Module 7 — Quantum Information

Quantum Foundations

hbar.university

Qubits and Quantum Registers

Position in Knowledge Graph

System: Quantum Foundations **Structural Domain:** Quantum Information **Node:** 7.1

This is the foundation node of the quantum information module. The qubit is the simplest quantum system — and the universal building block for quantum computation, communication, and cryptography. Every protocol in the rest of this module is built on registers of qubits.

Formal Definition

A **qubit** is a quantum system with state space $\mathcal{H} = \mathbb{C}^2$, equipped with a distinguished orthonormal basis $\{|0\rangle, |1\rangle\}$ called the **computational basis**.

A general qubit state is

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \qquad \alpha, \beta \in \mathbb{C}, \qquad |\alpha|^2 + |\beta|^2 = 1.$$

A **quantum register** of n qubits is a system with state space

$$\mathcal{H}^{\otimes n} = (\mathbb{C}^2)^{\otimes n} = \mathbb{C}^{2^n},$$

with computational basis

$$\{|x\rangle : x \in \{0,1\}^n\}$$

where $|x\rangle$ is shorthand for $|x_1\rangle \otimes |x_2\rangle \otimes \cdots \otimes |x_n\rangle$.

A general n-qubit state is

$$|\psi\rangle = \sum_{x \in \{0,1\}^n} c_x |x\rangle, \qquad \sum_x |c_x|^2 = 1,$$

with 2^n complex amplitudes (and $2^n - 1$ independent real parameters after normalization and global phase).

Intuition

A qubit is two classical bits' worth of "answers" entangled with one bit's worth of "information you can read." It is *not* a continuous classical bit; it is *not* a probabilistic classical bit; it is a structurally distinct object whose reduction-to-classical only happens at the moment of measurement.

The exponential dimensionality of registers — 2^n amplitudes for n qubits — is the structural source of quantum computational resources. A 50-qubit register has more amplitudes than there are atoms in the observable universe; this is the structural fact that quantum simulation exploits.

Structural Role

This node is the operational entry point to all quantum information protocols:

- No-cloning theorem (7.2) is a constraint on what can be done to qubit registers.
- Quantum channels (7.3) are the dynamics on density matrices on $\mathcal{H}^{\otimes n}$.
- Entropy and mutual information (7.4) measure resources on quantum registers.
- Teleportation and superdense coding (7.6) are protocols on small registers (2-3 qubits).

It is also the entry point to VQA: parameterized quantum circuits act on registers of this form.

Mathematical Development

Single-qubit gates

Single-qubit operations are unitaries on $\mathbb{C}^2$ — equivalently, elements of $SU(2)$ (up to phase). The Pauli operators X, Y, Z generate the Pauli group; the Hadamard H, phase S, and T gates generate the Clifford+T set, which is universal for single-qubit unitaries (approximately).

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \qquad S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}.$$

Two-qubit gates

The CNOT (controlled-NOT) gate is the canonical two-qubit entangling gate:

$$\text{CNOT} = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes X.$$

The set $\{H, S, \text{CNOT}\}$ plus any non-Clifford single-qubit gate (e.g., T) is **universal** for quantum computation: any unitary on n qubits can be approximated to arbitrary precision by a circuit of these gates.

Quantum registers vs classical registers

Property	Classical n-bit register	Quantum n-qubit register
State space	$\{0,1\}^n$, 2^n states	$\mathbb{C}^{2^n}$, $2 \cdot 2^n - 2$ real parameters
Operations	Boolean functions	Unitary operators
Read-out	Deterministic	Probabilistic (Born rule)
Composition	Cartesian product	Tensor product
Information content	n bits	Up to n qubits' worth of entanglement, but only n bits per measurement

The last row is Holevo's theorem (Node 7.5): you cannot extract more than n classical bits from n qubits, despite the exponentially larger state space.

Worked Example

Creating a Bell state

Starting from $|00\rangle$:

1. Apply H to qubit 1: $H \otimes I|00\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |10\rangle)$.
2. Apply CNOT (control = qubit 1, target = qubit 2): $\text{CNOT}\frac{1}{\sqrt{2}}(|00\rangle + |10\rangle) = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) = |\Phi^+\rangle$.

Two gates, one Bell state. This is the simplest non-trivial quantum protocol — and it is the entry point to teleportation, superdense coding, QKD, and Bell tests.

Cross-Links

- **Module 2:** State postulate (2.1) and tensor products (2.5) are the foundations.
- **Module 4:** Bell states (4.4) are the canonical entangled qubit states.
- **VQA:** All variational circuits are built from gates on qubit registers.
- **Linear Algebra:** The whole machinery of unitary matrices on $\mathbb{C}^{2^n}$.
- **Statistics:** Measurement of n qubits in the computational basis returns a sample from a distribution on $\{0,1\}^n$.

Structural Compression

Invariant: Qubit register dimension scales as 2^n; operations are unitaries; readouts are samples from a Born-rule distribution.

Mechanism: Tensor product of n copies of $\mathbb{C}^2$.

Cross-System Echo: Classical bit registers grow linearly in storage; quantum registers grow exponentially in state space dimension. The gap is the structural source of quantum advantage.

Exercises

1. Show that any single-qubit unitary can be written as $U = e^{i\alpha}R_z(\beta)R_y(\gamma)R_z(\delta)$ (Z-Y-Z decomposition).
2. Verify that the CNOT gate maps $|+\rangle|0\rangle \to \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$.
3. Show that an n-qubit pure state has $2^{n+1} - 2$ real parameters before normalization, and $2 \cdot 2^n - 2$ after global-phase removal.
4. Construct a 3-qubit GHZ state $(|000\rangle + |111\rangle)/\sqrt{2}$ using H and CNOT gates.
5. Argue why the exponential state-space dimension does *not* automatically translate to exponential computational speedup. (Hint: read-out is constrained by Holevo's bound, Node 7.5.)

No-Cloning Theorem

Position in Knowledge Graph

System: Quantum Foundations **Structural Domain:** Quantum Information **Node:** 7.2

The no-cloning theorem — Wootters–Zurek and Dieks, independently in 1982 — says that there is no physical process that can take an arbitrary unknown quantum state as input and produce two identical copies of it as output. Unlike classical bits, which can be freely copied, quantum states are "use-once" resources in a precise structural sense: whatever the state is, the only way to make a second instance of it is to already know it classically (in which case one can prepare the second copy directly).

No-cloning is the structural reason quantum key distribution is secure, the reason quantum error correction must work differently from classical repetition codes, and the reason there is no quantum amplifier that faithfully preserves a single-photon state. It is also one of the cleanest illustrations that quantum mechanics is not just "classical mechanics with extra ingredients" — it is a fundamentally different theory of information whose constraints are inherited directly from the linearity of unitary dynamics.

Formal Statement

The theorem

Theorem (Wootters–Zurek, Dieks 1982). There is no unitary operator U acting on $\mathcal{H} \otimes \mathcal{H}$ and no fixed "blank" state $|0\rangle \in \mathcal{H}$ such that, for every $|\psi\rangle \in \mathcal{H}$,

$$U(|\psi\rangle \otimes |0\rangle) = |\psi\rangle \otimes |\psi\rangle.$$

Proof

Suppose such a U exists. Apply it to two arbitrary states $|\psi\rangle$ and $|\phi\rangle$:

$$U(|\psi\rangle \otimes |0\rangle) = |\psi\rangle \otimes |\psi\rangle,$$
$$U(|\phi\rangle \otimes |0\rangle) = |\phi\rangle \otimes |\phi\rangle.$$

Unitary operators preserve inner products. Taking the inner product of the left-hand sides:

$$\langle\psi| \otimes \langle 0|\, U^\dagger U\, |\phi\rangle \otimes |0\rangle = \langle\psi|\phi\rangle\langle 0|0\rangle = \langle\psi|\phi\rangle.$$

And of the right-hand sides:

$$\big(\langle\psi| \otimes \langle\psi|\big)\big(|\phi\rangle \otimes |\phi\rangle\big) = \langle\psi|\phi\rangle^2.$$

Therefore $\langle\psi|\phi\rangle = \langle\psi|\phi\rangle^2$, which means $\langle\psi|\phi\rangle \in \{0, 1\}$. But $|\psi\rangle$ and $|\phi\rangle$ were arbitrary, and generic pairs of states have inner products strictly between 0 and 1. Contradiction. ■

Corollaries and refinements

- **No approximate universal cloning at perfect fidelity.** Even if one relaxes to imperfect cloning, the best achievable fidelity for universal cloners is bounded below 1 (Buzek–Hillery bound: $F = 5/6$ for optimal $1 \to 2$ qubit cloning).
- **No deletion.** A complementary theorem (Pati–Braunstein 2000): there is no unitary that takes $|\psi\rangle \otimes |\psi\rangle \to |\psi\rangle \otimes |0\rangle$ for arbitrary $|\psi\rangle$. Quantum information cannot be unconditionally deleted either.

