Holevo Bound

Quantum Foundations · Module 7 — Quantum Information

Node 7.5

Position in Knowledge Graph

System: Quantum Foundations **Structural Domain:** Quantum Information **Node:** 7.5

The **Holevo bound** — Alexander Holevo (1973) — is the fundamental upper bound on the amount of classical information that can be extracted from an ensemble of quantum states. It says: no matter how cleverly you measure, you cannot recover more than χ classical bits of information per quantum state, where χ is a specific entropic quantity associated with the ensemble. In particular, n qubits can carry *at most* n bits of classical information (when no prior entanglement is shared between sender and receiver).

The Holevo bound is the structural reason the exponentially large dimensionality of n-qubit state space does not translate directly into exponentially large classical information capacity. Quantum states have many amplitudes, but reading them out is constrained by the Born rule, and the Holevo bound makes that constraint sharp. It is the key to understanding why quantum advantage appears in *computation* (where interference between amplitudes matters) but not in *classical communication* beyond the bit-per-qubit rate unless entanglement assistance is available (which unlocks superdense coding's factor-of-2 speedup).

Formal Statement

The theorem

Let $\{(p_i, \rho_i)\}_{i=1}^n$ be an ensemble of quantum states: the sender Alice prepares state ρ_i with probability p_i and sends it to Bob. Let $\bar{\rho} = \sum_i p_i \rho_i$ be the average state of the ensemble. Bob performs some POVM measurement $\{E_j\}_{j=1}^m$ and obtains outcome j with probability $q_{j|i} = \mathrm{Tr}(E_j \rho_i)$.

Theorem (Holevo 1973). The classical mutual information between the input label I and the measurement outcome J satisfies

$$I(I:J) \leq \chi(\{p_i, \rho_i\}) := S(\bar{\rho}) - \sum_i p_i S(\rho_i),$$

where S is the von Neumann entropy. The right-hand side χ is called the **Holevo χ-quantity** of the ensemble.

Immediate corollaries

- **Bit-per-qubit bound.** If each ρ_i lives in a Hilbert space of dimension d, then $S(\bar{\rho}) \leq \log d$ and $\sum_i p_i S(\rho_i) \geq 0$, so $\chi \leq \log d$. For n qubits, this gives $\chi \leq n$ bits. No measurement strategy can extract more than n bits of classical information from n qubits.
- **Pure-state ensembles.** If all $\rho_i = |\psi_i\rangle\langle\psi_i|$ are pure, then $S(\rho_i) = 0$ and $\chi = S(\bar{\rho})$. The accessible information is bounded by the entropy of the average state.
- **Orthogonal states saturate.** If the $\{|\psi_i\rangle\}$ are mutually orthogonal, $S(\bar{\rho}) = H(\{p_i\})$ (the classical Shannon entropy), and the bound is saturated by measuring in the basis $\{|\psi_i\rangle\}$. Orthogonal states can be perfectly distinguished, and the Holevo bound reduces to the classical bound.

- **Non-orthogonal states have a gap.** For non-orthogonal pure states, $\chi < H(\{p_i\})$ strictly, and no measurement can achieve the classical source entropy. This is why non-orthogonal states cannot be perfectly distinguished — a restatement of the no-cloning theorem.

Achievability: HSW theorem

Holevo–Schumacher–Westmoreland (1996–1998): the χ-quantity is not just an upper bound on accessible information for a *single* use of a channel, but the *achievable* classical capacity of a quantum channel asymptotically. Explicitly, for a channel $\mathcal{N}$, its classical capacity is

$$C(\mathcal{N}) = \sup_{\{p_i,\rho_i\}} \chi\big(\{p_i, \mathcal{N}(\rho_i)\}\big)$$

(regularized over multiple channel uses for non-additive cases). The Holevo bound is thus tight in the appropriate asymptotic sense.

Intuition

An n-qubit state has 2^n complex amplitudes, so naively it contains $O(2^n)$ real numbers of information. The Holevo bound says this naive counting is wrong: you can put $O(2^n)$ amplitudes into n qubits, but you cannot *read them out* — any measurement returns at most n classical bits on average.

