{x}}\rangle$ as $|\widehat{\Psi_{\boldsymbol{x}}}\rangle$ for typical $\boldsymbol{x}$.

Typical States and Visible Compression

- Visible: the encoder Alice knows sequence x.
- She can compress with perfect fidelity the typical states.
- Instead of n qubits, she can transmit $nH(P)$ bits.
- The decoder Bob prepares $|\Psi_x\rangle$ as $|\widehat{\Psi_x}\rangle$ for typical x.
- Can $|\Psi_x\rangle$ be compressed to fewer than $nH(P)$ qubits so that
  - the compression is asymptotically lossless
  - Alice does not know x
  - Alice and Bob perform legal quantum operations

What Can be Done – Evolution (Reversible)

State ρ can be transformed to another state $\mathcal{E}(\rho)$ only by a physical process consistent with the lows of quantum theory:

What Can be Done – Evolution (Reversible)

State ρ can be transformed to another state $\mathcal{E}(\rho)$ only by a physical process consistent with the lows of quantum theory:

- unitary evolution:

$$\mathcal{E}(\rho) = U\rho U^\dagger \quad \text{where} \quad UU^\dagger = I,$$

What Can be Done – Evolution (Reversible)

State ρ can be transformed to another state $\mathcal{E}(\rho)$ only by a physical process consistent with the lows of quantum theory:

- unitary evolution:

$$\mathcal{E}(\rho) = U\rho U^\dagger \quad \text{where} \quad UU^\dagger = I,$$

- completely positive, trace-preserving map:

$$\mathcal{E}(\rho) = \sum_k E_k \rho E_k^\dagger \quad \text{where} \quad \sum_k E_k^\dagger E_k = I.$$

What Can be Done – Measurement (Irreversible)

- Von Neumann:
  - A set of pairwise orthogonal projection operators $\{\Pi_i\}$.
  - They form a complete resolution of the identity: $\sum_i \Pi_i = I$.
  - $|\psi_j\rangle$ is measured as $\Pi_i|\psi_j\rangle$ with probability $\langle\psi_j|\Pi_i|\psi_j\rangle$.

What Can be Done – Measurement (Irreversible)

- Von Neumann:
  - A set of pairwise orthogonal projection operators $\{\Pi_i\}$.
  - They form a complete resolution of the identity: $\sum_i \Pi_i = I$.
  - $|\psi_j\rangle$ is measured as $\Pi_i|\psi_j\rangle$ with probability $\langle\psi_j|\Pi_i|\psi_j\rangle$.
- Positive Operator-Valued Measure (POVM):
  - Any set of positive-semidefinite operators $\{E_i\}$.
  - They form a complete resolution of the identity: $\sum_i E_i = I$.
  - $|\psi_j\rangle$ is measured as $E_i|\psi_j\rangle$ with probability $\langle\psi_j|E_i|\psi_j\rangle$.

What Cannot be Done?

What Cannot be Done?

- The No-Cloning Principle:

  There is no physical process that leads to an evolution

$$|\phi\rangle \otimes |s\rangle \to |\phi\rangle \otimes |\phi\rangle$$

  where $|\phi\rangle$ is an arbitrary state and $|s\rangle$ is a fixed state.

What Cannot be Done?

- The No-Cloning Principle:

  There is no physical process that leads to an evolution

  $$|\phi\rangle \otimes |s\rangle \rightarrow |\phi\rangle \otimes |\phi\rangle$$

  where $|\phi\rangle$ is an arbitrary state and $|s\rangle$ is a fixed state.
- The No-Broadcasting Principle – generalization of no-cloning.

What Cannot be Done?

- The No-Cloning Principle:

  There is no physical process that leads to an evolution

  $$|\phi\rangle \otimes |s\rangle \longrightarrow |\phi\rangle \otimes |\phi\rangle$$

  where $|\phi\rangle$ is an arbitrary state and $|s\rangle$ is a fixed state.

- The No-Broadcasting Principle – generalization of no-cloning.

