On the Hardness of Detecting Macroscopic Superpositions

Scott Aaronson[1], Yosi Atia[1], and Leonard Susskind[2,3]

[1]Department of Computer Science, University of Texas at Austin, Austin, TX, USA
[2]SITP, Stanford University, Stanford, CA 94305, USA
[3]Google, Mountain View, CA 94043, USA

Abstract

When is decoherence "effectively irreversible"? Here we examine this central question of quantum foundations using the tools of quantum computational complexity. We prove that, if one had a quantum circuit to determine if a system was in an equal superposition of two orthogonal states (for example, the $|\text{Alive}\rangle$ and $|\text{Dead}\rangle$ states of Schrödinger's cat), then with only a slightly larger circuit, one could also *swap* the two states (e.g., bring a dead cat back to life). In other words, observing interference between the $|\text{Alive}\rangle$ and $|\text{Dead}\rangle$ states is a "necromancy-hard" problem, technologically infeasible in any world where death is permanent. As for the converse statement (i.e., ability to swap implies ability to detect interference), we show that it holds modulo a single exception, involving unitaries that (for example) map $|\text{Alive}\rangle$ to $|\text{Dead}\rangle$ but $|\text{Dead}\rangle$ to $-|\text{Alive}\rangle$. We also show that these statements are robust—i.e., even a *partial* ability to observe interference implies partial swapping ability, and vice versa. Finally, without relying on any unproved complexity conjectures, we show that all of these results are quantitatively tight. Our results have possible implications for the state dependence of observables in quantum gravity, the subject that originally motivated this study.

1 Introduction

Schrödinger's cat famously raised the question: how large does a quantum state have to be, before we can take it to represent actual events rather than just potentialities? In practice, the larger a state, the harder it is to keep track of all of its degrees of freedom, and the harder it is to prevent it from interacting with its environment; both effects can quickly make it infeasible to observe quantum coherence between different branches of the state. But is there some principled criterion for saying when two branches have become "macroscopically distinct," in the sense that observing interference between

them is now so technologically intractable that one might as well speak in terms of a "collapse" having happened?[1]

Brown and Susskind [3] conjectured that *relative complexity*, as defined below, characterizes how hard it is to observe coherence between two orthogonal quantum states $|x\rangle$ and $|y\rangle$—in the sense of performing a measurement that accepts the superposition $\frac{|x\rangle+|y\rangle}{\sqrt{2}}$ and rejects the superposition $\frac{|x\rangle-|y\rangle}{\sqrt{2}}$ with high probability. (We note that, by a convexity argument, this is essentially equivalent to distinguishing either of those two superpositions from the *classical mixture* $\frac{1}{2}(|x\rangle\langle x| + |y\rangle\langle y|)$.)

Definition 1. [Relative complexity, adapted from [4]]
Let $|x\rangle, |y\rangle$ be two n-qubit pure quantum states. Their relative complexity, $\mathcal{C}_\varepsilon(|x\rangle, |y\rangle)$, is the minimal number of gates in a circuit C such that

$$\left| \langle y| \langle 0 \dots 0| C |x\rangle |0 \dots 0\rangle \right|^2 \geq 1 - \varepsilon.$$

The gates are chosen from an arbitrary fixed universal set of 1-qubit and 2-qubit gates; ancilla qubits are allowed as long as they return the $|0 \dots 0\rangle$ state.

We will omit the dependence on ε when it is not necessary. It is easy to see that $\mathcal{C}_\varepsilon$ is a metric: it is symmetric; it is zero iff $|x\rangle = |y\rangle$; and it satisfies the triangle inequality (with ε increased):

$$\mathcal{C}(|x\rangle, |y\rangle) + \mathcal{C}(|y\rangle, |z\rangle) \geq \mathcal{C}(|x\rangle, |z\rangle)$$

By a counting argument, the relative complexity of almost any pair of n-qubit states is $2^{\Omega(n)}$. Importantly, two states could be orthogonal, but still extremely close in relative complexity distance: for example, $|0^n\rangle$ and $|0^{n-1}1\rangle$. Conversely, two states could be close in ℓ_2-distance but far in relative complexity: for example, consider the states $|0^n\rangle$ and $|\phi\rangle = \sqrt{1-\varepsilon}\,|0^n\rangle + \sqrt{\varepsilon}\,|\psi\rangle$, where $|\psi\rangle$ is Haar-random and ε is small. The ℓ_2-distance is between them is $O(\varepsilon)$, but for instance, $\mathcal{C}_{\varepsilon/4}(|\psi\rangle, |\phi\rangle)$ is exponential.

