On the Hardness of Detecting Macropscopic Superpositions

Scott Aaronson (UT Austin)
Shoulders of Everett Workshop,
July 9, 2021


Background: For the past decade, a bunch of physicists (Susskind, Preskill, Swingle, Harlow...) have been thinking about the black hole information problem, AdS/CFT, and quantum gravity more generally, using the
language of qubits, quantum circuits & complexity.

Often they're forced to consider wildly nonlocal changes of basis (e.g., from Hawking radiation to the black hole interior that it 'encodes'), or states like

$$\frac{1}{\sqrt{2}}\left|\text{BlackHole}\right\rangle + \left|\text{NoBlackHole}\right\rangle$$

for which it's extremely hard to avoid the question of Everettianism.

Central to their investigations has been the following distance metric on n-qubit states:

$$C_{\epsilon}(\left|\psi\right\rangle, \left|\phi\right\rangle) = \text{minimum \# of } 1\&2\text{-qubit gates in any quantum circuit } C \text{ such that } C\left|\psi\right\rangle \approx_{\epsilon} \left|\phi\right\rangle$$
Circuit C such that $U_{\text{nl}} = e^{\frac{\lambda}{2}}$

"Relative Circuit Complexity"

This is wildly different from the usual Hilbert space distance! E.g., $|0\ldots 00\rangle$ and $|0\ldots 01\rangle$ are close in complexity metric yet orthogonal.

If $|z\rangle$ is Haar-random, then

$|00\ldots 00\rangle$ and $\sqrt{0.99} |00\ldots 00\rangle + \sqrt{0.01} |z\rangle$

are close in inner product, yet exponentially far in complexity metric (w.h.p. over $|z\rangle$ and for small enough $\varepsilon$)

In Jan. 2016, Susskind (who uses me as a CS consultant) posed the following problem to me: given two orthogonal "macroscopic" states $|V\rangle$ and $|W\rangle$ (e.g.
geometries in quantum gravity), give a complexity criterion for when it's "hard" or "easy" to distinguish a superposition

\[
\frac{|V\rangle + |W\rangle}{\sqrt{2}}
\]

from an incoherent mixture of the two.

Susskind's conjecture: The circuit complexity of this distinguishing task should be characterized by \( C_{E}(|V\rangle, |W\rangle) \), the relative circuit complexity of \( |V\rangle \& |W\rangle \).

What we found: Susskind's conjecture is not quite true (we give an explicit counterexample). But the complexity of detecting macroscopic coherence is characterized—up to
a factor of $\leq$ by a closely related measure called "swap complexity":

$$S_c(\{\psi, \phi\}) = \# \text{ of gates in smallest circuit } C \text{ s.t. } |\psi_C| + |\phi_C| \geq 1 - \frac{\epsilon}{2}$$

I.e., the same circuit $C$ has to map $|\psi\rangle$ to $|\phi\rangle$ and $|\phi\rangle$ to $|\psi\rangle$, with no relative phase!

Observation: Distinguishing $\frac{|\psi\rangle + |\phi\rangle}{\sqrt{2}}$ from $\frac{|\psi\rangle - |\phi\rangle}{\sqrt{2}}$ is essentially equivalent to distinguishing either from $|\psi\rangle X |\psi\rangle + |\phi\rangle X |\phi\rangle$.

Proof of the Swapping/Distinguishing Connection

1) Easy distinguishing $\Rightarrow$ easy swapping

Let $|\psi\rangle = \frac{|\psi\rangle + |\phi\rangle}{\sqrt{2}}$ & $|\phi\rangle = \frac{|\psi\rangle - |\phi\rangle}{\sqrt{2}}$,
So \( |v\rangle = \frac{|u\rangle + i|\psi\rangle}{\sqrt{2}} \) \& \( |w\rangle = \frac{|u\rangle - i|\psi\rangle}{\sqrt{2}} \).

Suppose \( C|\psi\rangle = |0\rangle \ldots \) \& \( C|\phi\rangle = |1\rangle \ldots \).

Then consider:

![Circuit Diagram]

\( |v\rangle \) or \( |w\rangle \)

1. Easy swapping \( \Rightarrow \) easy distinguishing

Suppose \( C|v\rangle = |w\rangle \) \& \( C|w\rangle = |v\rangle \).

Then consider

![Circuit Diagram]

\( |y\rangle \) or \( |y'\rangle \)

We show that both implications are robust.

**Theorem:** Suppose we have a circuit \( C \)
\[ \text{s.t.} \quad \langle u | C | v \rangle = a \) \& \( \langle v | C | u \rangle = 6. \]
Using $C$, we can distinguish \( \frac{\ket{u} + \ket{v}}{\sqrt{2}} \) from \( \frac{\ket{u} - \ket{v}}{\sqrt{2}} \) with bias \( \frac{1 + 6\lambda}{2} \).