- The No-Deleting Principle:

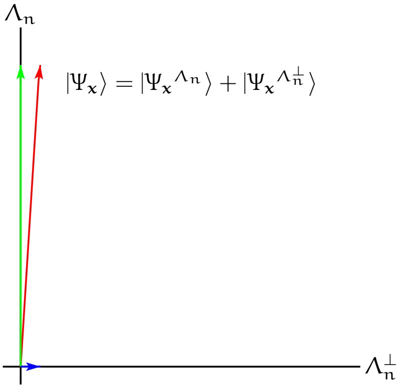
  There is no physical process that leads to an evolution

  $$|\phi\rangle \otimes |\phi\rangle \longrightarrow |\phi\rangle \otimes |s\rangle$$

  where $|\phi\rangle$ is an arbitrary state and $|s\rangle$ is a fixed state.

Typical Subspace

- Typical states $|\Psi_x\rangle \in \mathcal{H}^{2^n}$ "live" in the typical subspace Λ_n.



- The dimension of Λ_n is approximately $2^{nS(\rho)}$.

The Typical Subspace Λ_n

- We represent the source density matrix

$$\rho = \sum_{a \in \mathcal{X}} P(x)|\psi_x\rangle\langle\psi_x|$$

in terms of its eigenvectors and eigenvalues:

$$\rho = \lambda_0|\varphi_0\rangle\langle\varphi_0| + \lambda_1|\varphi_1\rangle\langle\varphi_1|.$$

- Note that $\lambda = \{\lambda_0, \lambda_1\}$ is a PD on $\{0, 1\}$ and $\langle\varphi_0|\varphi_1\rangle = 0$.
- T^n_λ denotes the set of λ-typical sequences.
- Λ_n is the subspace spanned by $|\Phi_z\rangle$, $z \in T^n_\lambda$.

Compression by Measurement

- Measurement is defined by $\Pi + \Pi^{\perp} = I_{2^n}$ where
  - $\Pi = \sum_{z \in T_\lambda^n} |\Phi_z\rangle\langle\Phi_z|$ is the projector to Λ_n.
  - $\Pi^{\perp} = \sum_{z \in \{0,1\}^n \setminus T_\lambda^n} |\Phi_z\rangle\langle\Phi_z|$ is the projector to $\Lambda_n^{\perp}$.
- State after measurement:
  - $\Pi \cdot |\Psi_x\rangle$ with probability $|\langle\Psi_x|\Pi|\Psi_x\rangle|^2$
  - $\Pi^{\perp} \cdot |\Psi_x\rangle$ with probability $|\langle\Psi_x|\Pi^{\perp}|\Psi_x\rangle|^2$

Compression by Measurement

- Measurement is defined by $\Pi + \Pi^{\perp} = I_{2^n}$ where
  - $\Pi = \sum_{z \in T_\lambda^n} |\Phi_z\rangle\langle\Phi_z|$ is the projector to Λ_n.
  - $\Pi^{\perp} = \sum_{z \in \{0,1\}^n \setminus T_\lambda^n} |\Phi_z\rangle\langle\Phi_z|$ is the projector to $\Lambda_n^{\perp}$.
- State after measurement:
  - $\Pi \cdot |\Psi_x\rangle$ with probability $|\langle\Psi_x|\Pi|\Psi_x\rangle|^2$
  - $\Pi^{\perp} \cdot |\Psi_x\rangle$ with probability $|\langle\Psi_x|\Pi^{\perp}|\Psi_x\rangle|^2$
- Expected probability of outcome $\Pi \cdot |\Psi_x\rangle$:

$$
\begin{aligned}
\sum_{x \in \mathcal{X}^n} P(x)|\langle\Psi_x|\Pi|\Psi_x\rangle|^2 \geqslant & -1 + 2\,\mathrm{Tr}(\Pi\rho^{\otimes n}) \\
= & -1 + 2\,\mathrm{Tr}\Big\{\Big[\sum_{z \in T_\lambda^n} |\Phi_z\rangle\langle\Phi_z|\Big] \cdot \Big[\sum_{z \in \{0,1\}^n} \lambda(z)|\Phi_z\rangle\langle\Phi_z|\Big]\Big\} \\
= & 1 - 2\epsilon_n
\end{aligned}
$$