In this paper we also consider a slightly different notion, the *swap complexity* of two states, which is at least as large as their relative complexity but could be larger.

Definition 2. Let $|x\rangle, |y\rangle$ be two quantum states. Their swap complexity, $\mathcal{S}_\varepsilon(|x\rangle, |y\rangle)$, is the minimal number of gates in a circuit C such that

$$\frac{|\langle x| \langle 0 \dots 0| C |y\rangle |0 \dots 0\rangle + \langle y| \langle 0 \dots 0| C |x\rangle |0 \dots 0\rangle|}{2} \geq 1 - \varepsilon$$

Ancilla qubits are allowed as long as they return the $|0 \dots 0\rangle$ state.

[1] Similarly, one of the questions raised in the context of the AdS/CFT correspondence is the "state dependence of observables" [1,2]. For example, in the discussion of black hole firewalls, which bulk operator an observer can apply depends on the spacetime background—e.g., is there a black hole or no black hole? This means that the set of measurements one can perform would depend on the quantum state of the spacetime background. While this seems to make little sense in the context of standard quantum mechanics, it is really about the dictionary between the standard quantum mechanics of the boundary holographic description and the incompletely understood bulk description of phenomena behind the horizon. Our results show that, if an observer cannot efficiently map one spacetime branch to the other, then she also cannot efficiently measure a superposition of the two branches, and thus effectively sees one or the other.

We show that the complexity of transforming $|x\rangle \leftrightarrow |y\rangle$, is (up to a constant factor) equal to the complexity of perfectly distinguishing between $|\psi\rangle = \frac{|x\rangle+|y\rangle}{2}$ and $|\phi\rangle = \frac{|x\rangle-|y\rangle}{2}$. The swap complexity is at least the relative complexity, and as we show later, there are cases where the swap complexity is exponential while the relative complexity is $O(1)$.

Going further, we show an equivalence even between *approximate* versions of swapping and distinguishing. Without relying on unproved complexity conjectures, we also show that our approximate equivalence theorem is optimal, in the sense that a result with better error parameters would be false.

Of course, qualitatively similar observations had been made before, but as far as we know, never in the sharp form here, which seems to require the formal notion of quantum circuit complexity or something similar. As one example, Aharonov and Rohrlich [5, Chapter 9] pointed out that the ability to measure a cat in the $\{|\text{Alive}\rangle \pm |\text{Dead}\rangle\}$ basis would imply the ability to revive a dead cat with success probability 1/2, a weaker statement than what we show here.

The equivalence between swap complexity and observing coherence is interesting in the context of the foundations of quantum mechanics. For example, in the Schrödinger's cat experiment, having the technological ability to *detect* that the cat was in superposition state at all, implies having the ability to perform a unitary that revives a dead cat, an ability that one could call "quantum necromancy." In other words, our results show that, if reviving a dead cat is considered "hard"—for essentially any reasonable definition of "hard"—then distinguishing Schrödinger's cat from a classical mixture is "hard" in that same sense, and the cat can be treated as effectively decohered. Similarly, in the Wigner's Friend thought experiment [6, 7], if Wigner can detect that his friend is in superposition then he can also swap his friend's mental states.

2 Main Result

2.1 Perfect case

We start by proving the equivalence (in circuit complexity) between a perfect swapper of two orthogonal states, and a perfect distinguisher for the corresponding conjugate states.

Theorem 1. *Let $|x\rangle, |y\rangle$ be n-qubit orthogonal quantum states, and let $|\psi\rangle = \frac{|x\rangle+|y\rangle}{\sqrt{2}}$ and $|\phi\rangle = \frac{|x\rangle-|y\rangle}{\sqrt{2}}$. The following two statements are equivalent:*

(i) There is a unitary U such that $U|x\rangle = |y\rangle$ and $U|y\rangle = |x\rangle$ with circuit complexity $O(T(n))$.

(ii) There is a unitary which perfectly distinguishes between $|\psi\rangle$ and $|\phi\rangle$ with circuit complexity $O(T(n))$.

Indeed, one can perfectly distinguish $|\psi\rangle$ from $|\phi\rangle$ given a single black-box access to a controlled swap of $|x\rangle$ and $|y\rangle$, and one can swap $|x\rangle$ and $|y\rangle$ given a single black-box

access to a unitary A that simulates a measurement perfectly distinguishing $|\psi\rangle$ from $|\phi\rangle$, as well as a single black-box access to $A^\dagger$.