- **Orthogonal states can be copied.** The theorem does not forbid cloning a known basis: if one knows $|\psi\rangle \in \{|0\rangle, |1\rangle\}$, a CNOT gate clones it perfectly. The theorem forbids a *single* unitary that clones *every* state.

Intuition

The intuitive story is that to clone a state, one would need to learn enough about it to reproduce it, and learning a quantum state means measuring it — which in general disturbs or destroys the state. But this intuition is only suggestive; the theorem is stronger. It says that no unitary can produce a clone, even if no classical information is extracted. The ban is on the *shape* of the required transformation, not on the information-theoretic resources of the cloner.

The cleanest way to see why is through linearity. Suppose a unitary cloner U exists. Then for any superposition $\alpha|\psi\rangle + \beta|\phi\rangle$, linearity forces

$$U\big((\alpha|\psi\rangle + \beta|\phi\rangle) \otimes |0\rangle\big) = \alpha\,|\psi\rangle|\psi\rangle + \beta\,|\phi\rangle|\phi\rangle.$$

But "two copies of the superposition" would require

$$(\alpha|\psi\rangle + \beta|\phi\rangle) \otimes (\alpha|\psi\rangle + \beta|\phi\rangle) = \alpha^2|\psi\rangle|\psi\rangle + \alpha\beta(|\psi\rangle|\phi\rangle + |\phi\rangle|\psi\rangle) + \beta^2|\phi\rangle|\phi\rangle.$$

These are manifestly different unless the cross-terms vanish, which happens only if $\alpha\beta = 0$ or the states are orthogonal. Cloning is incompatible with quantum superposition because it is a nonlinear operation on the state, and quantum dynamics is linear.

The philosophical upshot: classical information is "free" — it can be copied, broadcast, and distributed — because classical states are distinguishable pointers to definite facts. Quantum information is "private" — a state cannot be broadcast to multiple recipients without destroying its coherence — because quantum states are not pointers but the actual objects of physical interest, and linearity forbids the operation that would distribute them.

Structural Role

- **Cryptography foundation.** No-cloning is the structural reason BB84 and other QKD protocols are secure. An eavesdropper cannot tap a quantum channel without disturbing the states — and any disturbance is detectable by the legitimate users. Section 9.3 of Nielsen–Chuang develops this chain of reasoning in full.
- **Constraint on error correction.** Classical error correction uses repetition codes: to protect a bit, make three copies and majority-vote. Quantum error correction cannot use this directly because making copies is forbidden. Instead, quantum codes like the Shor code and Steane code distribute information *across* qubits via entanglement, in a way that does not require cloning.
- **No quantum signal amplification.** A classical amplifier boosts a weak signal by creating many high-energy copies. A quantum amplifier cannot do this for a single-photon state — any attempt to amplify introduces noise (this is the basis of the quantum limits on amplifier noise in nonlinear optics, and forms one pillar of the theory of quantum channels).
- **Contrast with classical channels.** A broadcast channel in classical information theory distributes one input to multiple receivers. The quantum analogue (the "quantum broadcast channel") is fundamentally different: one cannot split a quantum state into two independent copies, only into an entangled pair whose marginals have reduced purity.
- **Foundational significance.** No-cloning is one of a family of no-go theorems (no-cloning, no-deleting, no-broadcasting, no-hiding) that together characterize quantum information as a resource distinct from classical information. Each is proved by similar linearity arguments.

Mathematical Development

Proof via inner product preservation

The proof above uses the fact that unitary operators preserve inner products. This is the cleanest proof of no-cloning and generalizes easily. Any linear operator V satisfying $V|\psi\rangle|0\rangle = |\psi\rangle|\psi\rangle$ and $V|\phi\rangle|0\rangle = |\phi\rangle|\phi\rangle$ for two non-orthogonal states is inconsistent, regardless of unitarity. Unitarity is sufficient but not strictly necessary; linearity is the essential property.

Proof via linearity of superposition

For states $|\psi\rangle$ and $|\phi\rangle$ that can be cloned, linearity forces the superposition to map as

$$U\big((\alpha|\psi\rangle + \beta|\phi\rangle) \otimes |0\rangle\big) = \alpha|\psi\rangle|\psi\rangle + \beta|\phi\rangle|\phi\rangle,$$

which is an entangled state, not a product state of the form $(\alpha|\psi\rangle + \beta|\phi\rangle) \otimes (\alpha|\psi\rangle + \beta|\phi\rangle)$. So even if U could clone each of $|\psi\rangle$ and $|\phi\rangle$ individually, it cannot clone any proper superposition of them.

Approximate cloning: the Bužek–Hillery bound

Suppose one relaxes the requirement and asks for a "universal quantum cloner" that produces two output states, each approximately equal to the input, with maximum average fidelity over input states. Bužek and Hillery showed that for a qubit $1 \to 2$ universal cloner, the maximum achievable fidelity is

$$F = \frac{5}{6} \approx 0.833.$$

The optimal cloner produces entangled outputs whose reduced density matrices have this fidelity with the input. For d-dimensional systems, the bound is

$$F_{d,1\to 2} = \frac{1}{2} + \frac{1}{d+1}.$$

This converges to $1/2$ as $d \to \infty$ — high-dimensional states cannot even be copied to better-than-guessing fidelity.

State-dependent cloning

If one is willing to clone only a specific (known) pair of states rather than all states, better fidelity is possible. For two non-orthogonal states $|\psi_0\rangle, |\psi_1\rangle$ with inner product $\langle\psi_0|\psi_1\rangle = s$, the optimal state-dependent cloner has fidelity

$$F = \frac{1}{2}\left(1 + \sqrt{\frac{1-s^2}{1-s^2+s^2/2}}\right).$$

As $s \to 0$ (orthogonal states), $F \to 1$ — orthogonal states can be perfectly cloned. As $s \to 1$ (nearly identical states), $F \to 1/2$ — the cloner is no better than random guessing.

Broadcast generalization

No-broadcasting theorem (Barnum–Caves–Fuchs–Jozsa–Schumacher 1996): for mixed states, "broadcasting" — finding a joint state ρ_{AB} whose marginals are both equal to the input ρ — is impossible unless the input is classical (commuting with itself on some basis). This generalizes no-cloning from pure to mixed states.

Worked Example

Cannot clone $|0\rangle$ and $|+\rangle$

Suppose a hypothetical cloner exists with $U(|0\rangle \otimes |0\rangle) = |0\rangle|0\rangle$ and $U(|+\rangle \otimes |0\rangle) = |+\rangle|+\rangle$. Write out $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$:

$$U\left(\frac{|0\rangle + |1\rangle}{\sqrt{2}} \otimes |0\rangle\right) = \frac{1}{\sqrt{2}}U(|0\rangle|0\rangle) + \frac{1}{\sqrt{2}}U(|1\rangle|0\rangle).$$

For this to equal $|+\rangle|+\rangle = \frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle)$, we would need

$$\frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}U(|1\rangle|0\rangle) = \frac{1}{2}|00\rangle + \frac{1}{2}|01\rangle + \frac{1}{2}|10\rangle + \frac{1}{2}|11\rangle.$$

But the left-hand side has amplitude $1/\sqrt{2}$ on $|00\rangle$, and the right-hand side has amplitude $1/2$. Contradiction. No such U exists.

Equivalently, via the inner product argument: $\langle 0|+\rangle = 1/\sqrt{2}$, so cloning would require $1/\sqrt{2} = (1/\sqrt{2})^2 = 1/2$, which is false. The theorem applies in a concrete, verifiable way.

BB84 eavesdropping detection

In BB84, Alice sends qubits to Bob in randomly chosen bases $\{|0\rangle, |1\rangle\}$ or $\{|+\rangle, |-\rangle\}$. Eve, the eavesdropper, wants to learn the bit and forward a copy to Bob. If Eve could clone, she could copy each qubit, measure her copy, and send Bob an identical copy. No-cloning forbids this.

Eve's only option is to measure each qubit in some basis of her choice and send Bob a freshly prepared state. If Eve measures in the wrong basis (because she doesn't know which basis Alice used), she randomizes the state. Alice and Bob can detect this by comparing a random subset of their bits after the transmission: if the error rate is above the expected noise level, Eve is detected.

The structural chain is: no-cloning → Eve must disturb → disturbance is detectable → QKD is secure (modulo noise model assumptions). This is the cleanest application of no-cloning in a practical protocol.

Quantum money

Wiesner (1970) proposed "quantum money": bills encoded as specific quantum states such that the bank can verify authenticity (by measuring in the right basis) but counterfeiters cannot duplicate the states (no-cloning). The concept predates no-cloning's formal statement and motivated the theorem. Modern quantum money schemes build on this idea but require additional cryptographic assumptions beyond just no-cloning.

Cross-Links

- **Module 2.4 (Unitary evolution):** The proof relies essentially on the linearity of unitary dynamics. No-cloning is a direct consequence of this linearity.