Sources of Mixed Quantum States

- To a source letter $x \in \mathcal{X}$ corresponds quantum state $|\psi_y\rangle$, $y \in \mathcal{Y}$, with probability $W(y|x)$.
- Note that outputs are distributed as

$$Q(y) = \sum_{x \in \mathcal{X}} P(x)W(y|x).$$

Sources of Mixed Quantum States

- To a source letter $x \in \mathcal{X}$ corresponds quantum state $|\psi_y\rangle$, $y \in \mathcal{Y}$, with probability $W(y|x)$.
- Note that outputs are distributed as

$$Q(y) = \sum_{x \in \mathcal{X}} P(x) W(y|x).$$

- The density matrix corresponding to x is

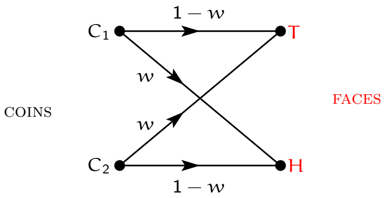
$$\rho_x = \sum_{b \in \mathcal{Y}} W(y|x) |\psi_y\rangle\langle\psi_y|, \;\; x \in \mathcal{X}.$$

- Compression is asymptotically lossless when

$$\sum_{\boldsymbol{x} \in \mathcal{X}^n} P(\boldsymbol{x}) F(\boldsymbol{\rho}_{\boldsymbol{x}}, \hat{\boldsymbol{\rho}}_{\boldsymbol{x}}) \to 1 \text{ as } n \to \infty.$$

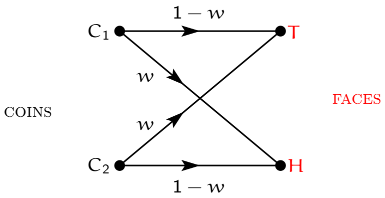
Sources of Mixed Quantum States

- Produce sequences of sources, *e.g.*, coins:



Sources of Mixed Quantum States

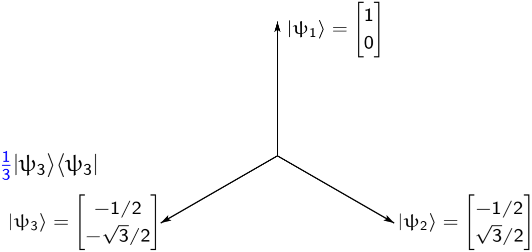
- Produce sequences of sources, *e.g.*, coins:



- A quantum example:

$$\rho_1 = \frac{2}{3}|\psi_1\rangle\langle\psi_1| + \frac{1}{3}|\psi_2\rangle\langle\psi_2|$$
$$\rho_2 = \frac{1}{3}|\psi_2\rangle\langle\psi_2| + \frac{2}{3}|\psi_3\rangle\langle\psi_3|$$

$$\begin{aligned}\rho &= \frac{1}{2}\rho_1 + \frac{1}{2}\rho_2 \\ &= \frac{1}{3}|\psi_1\rangle\langle\psi_1| + \frac{1}{3}|\psi_2\rangle\langle\psi_2| + \frac{1}{3}|\psi_3\rangle\langle\psi_3| \\ &= \frac{1}{2}I\end{aligned}$$



Distances Between Density Matrices ρ and σ

- Uhlman fidelity:

$$F(\sigma, \omega) = \left\{ \mathrm{Tr}\left[(\sqrt{\sigma}\omega\sqrt{\sigma})^{1/2} \right] \right\}^2.$$

Distances Between Density Matrices ρ and σ

- Uhlman fidelity:

$$F(\sigma, \omega) = \left\{ \mathrm{Tr}\left[(\sqrt{\sigma}\omega\sqrt{\sigma})^{1/2} \right] \right\}^2.$$

- Trace distance:

$$D(\sigma, \omega) = \frac{1}{2}\,\mathrm{Tr}\,|\sigma - \omega|,$$

$|A|$ denotes the positive square root of $A^\dagger A$.