Proof.
$(i) \to (ii)$
For distinguishing between $|\psi\rangle$ and $|\phi\rangle$, apply U to $|\psi\rangle$ or $|\phi\rangle$, conditioned on a $|+\rangle$ control qubit being $|1\rangle$ (see Fig. 1). Then check whether the control qubit becomes $|-\rangle$. Only $O(1)$ gates are added to U, hence the complexity is still $O(T(n))$.[2]

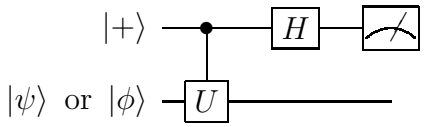

Figure 1: A circuit distinguishing $|\psi\rangle$ from $|\phi\rangle$ using a unitary U that swaps $|x\rangle$ with $|y\rangle$.

$(ii) \to (i)$
Suppose we had a unitary A such that $A|\psi\rangle = |0\rangle|g_\psi\rangle$ and $A|\phi\rangle = |1\rangle|g_\phi\rangle$ where $|g_\psi\rangle, |g_\phi\rangle$ are arbitrary states of the remaining $n-1$ qubits. Additionally, the circuit complexity of A is $O(T(n))$. Then to swap $|x\rangle = \frac{|\psi\rangle+|\phi\rangle}{\sqrt{2}}$ and $|y\rangle = \frac{|\psi\rangle-|\phi\rangle}{\sqrt{2}}$, we just apply A, then apply a Z gate on the first qubit, and finally uncompute by applying $A^\dagger$ (see Figure 2). Formally,

$$|x\rangle = \frac{|\psi\rangle+|\phi\rangle}{\sqrt{2}} \xrightarrow{A} \frac{|0\rangle|g_\psi\rangle+|1\rangle|g_\phi\rangle}{\sqrt{2}} \xrightarrow{Z_1} \frac{|0\rangle|g_\psi\rangle-|1\rangle|g_\phi\rangle}{\sqrt{2}} \xrightarrow{A^\dagger} \frac{|\psi\rangle-|\phi\rangle}{\sqrt{2}} = |y\rangle\,.$$

The total circuit complexity is twice the circuit complexity of A and another Z gate, which sums up to $O(T(n))$.

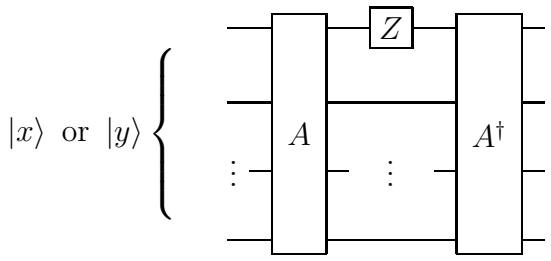

Figure 2: A circuit implementing $|x\rangle \leftrightarrow |y\rangle$ using a distinguishing circuit A.

□

[2]Note that in the definition of swap complexity (Definition 2), it was crucial that the ancilla qubits return to the all-0 state. Without that requirement, one can check that our construction of the distinguishing circuit in Figure 1 would fail, because the control qubit would be entangled with the ancilla qubits. By contrast, the circuit for the other direction of the proof (see Figure 2) is insensitive to the ancilla state. (We thank Daniel Gottesman for this observation.)

Theorem 1 is related to a known equivalence between Hamiltonian simulation and energy measurement [8]. In terms of [8], the swap unitary is a $\bar{\sigma}_x$ gate defined on the basis states $|\bar{0}\rangle = |x\rangle, |\bar{1}\rangle = |y\rangle$. Similarly, distinguishing between $|\phi\rangle$ and $|\psi\rangle$ is equivalent to an energy measurement by the Hamiltonian $\bar{\sigma}_x$ with precision $\Delta E = 1$. Indeed, the circuit in Figure 1 resembles the phase estimation (or energy measurement) circuit by Kitaev, Shen, and Vyalyi [9], with U as the simulation of the Hamiltonian $\bar{\sigma}_x$. Conversely, the circuit in Figure 2 is a simulation of the Hamiltonian $\bar{\sigma}_x$ using A, where the latter separates $\frac{|\bar{0}\rangle+|\bar{1}\rangle}{\sqrt{2}}$ from $\frac{|\bar{0}\rangle-|\bar{1}\rangle}{\sqrt{2}}$.