- **Module 7.1 (Qubits):** Qubits are the natural setting for the cleanest version of the theorem and for BB84-style applications.
- **Module 7.3 (Quantum channels):** The no-broadcasting theorem generalizes no-cloning to mixed states and channels.
- **Module 7.6 (Teleportation):** Teleportation moves a quantum state from Alice to Bob *without* violating no-cloning: the original is destroyed in the process. This is the no-cloning-compatible way to transmit quantum information.
- **VQA:** One cannot "checkpoint" an intermediate quantum state in a variational circuit for diagnostic purposes. The circuit must be re-run from the initial state each time.
- **Quantum error correction:** Must use entanglement rather than repetition.

Structural Compression

Invariant: There is no unitary U that maps $|\psi\rangle|0\rangle \to |\psi\rangle|\psi\rangle$ for all $|\psi\rangle$ in a Hilbert space of dimension ≥ 2. Quantum states cannot be copied, only transported (teleportation) or distributed via entanglement.

Mechanism: Unitarity preserves inner products, and the cloning relation forces $\langle\psi|\phi\rangle = \langle\psi|\phi\rangle^2$, which has no solutions for non-trivial superpositions. Equivalently, cloning is a nonlinear operation on the state, incompatible with linear quantum dynamics.

Cross-System Echo: Classical Shannon information is freely copiable — a bit stream can be duplicated and distributed arbitrarily. Quantum information is not, and this gap is the structural source of both quantum cryptography's security and quantum error correction's nontrivial form.

Exercises

1. Prove the no-cloning theorem using the inner-product argument for two distinct non-orthogonal states.
2. Prove the no-cloning theorem using the linearity-of-superposition argument. Where in the calculation does the contradiction appear?
3. Show that orthogonal states can be cloned by exhibiting the unitary (hint: CNOT clones computational-basis states).
4. Compute the Bužek–Hillery fidelity bound 5/6 for $1 \to 2$ qubit cloning explicitly by constructing the optimal cloner.
5. Describe how BB84 uses no-cloning to detect eavesdropping. What assumption about the noise model is required for the protocol to be secure?
6. State the no-deletion theorem and sketch its proof. Why is it a complementary result to no-cloning?
7. Argue, structurally, why no-cloning is a consequence of linearity and not of some deeper "measurement disturbance" principle. What would happen if quantum dynamics were nonlinear?

Quantum Channels and CPTP Maps

Position in Knowledge Graph

System: Quantum Foundations **Structural Domain:** Quantum Information **Node:** 7.3

A **quantum channel** is the most general physically realizable linear operation on quantum states. It generalizes unitary evolution (which describes closed systems) to arbitrary open-system processes that include noise, decoherence, measurement, and discarding of subsystems. The class of quantum channels coincides with the class of **completely positive trace-preserving (CPTP) maps**, and this characterization is one of the central results of quantum information theory.

Three equivalent representations of quantum channels — Stinespring dilation, Kraus operator sum, and the Choi matrix — give complementary views of the same mathematical object. Together they form the standard toolkit for analyzing any quantum process in which the system is not perfectly isolated. Every result in the subsequent modules about decoherence, measurement, noise, or quantum information flow is ultimately a statement about channels.

Formal Definition

CPTP maps

Let $\mathcal{B}(\mathcal{H})$ denote the space of bounded operators on a Hilbert space $\mathcal{H}$. A **quantum channel** is a linear map

$$\mathcal{E} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$$

satisfying two conditions:

1. **Trace preservation (TP):** For every $\rho \in \mathcal{B}(\mathcal{H}_A)$,

   $$\mathrm{Tr}(\mathcal{E}(\rho)) = \mathrm{Tr}(\rho).$$

   This ensures that probabilities sum to 1 after the channel acts.

2. **Complete positivity (CP):** For every ancillary Hilbert space $\mathcal{H}_R$ of any dimension, the extension $\mathcal{E} \otimes \mathrm{id}_R$ maps positive operators on $\mathcal{H}_A \otimes \mathcal{H}_R$ to positive operators on $\mathcal{H}_B \otimes \mathcal{H}_R$:

   $$X \geq 0 \implies (\mathcal{E} \otimes \mathrm{id}_R)(X) \geq 0.$$

Complete positivity is strictly stronger than ordinary positivity (which demands only $\mathcal{E}(\rho) \geq 0$ for $\rho \geq 0$). The standard counterexample is the transpose map $T : \rho \mapsto \rho^T$, which is positive but not completely positive — when tensored with the identity and applied to an entangled state, it can produce negative eigenvalues (this is the basis of the Peres PPT criterion in Node 4.2).

Stinespring dilation

Theorem (Stinespring 1955). A linear map $\mathcal{E} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ is CPTP if and only if there exist an environment Hilbert space $\mathcal{H}_E$, a pure state $|0\rangle_E \in \mathcal{H}_E$, and an isometry $V : \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_E$ such that

$$\mathcal{E}(\rho) = \mathrm{Tr}_E\left(V \rho V^\dagger\right).$$

Equivalently, every channel can be realized as a unitary on a larger system (system + environment) followed by tracing out the environment. The environment can always be chosen with dimension at most $\dim(\mathcal{H}_A) \cdot \dim(\mathcal{H}_B)$.

Kraus representation

Theorem (Kraus 1983). A linear map $\mathcal{E}$ is CPTP if and only if there exist operators $\{K_k\}_{k=1}^{N}$, $K_k : \mathcal{H}_A \to \mathcal{H}_B$, with

$$\sum_k K_k^\dagger K_k = I_A$$

(completeness condition, equivalent to trace preservation), such that

$$\mathcal{E}(\rho) = \sum_k K_k \rho K_k^\dagger.$$

The K_k are called **Kraus operators** (or operation elements). The representation is not unique — different Kraus sets can give the same channel, and the number of Kraus operators can be taken $\leq \dim(\mathcal{H}_A) \cdot \dim(\mathcal{H}_B)$.

Choi matrix

The **Choi matrix** of a channel $\mathcal{E}$ is the operator on $\mathcal{H}_A \otimes \mathcal{H}_B$ obtained by applying $\mathcal{E}$ to one half of a maximally entangled state:

$$C_\mathcal{E} = (\mathrm{id}_A \otimes \mathcal{E})(|\Omega\rangle\langle\Omega|), \qquad |\Omega\rangle = \sum_i |i\rangle_A \otimes |i\rangle_A.$$

By the **Choi–Jamiołkowski isomorphism**, $\mathcal{E}$ is CPTP if and only if $C_\mathcal{E} \geq 0$ and $\mathrm{Tr}_B(C_\mathcal{E}) = I_A$. The Choi matrix provides a finite-dimensional representation of the channel and is the basis for many numerical approaches in quantum information.

Intuition

A quantum channel is "anything physical that takes quantum states in and produces quantum states out." The defining properties are natural:

- Linearity: if ρ is a probabilistic mixture $p_1\rho_1 + p_2\rho_2$, the output must be the same mixture $p_1\mathcal{E}(\rho_1) + p_2\mathcal{E}(\rho_2)$. Otherwise the channel would "know" the decomposition, which is unphysical.
- Trace preservation: probabilities must add up to 1.
- Positivity: outputs must be valid density operators.
- Complete positivity: even if the input system is part of a larger entangled system, the channel acting on just one half must still produce a valid joint state. Without this, channels could produce "negative probabilities" on entangled inputs, which would be unphysical.

The three representations each emphasize a different aspect:

- **Stinespring** says: channels are just unitaries on bigger systems, then throwing away part. This is the physicist's picture — every "noisy" or "lossy" process in the lab is actually a clean unitary interaction with an environment we don't control.
- **Kraus** says: channels are sums of "branches," each weighted by an operator. This is useful for explicit calculations and for thinking of channels as generalized measurements (each Kraus operator corresponds to a possible measurement outcome, if the environment is measured).
- **Choi** says: channels are operators on a doubled Hilbert space. This is the most mathematically compact view and turns questions about channels into questions about fixed density matrices.

The key structural fact — that the CPTP class exactly coincides with physically realizable operations, independent of how one defines "physically realizable" — is a deep theorem. It says: there is no ambiguity

about what quantum operations can be done in principle. The mathematical definition captures physics exactly.

Structural Role

- **Unification of dynamics.** Unitary evolution, measurement (via generalized POVMs), discarding a subsystem (partial trace), noise, decoherence, and quantum error correction are all quantum channels. The structural payoff is that one framework handles all of them uniformly.
- **Foundation for open systems.** Module 8 (Lindblad equation) will introduce the continuous-time generator of a semigroup of channels. The Lindblad equation is to quantum channels what the master equation is to classical Markov processes.
- **Entropic bounds.** The data-processing inequality (Module 7.4) states that mutual information cannot increase under channel action: $I(A;B)_{(\mathrm{id}\otimes\mathcal{E})(\rho)} \leq I(A;B)_\rho$. This is the quantum analogue of the classical fact that noisy channels destroy information, and is central to quantum Shannon theory.
- **Noise models for real hardware.** Variational quantum circuits run on noisy quantum devices are modeled by inserting a noise channel after each gate. The choice of channel (depolarizing, amplitude damping, dephasing, etc.) determines the error model for quantum error correction and fault tolerance.
- **Capacity theorems.** Quantum channels have several inequivalent capacities: classical capacity, quantum capacity, private capacity, entanglement-assisted capacity. These are all rates at which specific resources can be transmitted through the channel, and they form the core of quantum Shannon theory.