- $1 - F(\sigma, \omega) \leqslant D(\sigma, \omega) \leqslant \sqrt{1 - F(\sigma, \omega)^2}$

Distances Between Density Matrices ρ and σ

- Uhlman fidelity:

$$F(\sigma, \omega) = \left\{ \mathrm{Tr}\left[(\sqrt{\sigma}\omega\sqrt{\sigma})^{1/2} \right] \right\}^2.$$

- Trace distance:

$$D(\sigma, \omega) = \frac{1}{2}\,\mathrm{Tr}\,|\sigma - \omega|,$$

  $|A|$ denotes the positive square root of $A^\dagger A$.

- $1 - F(\sigma, \omega) \leqslant D(\sigma, \omega) \leqslant \sqrt{1 - F(\sigma, \omega)^2}$

- Frobenius (Hilbert-Schmidt)?

Distances Between PD's – An Example

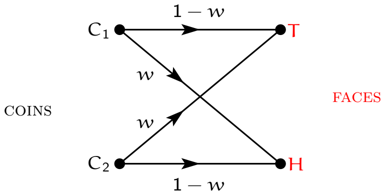
- $\mathcal{A}_N = \{a_1, \ldots, a_N\}$
- $N = 2^K$ and $n = 2^k$, with $k/K = c < 1$.
- Distributions P and Q:

$$P(a_i) = \begin{cases} 1/n, & 1 \leqslant i \leqslant n \\ 0 & n+1 \leqslant i \leqslant N \end{cases} \quad \text{and } Q(a_i) = \frac{1}{N}$$

- $Q(\{a_{n+1}, \ldots, a_N\}) \to 1$ as $k, K \to \infty$.
- $\frac{1}{2}\sum_i |P(a_i) - Q(a_i)| \to 1$ and $\sum_i |P(a_i) - Q(a_i)|^2 \to 0$.

Sources of Sources

- Produce sequences of sources, *e.g.*, coins:



- Example: $P(C_1) = P(C_2) = 1/2$ and $w = 2/3$.
- Sequences:

C_1	C_1	C_2	C_1	C_2	C_2
H	H	H	T	T	T
H	T	H	H	T	T
H	T	T	H	T	H

$\vdots$

Sequences of Sources

- For each source (coin) x in $\mathcal{X}$,
  we have a probability distribution $W(\cdot|x)$ over $\mathcal{Y}$ of letters (faces).
- For each sequence of coins $\boldsymbol{x}$ in $\mathcal{X}^n$,
  we have a probability distribution $W^n(\cdot|\boldsymbol{x})$ over $\mathcal{Y}^n$.
- Alice has coins $\boldsymbol{x}$ and sends N bits to Bob.
- Bob prepares faces $\boldsymbol{y}$ with probability $\widehat{W}^n(\boldsymbol{y}|\boldsymbol{x})$.
- $W^n(\cdot|\boldsymbol{x})$ is reproduced as $\widehat{W}^n(\cdot|\boldsymbol{x})$.

Compression by Sending Classical Information

- Fidelity between $W^n(\cdot|\mathbf{x})$ and $\widehat{W}^n(\cdot|\mathbf{x})$

$$F\left(W^n(\cdot|\mathbf{x}), \widehat{W}^n(\cdot|\mathbf{x})\right) = \sum_{\mathbf{y} \in \mathcal{Y}^n} \sqrt{W^n(\mathbf{y}|\mathbf{x}) \cdot \widehat{W}^n(\mathbf{y}|\mathbf{x})}$$

  is known as the Bhattacharyya-Wooters overlap.