The reason is structural. Measurement is a channel from quantum states to classical outcomes, and channels are subject to the data-processing inequality. Once you commit to a measurement strategy, the information flow is bounded by the Holevo quantity, which is itself bounded by $\log d$ in the dimension of the state space. The exponential amplitude structure is only accessible through interference — that is, through operations that act on superpositions before measurement — and even then, the final measurement is limited by the same bound.

This is why quantum advantage in computation is *subtle*. Shor's algorithm does not "read out" the exponentially many amplitudes of its quantum register; it arranges the amplitudes to interfere constructively at the correct answer and destructively elsewhere, so a single polynomial-sized measurement suffices to extract the answer. Grover's algorithm amplifies the amplitude of the target state. Neither violates the Holevo bound — both are consistent with it, because the classical information extracted from the final measurement is polynomially bounded.

By contrast, the Holevo bound makes quantum *memory* — using quantum states to store classical data for later retrieval — no better than classical memory, up to constant factors. A quantum hard drive is not an exponential improvement over a classical one (in terms of bits stored per physical system), even though the Hilbert space dimension is exponentially large.

The factor-of-2 loophole — superdense coding — exists because shared prior entanglement changes the counting: the channel is now "1 qubit + half a Bell pair," and this combined resource can transmit 2 classical bits. Holevo's bound applies to the single qubit *alone* without the prior entanglement. The bound is sharp.

Structural Role

- **Separating computational from communication advantage.** Quantum computation achieves exponential speedups over classical computation for certain problems (factoring, simulating physics), but quantum communication achieves only modest speedups for classical-information transmission (factor-of-2 via superdense coding). The Holevo bound is the structural reason for this asymmetry.
- **Channel capacity foundation.** The classical capacity of a quantum channel is expressed in terms of Holevo quantities. The Holevo–Schumacher–Westmoreland theorem is the quantum analogue of

Shannon's noisy-channel coding theorem: it gives the asymptotic rate at which classical information can be sent through a quantum channel with vanishing error.

- **Cryptographic bounds.** In QKD security proofs, the adversary's knowledge of the key is bounded by a Holevo quantity over the purifications of the key ensemble. Asymptotic security reduces to estimating this quantity.
- **Memory and storage.** Quantum random access codes (QRACs) allow encoding n bits into fewer than n qubits with some decoding error. Holevo's bound gives the tight rate at which this trade-off can be made.
- **Information-theoretic identity for mixed ensembles.** The Holevo quantity χ can be rewritten as a quantum mutual information on a classical-quantum state $\rho_{XQ} = \sum_i p_i |i\rangle\langle i|_X \otimes \rho_i^Q$: then $\chi = I(X:Q)_\rho$. This identification is the key to understanding why the Holevo bound is a data-processing inequality in disguise.

Mathematical Development

Proof via data processing inequality

Encode the ensemble as a classical-quantum state

$$\rho_{XQ} = \sum_i p_i |i\rangle\langle i|_X \otimes \rho_i^Q.$$

Compute the quantum mutual information:

$$I(X:Q)_\rho = S(X) + S(Q) - S(XQ) = H(\{p_i\}) + S(\bar{\rho}) - \left(H(\{p_i\}) + \sum_i p_i S(\rho_i) \right),$$

where the last equality uses the block-diagonal structure of ρ_{XQ} to get $S(XQ) = H(\{p_i\}) + \sum_i p_i S(\rho_i)$. So

$$I(X:Q)_\rho = S(\bar{\rho}) - \sum_i p_i S(\rho_i) = \chi.$$

The measurement channel $\mathcal{M} : Q \to J$ takes the quantum system to a classical measurement outcome. After measurement, the joint state becomes a classical-classical state ρ_{XJ} on a copy of X and the outcome J. The data processing inequality gives

$$I(X:J) \leq I(X:Q) = \chi.$$

But $I(X:J) = I(I:J)$ is exactly the classical mutual information between input label and measurement outcome. So $I(I:J) \leq \chi$. ■

The proof is essentially one application of the data-processing inequality applied to a classical-quantum state — a clean structural argument, the economy of which reflects the underlying simplicity of the bound.