Mathematical Development

Equivalence of representations

Start from Stinespring: $\mathcal{E}(\rho) = \mathrm{Tr}_E(V\rho V^\dagger)$. Choose an orthonormal basis $\{|e_k\rangle\}$ of $\mathcal{H}_E$ and define

$$K_k = \langle e_k|_E V : \mathcal{H}_A \to \mathcal{H}_B.$$

Then

$$\mathrm{Tr}_E(V\rho V^\dagger) = \sum_k \langle e_k|V\rho V^\dagger|e_k\rangle = \sum_k K_k \rho K_k^\dagger,$$

and the completeness relation $V^\dagger V = I_A$ translates to $\sum_k K_k^\dagger K_k = I_A$. This gives the Kraus form.

Conversely, given Kraus operators $\{K_k\}$, one can construct an isometry $V|\psi\rangle = \sum_k K_k|\psi\rangle \otimes |e_k\rangle$, which is exactly a Stinespring dilation. So the two representations are equivalent.

The Choi matrix can be read off either: from Kraus form,

$$C_\mathcal{E} = \sum_k (I \otimes K_k)|\Omega\rangle\langle\Omega|(I \otimes K_k^\dagger).$$

Its positivity (since each term is a rank-1 positive operator) reflects complete positivity of $\mathcal{E}$.

Complete positivity versus positivity: the transpose map

Consider the transpose map $T : \rho \mapsto \rho^T$ on a single qubit. T is positive: if $\rho \geq 0$ then $\rho^T \geq 0$ (since they have the same spectrum). But $T \otimes \mathrm{id}$ applied to a Bell state $|\Phi^+\rangle\langle\Phi^+|$ on two qubits gives

$$(T \otimes I)|\Phi^+\rangle\langle\Phi^+| = \frac{1}{2}\sum_{i,j}|i\rangle\langle j|^T \otimes |i\rangle\langle j| = \frac{1}{2}\sum_{i,j}|j\rangle\langle i| \otimes |i\rangle\langle j| = \frac{1}{2}\,\mathrm{SWAP},$$

which has eigenvalues $+1/2, +1/2, +1/2, -1/2$ — a negative eigenvalue. So T is positive but not completely positive, and therefore not a quantum channel.

This is exactly the mathematical obstruction exploited by the PPT (Peres–Horodecki) separability criterion in Node 4.2: applying the partial transpose to an entangled state can produce a non-positive matrix, detecting the entanglement.

Channel composition

If $\mathcal{E}_1 : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ and $\mathcal{E}_2 : \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_C)$ are channels, their composition $\mathcal{E}_2 \circ \mathcal{E}_1$ is also a channel, with Kraus operators given by products:

$$(K_k^{(2)} K_l^{(1)}) \in \{\text{Kraus of } \mathcal{E}_2 \circ \mathcal{E}_1\}.$$

Channels form a monoid under composition with the identity channel.

Unital vs non-unital channels

A channel $\mathcal{E}$ is **unital** if $\mathcal{E}(I) = I$. Equivalently, the maximally mixed state is a fixed point. Examples: unitary channels, Pauli channels, depolarizing channels. Non-unital examples: amplitude damping (which drives toward $|0\rangle\langle 0|$ rather than the maximally mixed state).

Unital channels form a subclass with many nice properties: they are the quantum analogue of doubly-stochastic classical channels (which preserve the uniform distribution) and are the channels for which entropy is non-decreasing.

Worked Example

Depolarizing channel

The depolarizing channel on a qubit with parameter $p \in [0, 1]$ is

$$\mathcal{E}_p(\rho) = (1-p)\rho + p \cdot \frac{I}{2}.$$

It leaves the state unchanged with probability $1 - p$ and replaces it with the maximally mixed state with probability p.

Kraus form: Using the Pauli operators,

$$\mathcal{E}_p(\rho) = \left(1 - \frac{3p}{4}\right)\rho + \frac{p}{4}\big(X\rho X + Y\rho Y + Z\rho Z\big),$$

so the Kraus operators are

$$K_0 = \sqrt{1 - \tfrac{3p}{4}}\, I, \quad K_1 = \sqrt{\tfrac{p}{4}}\, X, \quad K_2 = \sqrt{\tfrac{p}{4}}\, Y, \quad K_3 = \sqrt{\tfrac{p}{4}}\, Z.$$

Check: $\sum_k K_k^\dagger K_k = (1 - 3p/4)I + 3(p/4)I = I$. Trace-preserving.

Action on Bloch vector: $\mathcal{E}_p$ shrinks the Bloch vector by $(1-p)$, mapping the Bloch sphere to a smaller ball of radius $1-p$ centered at the origin. As $p \to 1$, all states collapse to the maximally mixed state. This is the canonical model of "isotropic noise."

Amplitude damping channel

The amplitude damping channel models spontaneous emission: a qubit in the excited state $|1\rangle$ decays to the ground state $|0\rangle$ with probability γ per unit time. The Kraus operators are

$$K_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix}, \qquad K_1 = \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix}.$$

The physical picture: K_0 is "no decay occurred" (the amplitude on $|1\rangle$ shrinks), K_1 is "decay occurred" (the qubit jumps from $|1\rangle$ to $|0\rangle$).

Trace-preservation check: $K_0^\dagger K_0 + K_1^\dagger K_1 = \mathrm{diag}(1, 1-\gamma) + \mathrm{diag}(0, \gamma) = I$. ✓

Effect on a state: Starting from $\rho = \begin{pmatrix} a & b \\ b^* & 1-a \end{pmatrix}$, one finds

$$\mathcal{E}(\rho) = \begin{pmatrix} a + \gamma(1-a) & \sqrt{1-\gamma}\, b \\ \sqrt{1-\gamma}\, b^* & (1-\gamma)(1-a) \end{pmatrix}.$$

The population $(1-a)$ on $|1\rangle$ decays exponentially; the coherences b decay as $\sqrt{1-\gamma}$. This is the source of the $T_2 \leq 2T_1$ relationship derived in Node 5.4: the coherence decay rate is at least half the population decay rate, because part of the dephasing comes from population decay itself.

The channel is non-unital: $\mathcal{E}(I) \neq I$ (the output is biased toward $|0\rangle\langle 0|$).

Stinespring dilation of amplitude damping

The amplitude damping channel can be realized as a unitary on the qubit plus an environment qubit, initially in $|0\rangle_E$:

$$V : |0\rangle_S \otimes |0\rangle_E \to |0\rangle_S \otimes |0\rangle_E,$$
$$V : |1\rangle_S \otimes |0\rangle_E \to \sqrt{1-\gamma}\, |1\rangle_S \otimes |0\rangle_E + \sqrt{\gamma}\, |0\rangle_S \otimes |1\rangle_E.$$

Tracing out the environment gives exactly the amplitude damping Kraus operators. The physical picture: the excited state $|1\rangle$ "leaks" into an entangled superposition where, with amplitude $\sqrt{\gamma}$, the system decays to ground and the environment absorbs a quantum of excitation. This is the quantum-mechanical content of spontaneous emission.

Measurement as a channel

A projective measurement with outcomes $\{P_i\}$ can be described as a channel whose Kraus operators are the projectors themselves (assuming outcomes are forgotten after measurement):

$$\mathcal{E}_{\text{meas}}(\rho) = \sum_i P_i \rho P_i.$$

This is a unital CPTP map that implements "dephasing in the measurement basis." If the outcome is retained as classical information, the channel becomes a *quantum-classical channel* $\rho \mapsto \sum_i \mathrm{Tr}(P_i \rho) |i\rangle\langle i|$, which destroys quantum coherence entirely and outputs a classical distribution.

Cross-Links

- **Module 5.2 (Reduced density operator):** The partial trace is a specific channel (tracing out subsystems). Every channel arises this way via Stinespring.
- **Module 5.4 (Decoherence mechanisms):** Dephasing and amplitude damping channels are the canonical quantum noise models, and are decoherence channels.
- **Module 7.4 (Von Neumann entropy):** Data-processing inequality $S(\mathcal{E}(\rho)) \geq S(\rho)$ for unital channels; mutual information is contractive under channels.
- **Module 8.2 (Lindblad equation):** The generator of a continuous-time semigroup of channels. Lindblad operators are the infinitesimal analogue of Kraus operators.
- **Module 4.2 (Separability):** Partial transpose is positive but not completely positive — it is not a channel, which is why PPT can detect entanglement.
- **Statistics:** Classical stochastic (Markov) channels are the commutative analogue: linear, completely positive (trivially), trace-preserving maps on probability distributions.