- Compression fidelity

$$\sum_{\mathbf{x} \in \mathcal{X}^n} P(\mathbf{x}) F\left(W^n(\cdot|\mathbf{x}), \widehat{W}^n(\cdot|\mathbf{x})\right)$$

  should approach 1 as $n \to \infty$.

Compression Algorithm for a Typical x

- Alice and Bob
  - have identical random number generators
  - which they use to form a list of N_l typical ys.
- If $N_l > 2^{nI(P,W)}$, then with high probability there will be at least one y on the list which is conditionally typical with respect to x.
- Alice sends $\log N_l$ bits to Bob identifying y.
- Compression rate $\log N_l / n$ approaches $I(P, W)$.

Proof Idea

- $W^n(\cdot|x)$ is roughly a uniform distribution over ys that are conditionally typical given x.
- For a typical x, there are about $2^{nH(W/P)}$ such ys.
- These ys are typical.
- There are about $2^{nH(Q)}$ typical ys.
- A randomly chosen y will be conditionally typical with respect to any typical x with probability of about

$$\frac{2^{nH(W/P)}}{2^{nH(Q)}} = \frac{1}{2^{nI(P,W)}}.$$

A Related Problem

For each Alice's sequence $C_{\mathbf{x}}$ of coins,

Bob prepares a predetermined sequence $\mathbf{y}(\mathbf{x})$ of faces such that

$$\bar{F} = \sum_{\mathbf{x} \in \mathcal{X}^n} P(\mathbf{x}) F_{\mathcal{X}\times\mathcal{Y}}\left(P_{\mathbf{x}} W(\cdot|\cdot), P_{\mathbf{x},\mathbf{y}(\mathbf{x})}\right)$$

$$\begin{aligned} F_{\mathcal{X}\times\mathcal{Y}}\left(P_{\mathbf{x}} W(\cdot|\cdot), P_{\mathbf{x},\mathbf{y}(\mathbf{x})}\right) &= \left[\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \sqrt{P_{\mathbf{x}}(x) W(y|x) \cdot P_{\mathbf{x},\mathbf{y}(\mathbf{x})}(x,y)}\right]^2 \\ &= \left[\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \frac{1}{n} \sqrt{N(x|\mathbf{x}) W(y|x) \cdot N(x,y|\mathbf{x},\mathbf{y}(\mathbf{x}))}\right]^2. \end{aligned}$$

How large is Bob's codebook?

The Original Quantum Problem

- The compression rate can not go below the Holevo quantity χ:

$$\chi = \underbrace{-\operatorname{Tr}\rho\log\rho}_{S(\rho)} - \sum_{a\in\mathcal{X}} P(a)S(\rho_a)$$

The Original Quantum Problem

- The compression rate can not go below the Holevo quantity χ:

$$\chi = \underbrace{-\operatorname{Tr}\rho\log\rho}_{S(\rho)} - \sum_{a\in\mathcal{X}} P(a)S(\rho_a)$$

- $I(P, W)$ is achievable by sending classical information.
  The proof uses
  1. the equivalence of the Uhlman fidelity and the trace distance
  2. the strong convexity of the trace distance:

$$D\Big(\sum_i p_i\omega_i, \sum_i q_i\sigma_i\Big) \leqslant D(\{p_i\},\{q_i\}) + \sum_i p_i D(\omega_i, \sigma_i).$$

  3. the method of types

The Original Quantum Problem

- The compression rate can not go below the Holevo quantity χ:

$$\chi = \underbrace{-\operatorname{Tr} \rho \log \rho}_{S(\rho)} - \sum_{a \in \mathcal{X}} P(a) S(\rho_a)$$

- $I(P, W)$ is achievable by sending classical information.

  The proof uses

  1. the equivalence of the Uhlman fidelity and the trace distance
  2. the strong convexity of the trace distance:

$$D\Big(\sum_i p_i \omega_i, \sum_i q_i \sigma_i\Big) \leqslant D(\{p_i\}, \{q_i\}) + \sum_i p_i D(\omega_i, \sigma_i).$$

  3. the method of types

- Can the gap be closed by qubits or bits?