Tightness via HSW

The HSW theorem shows the bound is asymptotically tight. Given an ensemble $\{(p_i, \rho_i)\}$, one can construct codes using n copies of the channel that transmit $n\chi$ bits of classical information with arbitrarily small error as $n \to \infty$. The proof uses a random coding argument analogous to Shannon's original proof of the classical noisy-channel coding theorem, but with classical codes replaced by quantum ones.

Accessible information

The **accessible information** of an ensemble is

$$I_{\text{acc}}(\{p_i, \rho_i\}) = \sup_{\text{POVMs } \{E_j\}} I(I:J),$$

i.e., the supremum of classical mutual information over all measurement strategies. The Holevo bound says $I_{\text{acc}} \leq \chi$. In general, the gap can be strict — for some ensembles of non-orthogonal states, no single-shot measurement achieves χ, and the bound is achievable only in the asymptotic limit of block coding.

Example: two non-orthogonal states

Consider two non-orthogonal qubit states $|\psi_0\rangle = |0\rangle$ and $|\psi_1\rangle = \cos\theta|0\rangle + \sin\theta|1\rangle$, with equal prior probabilities $p_0 = p_1 = 1/2$.

The average state is

$$\bar{\rho} = \frac{1}{2}|\psi_0\rangle\langle\psi_0| + \frac{1}{2}|\psi_1\rangle\langle\psi_1| = \begin{pmatrix} \frac{1+\cos^2\theta}{2} & \frac{\cos\theta\sin\theta}{2} \\ \frac{\cos\theta\sin\theta}{2} & \frac{\sin^2\theta}{2} \end{pmatrix}.$$

Its eigenvalues are $\lambda_\pm = (1 \pm |\cos\theta|)/2$, and

$$\chi = S(\bar{\rho}) = h\left(\frac{1+|\cos\theta|}{2}\right),$$

where h is the binary entropy. For $\theta = \pi/2$ (orthogonal states), $\chi = 1$ bit — full bit of information. For $\theta = 0$ (identical states), $\chi = 0$ — no information possible. For intermediate θ, $\chi < 1$ strictly, and no measurement can recover the full 1-bit classical source.

The optimal single-shot measurement (the Helstrom measurement) achieves an error probability $P_e = (1 - \sin\theta)/2$, and the corresponding mutual information is $1 - h(P_e)$, which is strictly less than χ for $0 < \theta < \pi/2$. Asymptotically, however, block coding over many copies of the channel achieves χ in the HSW sense.

Worked Example

n qubits cannot store more than n bits

Suppose Alice wants to encode m classical bits into n qubits. By Holevo, the accessible information is at most $\chi \leq n$ bits. So if $m > n$, Bob cannot reliably recover all m bits from a measurement of the n qubits. In the asymptotic limit of many repetitions, the rate m/n is bounded by 1, so quantum storage has the same bit-per-qubit density as classical storage. No exponential quantum storage advantage exists.

Superdense coding — why it doesn't violate Holevo

In superdense coding (Node 7.6), Alice and Bob share a Bell pair. Alice applies one of four Pauli operations $\{I, X, Y, Z\}$ to her qubit, then sends it to Bob. Bob performs a Bell measurement on both qubits and recovers 2 classical bits.

Does this violate Holevo? No — because the shared entanglement is additional resources beyond the single qubit. The relevant ensemble is on *two* qubits (Alice's encoded qubit + Bob's retained Bell-pair half), and the effective dimension is 4, giving a Holevo bound of 2 bits, which is exactly what superdense coding achieves.

Without prior entanglement, a single qubit sent from Alice to Bob has $\chi \leq 1$ and cannot transmit more than 1 bit. The shared Bell pair is the "loophole" — it changes the counting. This is consistent with the Holevo bound being a statement about an ensemble of states, and the ensemble in superdense coding lives in a 4-dimensional space.

Quantum random access codes

A $(2 \to 1)$-QRAC encodes 2 classical bits into 1 qubit such that either bit can be recovered with probability $\geq p$ (but not both simultaneously with certainty). The optimal QRAC achieves $p = \cos^2(\pi/8) \approx 0.854$ per bit.