Structural Compression

Invariant: A quantum channel is a completely positive trace-preserving linear map. This is the unique class of transformations representing physically realizable operations on quantum states — neither smaller nor larger.

Mechanism: Three equivalent representations: Stinespring dilation (unitary on a larger system + partial trace), Kraus operator sum ($\sum_k K_k \rho K_k^\dagger$ with $\sum K_k^\dagger K_k = I$), and the Choi matrix via the Choi–Jamiołkowski isomorphism. Complete positivity is strictly stronger than positivity, forced by the existence of entangled ancillas.

Cross-System Echo: Stochastic matrices on probability distributions are the classical (commutative) analogue: linear, positive, trace-preserving. CPTP maps are the non-commutative generalization. Complete positivity — the strengthening of positivity to allow entangled inputs — is the specifically quantum ingredient.

Exercises

1. Show that the transpose map $T : \rho \mapsto \rho^T$ is positive but not completely positive by computing $(T \otimes I)|\Phi^+\rangle\langle\Phi^+|$ and finding its eigenvalues.
2. Verify that the Kraus operators of the depolarizing channel satisfy $\sum_k K_k^\dagger K_k = I$, and compute the action on the Bloch vector.
3. Derive the amplitude damping Kraus operators from a Stinespring dilation using a qubit environment.
4. Compute the Choi matrix for the depolarizing channel with $p = 1/2$. Verify that it is positive and satisfies the trace condition.
5. Show that the composition of two channels is a channel by constructing its Kraus representation from the individual Kraus operators.
6. Define a non-unital channel and discuss its fixed points. What is the fixed point of the amplitude damping channel?
7. Argue, structurally, why complete positivity (rather than mere positivity) is the correct mathematical requirement for physical channels. What would go wrong if we only required positivity?

Von Neumann Entropy and Mutual Information

Position in Knowledge Graph

System: Quantum Foundations **Structural Domain:** Quantum Information **Node:** 7.4

The **von Neumann entropy** $S(\rho) = -\operatorname{Tr}(\rho \log \rho)$ is the quantum analogue of Shannon entropy. It measures the information content (or equivalently, the mixedness, or the uncertainty) of a quantum state and is the foundational information-theoretic quantity for quantum systems. From it are derived quantum relative entropy, quantum mutual information, and the entropies that control channel capacities, entanglement measures, and thermodynamic bounds.

The quantum theory differs from the classical theory in two important ways. First, entropies can exhibit strict subadditivity with surprising strength — joint entropy can be *less* than the entropy of a single subsystem, a phenomenon impossible in classical Shannon theory. Second, quantum mutual information can exceed twice the corresponding classical mutual information, reflecting the extra correlations available through entanglement. Together these give von Neumann entropy a character distinct from its classical cousin, even though it reduces to Shannon entropy on diagonal states.

Formal Definition

Von Neumann entropy

For a density operator ρ on a finite-dimensional Hilbert space $\mathcal{H}$,

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

where $\{\lambda_i\}$ are the eigenvalues of ρ (with $0 \log 0 \equiv 0$ by continuity). The logarithm is base 2 for entropies measured in bits, base e for nats.

Basic properties:

- **Non-negativity.** $S(\rho) \geq 0$, with equality iff ρ is a pure state.
- **Upper bound.** $S(\rho) \leq \log d$, where $d = \dim \mathcal{H}$, with equality iff $\rho = I/d$.
- **Unitary invariance.** $S(U\rho U^\dagger) = S(\rho)$ for any unitary U.
- **Concavity.** $S(\lambda\rho_1 + (1-\lambda)\rho_2) \geq \lambda S(\rho_1) + (1-\lambda)S(\rho_2)$.
- **Additivity over products.** $S(\rho_A \otimes \rho_B) = S(\rho_A) + S(\rho_B)$.

Relative entropy

The **quantum relative entropy** (Umegaki) is

$$S(\rho\|\sigma) = \operatorname{Tr}(\rho \log \rho) - \operatorname{Tr}(\rho \log \sigma),$$

defined when $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$ and $+\infty$ otherwise. Klein's inequality: $S(\rho\|\sigma) \geq 0$, with equality iff $\rho = \sigma$. Relative entropy is the fundamental "distance" in quantum information, underlying hypothesis testing, resource theories, and thermodynamics.

Joint, conditional, and mutual information

For a bipartite state ρ_{AB} with marginals $\rho_A = \operatorname{Tr}_B \rho_{AB}$ and $\rho_B = \operatorname{Tr}_A \rho_{AB}$:

- **Joint entropy:** $S(AB) = S(\rho_{AB})$.
- **Conditional entropy:** $S(A|B) = S(AB) - S(B)$.
- **Mutual information:** $I(A:B) = S(A) + S(B) - S(AB) = S(\rho_{AB}\|\rho_A \otimes \rho_B)$.

A striking quantum fact: $S(A|B)$ can be *negative*. For example, if ρ_{AB} is a pure entangled state, then $S(AB) = 0$ but $S(B) > 0$, so $S(A|B) = -S(B) < 0$. Negative conditional entropy has no classical analogue and is closely related to the ability to perform quantum protocols like teleportation.

Intuition

Von Neumann entropy measures "how mixed" a state is, equivalently "how much classical uncertainty is encoded in ρ." A pure state has $S = 0$ — it is a definite quantum state, and although measurements give probabilistic outcomes, the state itself is maximally known. A maximally mixed state $\rho = I/d$ has $S = \log d$ — it is the state of maximum ignorance, corresponding to the uniform classical distribution over a basis of size d.

What makes von Neumann entropy non-classical is the behavior of bipartite entropy under entanglement. For a pure entangled state $|\psi\rangle_{AB}$, the whole system has $S(AB) = 0$, but each subsystem has $S(A) = S(B) > 0$. This is structurally impossible in classical information theory: a classical joint distribution with zero entropy is a delta function, which has marginals with zero entropy. Quantum joint purity does not imply marginal purity because the whole system is more than the sum of its parts — the entanglement is "in between" the subsystems, not localized in either.

Mutual information is the natural measure of how much two subsystems are correlated. Classically, $I(A : B) \leq \min(H(A), H(B))$. Quantumly, $I(A : B) \leq 2\min(S(A), S(B))$ — a factor of 2 more, achieved by maximally entangled states. This extra factor is one of the structural signatures of quantum correlations and is tightly connected to why quantum channels can transmit more classical information per qubit (when assisted by entanglement) than classical channels can transmit per bit.

The practical slogan: entropy measures the number of (qu)bits needed on average to describe the state classically, given some reference basis. A pure state needs zero bits (you already know it). The maximally mixed state needs $\log d$ bits (you know nothing). An arbitrary state lies somewhere in between, and its position is exactly the entropy.

Structural Role

- **Foundational quantity.** Von Neumann entropy is *the* fundamental information-theoretic quantity for quantum systems, analogous to Shannon entropy in the classical case. Every capacity theorem, compression theorem, and information inequality in quantum information theory is formulated in terms of $S(\rho)$ or related quantities.
- **Entanglement entropy.** For a bipartite pure state, $S(\rho_A) = S(\rho_B)$ is the *entanglement entropy* — the unique measure of entanglement for pure states (Node 4.5). Entanglement entropy is the operational measure used in condensed matter and holography (area laws, entanglement entropy of ground states).
- **Channel capacities.** Classical and quantum capacities of quantum channels are expressed in terms of entropic quantities: the Holevo capacity $\chi = S(\bar{\rho}) - \sum_i p_i S(\rho_i)$ (Node 7.5) and the coherent information $I_c = S(B) - S(AB)$ are the fundamental capacity formulas.
- **Thermodynamic bridge.** Thermal states minimize free energy $F = E - TS$ at fixed temperature, where S is the von Neumann entropy. This gives the quantum-statistical-mechanics link between information and heat, developed in Module 9.
- **Data processing.** Mutual information cannot increase under local channels: $I(A : B)_{(\mathrm{id} \otimes \mathcal{E})(\rho)} \leq I(A : B)_\rho$. This is the data-processing inequality, the quantum generalization of the classical statement that post-processing cannot create information. It is equivalent to monotonicity of quantum relative entropy under CPTP maps, which in turn is one of the deepest theorems in quantum information (Lindblad 1975, Uhlmann).
- **Strong subadditivity.** Lieb–Ruskai (1973): $S(ABC) + S(B) \leq S(AB) + S(BC)$ for any tripartite state. A nontrivial and highly structural inequality; equivalent to monotonicity of relative entropy and to the data-processing inequality for quantum mutual information.