Theorem: Suppose we can distinguish \( \frac{\ket{u} + \ket{v}}{\sqrt{2}} \) from \( \frac{\ket{u} - \ket{v}}{\sqrt{2}} \) with bias \( \delta \), using a circuit with \( T \) gates. Then \( S_E(\ket{u}, \ket{v}) \leq 2T + 1 \).

We show that the error parameters in both theorems are tight.

---

**Remarks**

Swappability of \( \ket{u} \) & \( \ket{v} \) is **not** equivalent to distinguishability of \( \ket{u} \) & \( \ket{v} \).

Let \( \ket{z} \) be Haar-random, orthogonal to \( \ket{0} \).
Then \( |0\cdots 0; 1, 2\rangle\): easy to distinguish, hard to swap.

\[ \sqrt{2} |0\cdots 0; 1, 2\rangle, \sqrt{2} |0\cdots 0; -1, 2\rangle: \text{easy to swap, hard to distinguish.} \]

Distinguishability in one basis is equivalent to swappability in a complementary basis.

Why did we need to use swap complexity rather than relative complexity?

Suppose we have a small circuit \( C \) s.t.

\[ C|u\rangle = |v\rangle, \quad C|v\rangle = -|u\rangle \]

This need not help at all in distinguishing \( \frac{|u\rangle + |v\rangle}{\sqrt{2}} \) from \( \frac{|u\rangle - |v\rangle}{\sqrt{2}} \)?

The latter can even remain exponentially
hard, e.g. it

\[ |u\rangle = \frac{|0\rangle |2\rangle + |1\rangle |0\rangle}{\sqrt{2}} \]

\[ |v\rangle = \frac{i |0\rangle |2\rangle - i |1\rangle |0\rangle}{\sqrt{2}} \]

|z\rangle \text{ Haar-random & orthogonal to } |0\rangle \text{.}

This won't happen if e.g.

\[ |u\rangle = |\evil\rangle, \quad |v\rangle = |\happy\rangle \]

Since it's easy to check if a cat is dead & apply a $-1$ phase if so.

But it could happen for $|u\rangle, |v\rangle$ that are hard to distinguish.

Why Should Any of You Care About Any of This?
You've been fed misinformation about Schrödinger's cat!

\[ |\text{Alive}\rangle + |\text{Dead}\rangle \]
\[ \sqrt{2} \]

Macabre! Animal cruelty!

**THE REALITY:** If you had the technological ability (as measured by circuit size) to distinguish a superposition of Alive & Dead from an incoherent mixture, you'd necessarily also have the technological ability to bring a dead cat back to life.

"The Schrödinger's cat experiment is necromancy-hard!"
(If it rotating between \( |\text{Alive}\rangle & |\text{Dead}\rangle \)
is so easy, was the cat ever “dead” at all?)

People have made related observations before.
E.g., Aharonov & Rohrlich 2008: You could revive a dead cat with probability \( \frac{1}{2} \).
This is weaker than what we show.

② David Deutsch’s thought experiment.
“Testing” the MWI by running an AGI on a QC, and proving it was in a state like:

\[
\frac{1}{\sqrt{2}} (|\text{Perception}_1\rangle + |\text{Perception}_2\rangle)
\]

My view: Alas, people would keep arguing about MWI even then.
Just like, they’d keep arguing about
the Hard Problem of Consciousness—and for similar reasons.

One possible view: Consciousness, at least of any sort we know about from experience, is a phenomenon intimately bound up with the Arrow of Time, the Second Law & irreversibility.

If an external agent could rotate you at will from $|\text{Perception}_1\rangle$ to $|\text{Perception}_2\rangle$ and back to $|\text{Perception}_1\rangle$, would your "perceptions" mean anything? Would they even be perceptions at all?

And yet, if that external agent could measure interference between the $|\text{Perception}_1\rangle$ & $|\text{Perception}_2\rangle$ branches, then they'd necessarily have
branches, then they'd necessarily have this ability to rotate.

For more, see my "Ghost In The Quantum Turing Machine" essay (arXiv:1306.0159), though it simply accepts as folklore the connection between swapping & detecting interference that this work made more formal.

Discussion Prompts
① Is quantum circuit complexity a suitable proxy for "difficulty"?
② Is "difficulty" (e.g., of recohering two branches) ever relevant to quantum foundations?
③ If not, then what's a better,
1. criterion for two Everett branches to have independent existence?

4. Does the Everett interpretation require it to be possible, at least in principle, to recollect two different mental states of a conscious being?

5. Anything else we/I should try to prove that might bear on any of these questions?