2.2 Imperfect case

Note that there are cases where U efficiently maps $|x\rangle$ to $|y\rangle$, and a corresponding $U^\dagger$ maps $|y\rangle$ to $|x\rangle$, but the same U doesn't do both. For example, $|x\rangle = |0^n\rangle$ and $|y\rangle = C|0^n\rangle$, where C is some random quantum circuit (see Section 4 for a comparison of swap complexity and relative state complexity). In such cases, a natural question arises: how well can the circuit C be used to distinguish $|\psi\rangle$ from $|\phi\rangle$?

More generally, we might wonder: if we have some "imperfect" or "partial" ability to swap $|x\rangle$ and $|y\rangle$ (for example, a unitary U such that $U|x\rangle = |y\rangle$ but $U|y\rangle \neq |x\rangle$), what does that imply about our ability to measure the relative phase in $\frac{|x\rangle\pm|y\rangle}{\sqrt{2}}$? Conversely, what does an imperfect ability to measure the relative phase imply about our ability to swap?

Our main result is as follows:

Theorem 2.

(i) Let $|x\rangle, |y\rangle$ be orthogonal n-qubit states, and suppose that $\langle y|U|x\rangle = a$ and $\langle x|U|y\rangle = b$. Then using a single black-box access to controlled-U, plus $O(1)$ additional gates, we can distinguish $|\psi\rangle = \frac{|x\rangle+|y\rangle}{\sqrt{2}}$ from $|\phi\rangle = \frac{|x\rangle-|y\rangle}{\sqrt{2}}$ with bias $\Delta = \frac{|a+b|}{2}$.

(ii) Let $|\psi\rangle, |\phi\rangle$ be orthogonal n-qubit states, and suppose the procedure A accepts $|\psi\rangle$ with probability p and $|\varphi\rangle$ with probability $p-\Delta$, i.e. A distinguishes $|\psi\rangle$ from $|\phi\rangle$ with bias Δ. Then using a single black-box access each to A and $A^\dagger$, and a single additional gate, we can apply a unitary U such that

$$\frac{|\langle y|U|x\rangle + \langle x|U|y\rangle|}{2} = \Delta$$

where $|x\rangle = \frac{|\psi\rangle+|\phi\rangle}{\sqrt{2}}$ and $|y\rangle = \frac{|\psi\rangle-|\phi\rangle}{\sqrt{2}}$.

Note that the parameters in the two parts of the theorem are equivalent. Given a, b in Theorem 2(i), we can distinguish $|\psi\rangle, |\phi\rangle$ with bias $\Delta = \frac{|a+b|}{2}$, while with the same distinguishibility bias Δ, we can create a swap with parameters $\widetilde{a}, \widetilde{b}$ such that $\frac{|\widetilde{a}+\widetilde{b}|}{2} = \Delta$.

Proof. For part (i), let U be as follows:

$$U|x\rangle = (a|y\rangle + c|x\rangle + f|w\rangle)$$
$$U|y\rangle = (b|x\rangle + d|y\rangle + g|z\rangle),$$

where $|w\rangle, |z\rangle$ are states orthogonal to both $|x\rangle$ and $|y\rangle$. We add a global phase to U, and denote it $\widetilde{U} = e^{i\theta}U$ (alternatively we initialize the ancilla qubit to $|0\rangle + e^{i\theta}|1\rangle$).

Using the same procedure as in Fig. 1 on the input $|\psi\rangle$, with $\widetilde{U}$, we get

$$\begin{aligned}\Pr(|+\rangle) &= \frac{1}{2} + \frac{1}{2}\mathrm{Re}\left(\langle\psi|\,\widetilde{U}\,|\psi\rangle\right) \\ &= \frac{1}{2} + \frac{1}{4}\mathrm{Re}\left[e^{i\theta}\left(\langle x| + \langle y|\right)\left(a\,|y\rangle + c\,|x\rangle + f\,|w\rangle + b\,|x\rangle + d\,|y\rangle + g\,|z\rangle\right)\right] \\ &= \frac{1}{2} + \mathrm{Re}\left(e^{i\theta}\cdot\frac{a+b+c+d}{4}\right),\end{aligned}$$

whereas on input $|\phi\rangle$, we get

$$\begin{aligned}\Pr(|+\rangle) &= \frac{1}{2} + \frac{1}{2}\mathrm{Re}\left(\langle\phi|\,\widetilde{U}\,|\phi\rangle\right) \\ &= \frac{1}{2} + \frac{1}{4}\mathrm{Re}\left[e^{i\theta}\left(\langle x| - \langle y|\right)\left(a\,|y\rangle + c\,|x\rangle + f\,|w\rangle - b\,|x\rangle - d\,|y\rangle - g\,|z\rangle\right)\right] \\ &= \frac{1}{2} + \mathrm{Re}\left(e^{i\theta}\cdot\frac{-a-b+c+d}{4}\right).\end{aligned}$$

The difference is $\left|\frac{\mathrm{Re}\left[e^{i\theta}(a+b)\right]}{2}\right|$, but we can improve it to $\frac{|a+b|}{2}$ by choosing $\theta = -\arg(a+b)$.