Mathematical Development

Klein's inequality

For density operators ρ, σ,

$$S(\rho\|\sigma) = \mathrm{Tr}(\rho(\log\rho - \log\sigma)) \geq 0,$$

with equality iff $\rho = \sigma$. Proof: use the operator inequality $x\log x - x\log y \geq x - y$ (from concavity of log) on eigenvalues. Klein's inequality is the quantum analogue of non-negativity of classical KL divergence and underlies most entropic inequalities.

Subadditivity

For any bipartite state ρ_{AB},

$$S(AB) \leq S(A) + S(B),$$

with equality iff $\rho_{AB} = \rho_A \otimes \rho_B$. This follows from Klein's inequality applied to ρ_{AB} and $\rho_A \otimes \rho_B$:

$$0 \leq S(\rho_{AB}\|\rho_A \otimes \rho_B) = -S(AB) + S(A) + S(B).$$

Subadditivity implies mutual information is non-negative: $I(A:B) = S(A) + S(B) - S(AB) \geq 0$.

Strong subadditivity (Lieb–Ruskai)

For any tripartite state ρ_{ABC},

$$S(ABC) + S(B) \leq S(AB) + S(BC).$$

This is the quantum version of the classical inequality $H(ABC) + H(B) \leq H(AB) + H(BC)$, but it is vastly harder to prove in the quantum case. It was conjectured by Lanford and Robinson and proved by Lieb and Ruskai in 1973 using a key operator convexity theorem of Lieb. Strong subadditivity implies monotonicity of mutual information under local channels, Markov chain conditions for quantum states, and many other central results.

Data processing inequality

For any CPTP map $\mathcal{E}$ on system B and any bipartite state ρ_{AB},

$$I(A:B)_{(\mathrm{id}_A \otimes \mathcal{E})(\rho_{AB})} \leq I(A:B)_{\rho_{AB}}.$$

Equivalently: quantum relative entropy is monotone under CPTP maps,

$$S(\mathcal{E}(\rho)\|\mathcal{E}(\sigma)) \leq S(\rho\|\sigma).$$

This is the fundamental information-theoretic bound on how much correlation between systems can survive processing. It implies the second law of thermodynamics (when applied to thermal reference states) and constrains all channel capacities.

Bounds on mutual information

For any bipartite state ρ_{AB} on $\mathcal{H}_A \otimes \mathcal{H}_B$,

$$0 \leq I(A:B) \leq 2\min(\log d_A, \log d_B).$$

The upper bound is achieved by maximally entangled pure states, where $S(A) = S(B) = \log d$ and $S(AB) = 0$, giving $I(A:B) = 2\log d$. Classical mutual information is bounded by $\min(H(A), H(B))$ (a factor of 2 smaller), reflecting the non-classical correlations available through entanglement.

Joint concavity and convexity

- **Concavity of entropy:** $S(\sum_i p_i\rho_i) \geq \sum_i p_i S(\rho_i)$. Mixing states increases entropy on average.
- **Joint convexity of relative entropy:** $S(\sum_i p_i\rho_i \| \sum_i p_i\sigma_i) \leq \sum_i p_i S(\rho_i\|\sigma_i)$.
- **Araki–Lieb triangle inequality:** $|S(A) - S(B)| \leq S(AB)$. This is the lower bound complementing subadditivity.

Worked Example

Entropies of basic states

- **Pure state:** $\rho = |\psi\rangle\langle\psi|$ has eigenvalues $(1, 0, 0, \ldots)$, so $S(\rho) = 0$.
- **Maximally mixed qubit:** $\rho = I/2$ has eigenvalues $(1/2, 1/2)$, so $S(\rho) = 1$ bit.
- **Maximally mixed d-dim state:** $S(I/d) = \log d$. For a qubit register of n qubits, $S(I/2^n) = n$ bits.
- **Mixed state $p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1|$:** $S = h(p) = -p\log p - (1-p)\log(1-p)$, the binary entropy function.
- **Thermal state at temperature T:** $\rho_{\text{th}} = e^{-\beta H}/Z$ with $\beta = 1/k_B T$. Then $S(\rho_{\text{th}}) = \beta\langle H\rangle + \log Z$, which is the standard thermodynamic entropy formula.

Bell state

The Bell state $|\Phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ has $S(\rho_{AB}) = 0$ (pure). The reduced density matrix of either qubit is $\rho_A = I/2$, with $S(A) = S(B) = 1$ bit. Therefore:

- $I(A:B) = 1 + 1 - 0 = 2$ bits — the maximum possible mutual information for a qubit pair.
- $S(A|B) = S(AB) - S(B) = 0 - 1 = -1$ bit — negative conditional entropy, a purely quantum phenomenon.

For comparison, the maximum classical mutual information for two bits is 1 bit (perfectly correlated). The Bell state doubles this.

Mixed Bell state

Consider $\rho_{AB} = p|\Phi^+\rangle\langle\Phi^+| + (1-p)I/4$ (a Werner-like state). One finds:

- $\rho_A = \rho_B = I/2$, so $S(A) = S(B) = 1$ bit independent of p.
- ρ_{AB} has eigenvalues $(p + (1-p)/4, (1-p)/4, (1-p)/4, (1-p)/4)$, giving

$$S(AB) = -\left(p + \frac{1-p}{4}\right)\log\left(p + \frac{1-p}{4}\right) - 3\cdot\frac{1-p}{4}\log\frac{1-p}{4}.$$

- $I(A:B) = 2 - S(AB)$. For $p = 1$, $S(AB) = 0$, $I = 2$. For $p = 0$, $S(AB) = 2$, $I = 0$. The transition is smooth and nonlinear.

Data processing on a Bell state

Apply a dephasing channel $\mathcal{E}(\rho) = (\rho + Z\rho Z)/2$ to qubit B of a Bell state. The resulting state is

$$\rho'_{AB} = \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|),$$

a classical-quantum correlated state with $S(AB) = 1$ (not zero anymore) and $I(A:B) = 1 + 1 - 1 = 1$ bit. The dephasing channel has halved the mutual information, consistent with the data processing inequality: we lost exactly the "coherent" part of the correlation, retaining only the "classical" part.

Cross-Links

- **Module 5.1 (Density operators):** The object on which entropy acts. Pure states have zero entropy; mixing always increases it.
- **Module 4.5 (Entanglement measures):** Entanglement entropy $S(\rho_A)$ is the canonical entanglement measure for pure bipartite states, and is the special case of von Neumann entropy applied to reduced density matrices.
- **Module 7.3 (Quantum channels):** Data-processing inequality expresses the monotonicity of entropic quantities under CPTP maps.
- **Module 7.5 (Holevo bound):** Bounds classical information extractable from quantum states in terms of entropic quantities.
- **Module 9.4 (Quantum thermodynamics):** Von Neumann entropy is the thermodynamic entropy of quantum states; equilibrium and non-equilibrium thermodynamics are built on its behavior.
- **Module 10.5 (Quantum resource theories):** Entropies quantify resource content (entanglement, coherence, athermality).
- **Statistics:** Shannon entropy is the classical restriction. Quantum relative entropy generalizes KL divergence; quantum mutual information generalizes classical mutual information.

Structural Compression

Invariant: $S(\rho) = -\operatorname{Tr}(\rho \log \rho)$ is the unique (up to normalization) entropy functional satisfying concavity, additivity on products, and continuity. It is zero for pure states, maximized at $\log d$ for the maximally mixed state, invariant under unitaries, non-decreasing under unital channels, and monotone-decreasing for relative entropy under any CPTP map.

Mechanism: Eigendecomposition of the density matrix: entropy is Shannon entropy of the eigenvalue spectrum. All quantum-specific phenomena (subadditivity with negative conditional entropy, doubling of mutual information, entanglement entropy of pure states) arise from the non-commuting structure of the density operators involved.

Cross-System Echo: Shannon entropy in classical information theory is the diagonal restriction; KL divergence is the classical analogue of relative entropy; classical mutual information sits inside quantum mutual information. The factor-of-2 gap in mutual information upper bounds quantifies the specifically quantum contribution from entanglement.

Exercises

1. Compute $S(\rho)$ for a qubit state $\rho = (I + \vec{r} \cdot \vec{\sigma})/2$ with Bloch vector of magnitude r. Show that $S = h((1+r)/2)$.
2. Show that $S(AB) \leq S(A) + S(B)$ using Klein's inequality applied to ρ_{AB} and $\rho_A \otimes \rho_B$.

3. Verify the Araki–Lieb inequality $|S(A) - S(B)| \leq S(AB)$ for a Bell state.
4. Compute the quantum mutual information of a Werner state $\rho = p|\Phi^+\rangle\langle\Phi^+| + (1-p)I/4$ as a function of p. At what value of p does it become zero?
5. State the strong subadditivity inequality and show how it implies monotonicity of quantum mutual information under local channels.
6. For a thermal state $\rho_{\text{th}} = e^{-\beta H}/Z$, derive the formula $S = \beta\langle H\rangle + \log Z$. How does this connect to the thermodynamic identity $S = (E - F)/T$?
7. Argue, structurally, why negative conditional entropy is possible quantumly but not classically. What does $S(A|B) < 0$ mean operationally?