Holevo's bound is consistent with this: while the accessible information is bounded by $\chi \leq 1$ bit, the QRAC does not violate the bound because the "information" recovered per measurement is not a full bit. The mutual information per recovered bit is $1 - h(0.854) \approx 0.4$, and the total mutual information across the two bits is also bounded by 1. The QRAC distributes the 1 bit of accessible information across the two target bits.

BB84 key rate

In BB84 QKD, Alice sends qubits in 4 possible states (2 bases $\times$ 2 outcomes). Holevo bounds Eve's maximum knowledge of the key given her access to the quantum channel. In the collective-attack model, the secure key rate is bounded by

$$r \geq I(\text{Alice} : \text{Bob}) - \chi(\text{Alice} : \text{Eve}),$$

where χ is the Holevo quantity over Eve's ensemble. Asymptotic security proofs reduce to estimating χ from the observed error rate.

Cross-Links

- **Module 7.1 (Qubits):** Justifies the last row of the qubit-vs-classical-register comparison table: "n qubits carry n bits of classical information."
- **Module 7.2 (No-cloning):** Holevo and no-cloning are related — both express the inaccessibility of non-orthogonal quantum state information.
- **Module 7.3 (Quantum channels):** The proof uses the data-processing inequality for classical-quantum states under CPTP maps (measurement is a channel).
- **Module 7.4 (Von Neumann entropy):** χ is built from entropies. The Holevo bound is a corollary of the data-processing inequality for quantum mutual information.
- **Module 7.6 (Teleportation and superdense coding):** Superdense coding achieves the Holevo bound on 2-qubit ensembles (assisted by prior entanglement); teleportation transmits a qubit using 2 classical bits and shared entanglement.
- **Module 10.5 (Quantum resource theories):** Holevo quantity is a monotone under free operations in the resource theory of classical communication.
- **Statistics / Shannon theory:** Classical capacity theorems generalize Shannon's noisy-channel coding theorem.

Structural Compression

Invariant: No measurement can extract more than $\chi = S(\bar{\rho}) - \sum_i p_i S(\rho_i)$ bits of classical information from a quantum ensemble $\{(p_i, \rho_i)\}$. In particular, n qubits carry at most n bits of classical information, regardless of encoding.

Mechanism: Application of the data-processing inequality to the classical-quantum state $\rho_{XQ} = \sum_i p_i |i\rangle\langle i| \otimes \rho_i$ under the measurement channel. χ is exactly the quantum mutual information $I(X:Q)_\rho$ of this state, and measurement is a CPTP map from Q to a classical outcome register.

Cross-System Echo: Shannon's channel capacity theorem bounds the classical rate through a classical channel; Holevo's bound is the quantum generalization. The HSW theorem closes the analogy by showing that the Holevo quantity is asymptotically achievable, making it the quantum Shannon capacity for classical communication.

Exercises

1. Verify that for an ensemble of pure states $\{(p_i, |\psi_i\rangle\langle\psi_i|)\}$, the Holevo quantity reduces to $\chi = S(\bar{\rho})$.
2. Compute the Holevo quantity for two non-orthogonal qubit states at angle θ (as in the worked example) and verify that it is less than 1 bit for $0 < \theta < \pi/2$.
3. Show that the Holevo bound reduces to the classical source-coding bound when the states $\{|\psi_i\rangle\}$ are mutually orthogonal.
4. Prove the Holevo bound via the data-processing inequality applied to the classical-quantum state ρ_{XQ} under a measurement channel.
5. Explain why superdense coding does not violate the Holevo bound. What is the correct ensemble to apply the bound to?
6. Describe how the Holevo bound constrains the security of BB84 quantum key distribution. What is Eve's role in the relevant ensemble?
7. Argue, structurally, why the Holevo bound is the reason quantum advantage appears in computation but not in classical communication (without entanglement assistance). What feature of quantum computation allows it to evade the bound's implications?

Quantum Foundations · Module 7 — Quantum Information · Node 7.5
Concepts: von-neumann-entropy, channel-capacity, mutual-information, information-theory
Prerequisites: 7.4

hbar.university — own.your.cognition