Holevo Bound

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?

Teleportation and Superdense Coding

Position in Knowledge Graph

System: Quantum Foundations **Structural Domain:** Quantum Information **Node:** 7.6

Quantum teleportation (Bennett, Brassard, Crépeau, Jozsa, Peres, Wootters 1993) and **superdense coding** (Bennett–Wiesner 1992) are two protocols built on the same resource — a shared entangled Bell pair — that trade classical and quantum communication against each other in exactly reciprocal ways. Teleportation spends one Bell pair plus two classical bits to transmit one qubit. Superdense coding spends one Bell pair plus one qubit to transmit two classical bits. Together they establish entanglement as a *tradeable resource* that can be converted into classical or quantum communication and back, and they form the operational cornerstone of the theory of quantum information as a resource theory.

Beyond their foundational role, teleportation is the basic primitive for quantum networking, distributed quantum computation, and quantum state transfer over noisy channels. It is the closest real physical phenomenon to "beaming" — with the caveat that the original is destroyed in the process (required by no-cloning) and that no faster-than-light signaling is possible (classical information must accompany the entanglement).

Formal Statement

Teleportation protocol

Alice wants to transmit an unknown qubit state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ to Bob. They share a Bell pair $|\Phi^+\rangle_{A'B} = (|00\rangle + |11\rangle)/\sqrt{2}$, with qubit A' held by Alice and B held by Bob. The protocol:

1. Alice performs a Bell-basis measurement on her two qubits (the unknown $|\psi\rangle$ and her half of the Bell pair). The measurement has four possible outcomes: $\Phi^+, \Phi^-, \Psi^+, \Psi^-$.
2. Alice sends the 2-bit classical outcome to Bob.
3. Depending on the outcome, Bob applies a Pauli correction to his half of the Bell pair:
   - $\Phi^+ \to I$ (no correction),
   - $\Phi^- \to Z$,
   - $\Psi^+ \to X$,
   - $\Psi^- \to XZ$ (or equivalently Y up to phase).

After the correction, Bob's qubit is in the state $|\psi\rangle$. Alice's original qubit, after the Bell measurement, is no longer in the state $|\psi\rangle$ — consistent with no-cloning.

Resource accounting: 1 qubit teleported = 1 Bell pair (1 "ebit") + 2 classical bits.

Superdense coding protocol

Alice wants to transmit two classical bits $(a, b) \in \{0, 1\}^2$ to Bob. They share a Bell pair $|\Phi^+\rangle_{AB}$. The protocol:

1. Alice applies one of four Pauli operators to her half of the Bell pair, depending on (a, b):
   - $(0, 0) \to I$
   - $(0, 1) \to X$
   - $(1, 0) \to Z$
   - $(1, 1) \to XZ$
2. Alice sends her half of the Bell pair to Bob via a quantum channel.
3. Bob performs a Bell-basis measurement on both qubits and recovers the 2-bit message.

The four possible operations map $|\Phi^+\rangle$ into the four Bell states $\{\Phi^+, \Phi^-, \Psi^+, \Psi^-\}$, which Bob can perfectly distinguish because they are orthogonal.

Resource accounting: 2 classical bits transmitted = 1 Bell pair + 1 qubit.

Reciprocity

The two protocols are exactly reciprocal in their resource conversions: - Teleportation: 2 cbits + 1 ebit → 1 qubit. - Superdense coding: 1 qubit + 1 ebit → 2 cbits.

A Bell pair (ebit) is the pivot, and the direction of resource flow depends on whether you are spending ebits to send classical information or quantum information.

Intuition

Teleportation's surprising feature is that Alice can transmit an *unknown* quantum state — one she cannot even describe, since describing it would require infinitely many classical bits (continuous amplitudes) — using only 2 classical bits. The "extra" information required to specify $|\psi\rangle$ is pre-packaged in the entanglement of the Bell pair. The Bell pair is a kind of "shared randomness with quantum flavor," and once Alice performs her Bell measurement, her measurement outcome is maximally random (each outcome has probability 1/4) regardless of $|\psi\rangle$. The specific random outcome, when combined with the Bell pair's pre-existing entanglement, picks out the correct unitary correction at Bob's end.

No faster-than-light signaling is possible. Before receiving Alice's 2 classical bits, Bob's qubit is the reduced density matrix $I/2$ — the maximally mixed state — which contains no information about $|\psi\rangle$ whatsoever. Only after Bob applies the correct Pauli correction (which requires knowing Alice's outcome) does his qubit become $|\psi\rangle$. The information is carried by the combination of the entanglement (already in place) and the classical message (traveling at most at light speed).

Superdense coding's surprise is that Alice can transmit *two* classical bits by sending only *one* qubit. Without shared entanglement, this would violate the Holevo bound. But with the Bell pair, Alice's qubit is never "alone" — it is part of a 4-dimensional entangled system, and the operations Alice applies to her half reveal themselves as distinct orthogonal states when measured jointly with Bob's half. The effective information capacity is $\log 4 = 2$ bits, matching what superdense coding achieves.

The reciprocity is structural: both protocols exploit the fact that the Bell basis is a 4-dimensional space in which any of the four distinct orthogonal states encodes 2 bits of classical information. Teleportation uses Alice's Bell measurement to "read out" which state the joint system is in (giving 2 classical bits); superdense coding uses Alice's Pauli operation to "write in" one of 4 Bell states (encoding 2 classical bits). The two are Fourier-dual in a sense.

Structural Role

- **Entanglement as a resource.** Teleportation and superdense coding establish that entanglement is not just a static correlation but a dynamical resource that can be converted into communication. This observation is the seed of quantum resource theories (Module 10.5), in which entanglement is quantified by how much it can do under restricted operations (LOCC, typically).
- **Classical vs quantum communication trade-offs.** The two protocols give the canonical answer to "how do classical bits and quantum bits trade against each other?" — through the intermediary of shared ebits. The full theory of such trade-offs (classical capacity, quantum capacity, entanglement-assisted capacity) is built on these primitives.
- **Quantum networking.** Teleportation is the basic primitive for sending quantum states over long distances. Quantum repeaters — the enabling technology for long-distance quantum communication — work by chaining multiple short-distance teleportation steps together, with entanglement distillation in between to combat noise.
- **Distributed quantum computation.** Gate teleportation generalizes state teleportation: a quantum gate can be "teleported" from one location to another using pre-shared entanglement and classical communication, enabling distributed quantum computation across separated processors.

- **Consistency with no-cloning.** Teleportation is not cloning: Alice's original qubit is destroyed by her Bell measurement, and only Bob's qubit retains $|\psi\rangle$. No-cloning would forbid any protocol that produced two copies; teleportation produces one copy and destroys the original, staying compatible.
- **Measurement-based quantum computation.** One-way quantum computation (Raussendorf–Briegel) is built on the idea that teleportation-like steps, repeated through a large entangled resource state, can implement arbitrary quantum computations. Teleportation is the local primitive; a graph state is the global resource.

Mathematical Development

Teleportation derivation

Alice holds qubits S (the unknown state) and A' (her half of the Bell pair); Bob holds qubit B (the other half). The total state is

$$|\psi\rangle_S \otimes |\Phi^+\rangle_{A'B} = (\alpha|0\rangle_S + \beta|1\rangle_S) \otimes \frac{1}{\sqrt{2}}(|00\rangle_{A'B} + |11\rangle_{A'B}).$$

Expand and regroup in the Bell basis of qubits S and A':

$$\begin{aligned} &= \frac{1}{2}\Big[|\Phi^+\rangle_{SA'}(\alpha|0\rangle + \beta|1\rangle)_B + |\Phi^-\rangle_{SA'}(\alpha|0\rangle - \beta|1\rangle)_B \\ &+|\Psi^+\rangle_{SA'}(\alpha|1\rangle + \beta|0\rangle)_B + |\Psi^-\rangle_{SA'}(\alpha|1\rangle - \beta|0\rangle)_B\Big]. \end{aligned}$$

The four terms correspond to the four possible Bell measurement outcomes, each with probability $1/4$.

Alice measures qubits S and A' in the Bell basis. Suppose she obtains outcome Ψ^+. Bob's qubit is projected to $\alpha|1\rangle + \beta|0\rangle$, which equals $X|\psi\rangle$ — so Bob must apply X to recover $|\psi\rangle$.

The full outcome-to-correction map:

Alice's outcome	Bob's state	Correction
Φ^+	$\alpha\|0\rangle + \beta\|1\rangle$	I
Φ^-	$\alpha\|0\rangle - \beta\|1\rangle$	Z
Ψ^+	$\alpha\|1\rangle + \beta\|0\rangle$	X
Ψ^-	$\alpha\|1\rangle - \beta\|0\rangle$	XZ

After the appropriate correction, Bob's qubit is in state $|\psi\rangle$, regardless of the original amplitudes α, β.

Reduced state of Bob before classical communication

Before Bob receives Alice's outcome, his qubit's state is the reduced density matrix

$$\rho_B = \mathrm{Tr}_{SA'}\left(|\psi\rangle\langle\psi|_S \otimes |\Phi^+\rangle\langle\Phi^+|_{A'B}\right) = \frac{I}{2}.$$

This is independent of $|\psi\rangle$ — Bob has no access to the teleported state until Alice's 2 classical bits arrive. The protocol is consistent with no-signaling: classical communication is required to complete it.

Superdense coding derivation

Start with a Bell pair $|\Phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ shared between Alice (qubit A) and Bob (qubit B). Alice applies one of four Pauli operations to her qubit:

- $I \otimes I|\Phi^+\rangle = |\Phi^+\rangle$
- $Z \otimes I|\Phi^+\rangle = (|00\rangle - |11\rangle)/\sqrt{2} = |\Phi^-\rangle$
- $X \otimes I|\Phi^+\rangle = (|10\rangle + |01\rangle)/\sqrt{2} = |\Psi^+\rangle$
- $XZ \otimes I|\Phi^+\rangle = (|10\rangle - |01\rangle)/\sqrt{2} = |\Psi^-\rangle$ (up to an overall sign)

The four resulting Bell states are mutually orthogonal. Alice then sends qubit A to Bob, who now holds both qubits and performs a Bell measurement, recovering with certainty which of the four Bell states is present and hence which 2-bit message Alice encoded.

Resource accounting via coherent information

The coherent information $I_c(A\rangle B) = S(B) - S(AB)$ measures the quantum capacity of a channel. For a noiseless channel with prior ebit assistance, the coherent information of 1 qubit is $\log 2 = 1$, matching teleportation's rate. Entanglement-assisted classical capacity (Bennett–Shor–Smolin–Thapliyal 2002) gives $2 \log d$ for a noiseless channel, consistent with superdense coding.

These are special cases of the **resource calculus** of quantum Shannon theory, in which inequalities relate rates of qubit, cbit, and ebit consumption. Teleportation and superdense coding are the tightest extremal resource conversions in this calculus.

Worked Example

Teleporting $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ explicitly

Initial state (Alice holds qubits 1 and 2, Bob holds qubit 3):

$$|\psi\rangle_1|\Phi^+\rangle_{23} = (\alpha|0\rangle + \beta|1\rangle) \otimes \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle).$$

Expand:

$$= \frac{1}{\sqrt{2}}\big(\alpha|0\rangle|00\rangle + \alpha|0\rangle|11\rangle + \beta|1\rangle|00\rangle + \beta|1\rangle|11\rangle\big).$$

Rewrite the joint state of qubits 1 and 2 in the Bell basis:

$$|00\rangle = \frac{1}{\sqrt{2}}(|\Phi^+\rangle + |\Phi^-\rangle), \quad |11\rangle = \frac{1}{\sqrt{2}}(|\Phi^+\rangle - |\Phi^-\rangle),$$
$$|01\rangle = \frac{1}{\sqrt{2}}(|\Psi^+\rangle + |\Psi^-\rangle), \quad |10\rangle = \frac{1}{\sqrt{2}}(|\Psi^+\rangle - |\Psi^-\rangle).$$

Substituting and collecting terms (tedious but mechanical):

$$= \frac{1}{2}\big[|\Phi^+\rangle_{12}(\alpha|0\rangle_3 + \beta|1\rangle_3) + |\Phi^-\rangle_{12}(\alpha|0\rangle_3 - \beta|1\rangle_3)$$
$$+|\Psi^+\rangle_{12}(\alpha|1\rangle_3 + \beta|0\rangle_3) + |\Psi^-\rangle_{12}(\alpha|1\rangle_3 - \beta|0\rangle_3)\big].$$

Alice measures qubits 1 and 2 in the Bell basis. Each outcome has probability $1/4$. If the outcome is Φ^+, qubit 3 is already in $|\psi\rangle$; no correction needed. If it is Φ^-, Bob applies Z, turning $\alpha|0\rangle - \beta|1\rangle$ into $\alpha|0\rangle + \beta|1\rangle = |\psi\rangle$. Similarly for the other outcomes. In all four cases, after the classical communication and Pauli correction, Bob's qubit is $|\psi\rangle$.

Superdense coding, two-bit message "11"

Alice wants to send $(1, 1)$. She applies XZ to her half of $|\Phi^+\rangle$:

$$(XZ \otimes I)|\Phi^+\rangle = (XZ \otimes I)\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) = \frac{1}{\sqrt{2}}(|10\rangle - |01\rangle) = |\Psi^-\rangle \text{ (up to sign).}$$

Alice sends her qubit to Bob. Bob performs a Bell measurement and obtains Ψ^- with certainty, decoding "11." Total cost: 1 physical qubit transmitted + 1 Bell pair used = 2 bits of classical information delivered, at a rate of 2 cbits per qubit — double what any unassisted quantum channel could do, and matching the Holevo bound for 2-qubit systems.

Entanglement swapping

Combine teleportation with a second entangled pair to perform **entanglement swapping**: Alice holds qubits 1 and 2 (with 1 entangled with a distant qubit 0, and 2 entangled with Bob's qubit 3). Alice performs a Bell measurement on qubits 1 and 2. After classical communication and Bob's correction, qubits 0 and 3 — which have never interacted directly — become entangled. This is the primitive for quantum repeaters: long-distance entanglement is built up by swapping a chain of short-distance entangled pairs.

Cross-Links

- **Module 4.4 (Bell states and CHSH):** The Bell basis is the core ingredient in both protocols. Bell states are both the resource (Bell pair in teleportation) and the measurement basis.
- **Module 7.1 (Qubits):** The basic building blocks for the protocols.
- **Module 7.2 (No-cloning):** Teleportation is consistent with no-cloning because the original state is destroyed by Alice's Bell measurement.
- **Module 7.5 (Holevo bound):** Superdense coding saturates the Holevo bound for 2-qubit systems. It does not violate the bound because the relevant ensemble is on two qubits, not one.
- **Module 4.6 (Monogamy of entanglement):** Monogamy explains why teleportation uses up the Bell pair — the entanglement is "spent" in the Bell measurement and the pair becomes unentangled afterwards.
- **Module 10.5 (Quantum resource theories):** Teleportation and superdense coding are the foundational conversions in the resource theory of entanglement.
- **Statistics — shared randomness:** Classical shared randomness allows analogous (but much weaker) protocols; the quantum extension via shared entanglement gives exponentially greater power.

Structural Compression

Invariant: Shared entanglement is a tradeable resource that converts between classical and quantum communication. Teleportation: 1 ebit + 2 cbits $\to$ 1 qubit. Superdense coding: 1 ebit + 1 qubit $\to$ 2 cbits. The two protocols are reciprocal and exploit the 4-dimensional structure of the Bell basis.

Mechanism: Bell-basis measurement (teleportation) or Pauli encoding (superdense coding) on a pre-shared Bell pair, followed by classical communication (teleportation) or quantum transmission (superdense coding), followed by Pauli correction (teleportation) or Bell measurement (superdense coding).

Cross-System Echo: Shared randomness in classical communication allows rate-increasing protocols analogous to superdense coding (but much weaker, since shared random bits cannot transmit 2 cbits per 1 cbit of channel use). The quantum case is strictly stronger because Bell states have 4 orthogonal alternatives rather than 2, doubling the channel capacity with entanglement assistance.

Exercises

1. Derive the teleportation protocol by expanding $|\psi\rangle_S|\Phi^+\rangle_{A'B}$ in the Bell basis of SA'. Verify all four Pauli corrections.
2. Show explicitly that the reduced state of Bob's qubit before classical communication is $I/2$, independent of $|\psi\rangle$. Use this to argue that teleportation is consistent with no-signaling.
3. Verify that the four states $\{(I\otimes I)|\Phi^+\rangle, (Z\otimes I)|\Phi^+\rangle, (X\otimes I)|\Phi^+\rangle, (XZ\otimes I)|\Phi^+\rangle\}$ are mutually orthogonal and form a complete basis of two qubits.
4. Construct the entanglement swapping protocol: given Bell pairs $(0,1)$ and $(2,3)$, describe the measurement and corrections that entangle qubits 0 and 3.
5. Compute the success probability of teleportation if the shared Bell pair is imperfect, e.g., a Werner state $\rho = p|\Phi^+\rangle\langle\Phi^+| + (1-p)I/4$. At what p does teleportation fidelity drop below classical?
6. Describe gate teleportation: how to apply a known single-qubit gate U to a distant qubit using pre-shared entanglement and classical communication.
7. Argue, structurally, why teleportation and superdense coding are reciprocal in their resource consumption. What role does the 4-dimensional Bell basis play in both directions?