□

For part (ii), we use the same circuit as in Fig. 2. Let

$$\begin{aligned}A\,|\psi\rangle &= \sqrt{p}\,|1\rangle\,|\psi_1\rangle + \sqrt{1-p}\,|0\rangle\,|\psi_0\rangle \\ A\,|\phi\rangle &= \sqrt{1-p+\Delta}\,|0\rangle\,|\phi_0\rangle + \sqrt{p-\Delta}\,|1\rangle\,|\phi_1\rangle.\end{aligned}$$

Then,

$$A\,|x\rangle = \frac{1}{\sqrt{2}}\left[\sqrt{p}\,|1\rangle\,|\psi_1\rangle + \sqrt{1-p}\,|0\rangle\,|\psi_0\rangle + \sqrt{1-p+\Delta}\,|0\rangle\,|\phi_0\rangle + \sqrt{p-\Delta}\,|1\rangle\,|\phi_1\rangle\right],$$

which after a phase flip on the first qubit (Z_1) yields

$$Z_1 A\,|x\rangle = \frac{1}{\sqrt{2}}\left[-\sqrt{p}\,|1\rangle\,|\psi_1\rangle + \sqrt{1-p}\,|0\rangle\,|\psi_0\rangle + \sqrt{1-p+\Delta}\,|0\rangle\,|\phi_0\rangle - \sqrt{p-\Delta}\,|1\rangle\,|\phi_1\rangle\right].$$

Meanwhile,

$$A\,|y\rangle = \frac{1}{\sqrt{2}}\left[\sqrt{p}\,|1\rangle\,|\psi_1\rangle + \sqrt{1-p}\,|0\rangle\,|\psi_0\rangle - \sqrt{1-p+\Delta}\,|0\rangle\,|\phi_0\rangle - \sqrt{p-\Delta}\,|1\rangle\,|\phi_1\rangle\right].$$

The inner product is the following:

$$\begin{aligned}\langle y|\,A^\dagger Z_1 A\,|x\rangle = \langle x|\,A^\dagger Z_1 A\,|y\rangle^* = \frac{1}{2}\Big[&-p + (1-p) - (1-p+\Delta) + (p-\Delta) \\ &+ \sqrt{p(p-\Delta)}(\langle\phi_1|\psi_1\rangle - \langle\psi_1|\phi_1\rangle) \\ &+ \sqrt{(1-p)(1-p+\Delta)}(\langle\psi_0|\phi_0\rangle - \langle\phi_0|\psi_0\rangle)\Big].\end{aligned}$$

Hence,

$$\frac{\left|\langle y|\, A^\dagger Z_1 A\, |x\rangle + \langle x|\, A^\dagger Z_1 A\, |y\rangle\right|}{2} = \Delta.$$

□

One implication of Theorem 2(i) is that if $U|x\rangle = |y\rangle$, while $U|y\rangle$ is orthogonal to both $|x\rangle$ and $|y\rangle$, then we can distinguish $|\psi\rangle$ from $|\phi\rangle$ with bias $\frac{1}{2}$ (because $|a+b|=1$).

Another implication is that, if $U|x\rangle = |y\rangle$ but $U|y\rangle = -|x\rangle$, then $|a+b|=0$ and we get no distinguishing power at all by the method of Theorem 2(i). One might wonder: is this just an artifact of our proof, or are there actual examples of $|x\rangle$ and $|y\rangle$ that are easy to swap with a -1 phase, but exponentially hard to swap with any other phase (or equivalently, for which it's exponentially hard to distinguish $\frac{|x\rangle+|y\rangle}{\sqrt{2}}$ from $\frac{|x\rangle-|y\rangle}{\sqrt{2}}$)? Perhaps surprisingly, we will show in Section 3 that the answer is the latter.

We stress that the ability to swap $|x\rangle$ and $|y\rangle$ is *not* equivalent to the ability to distinguish $|x\rangle$ and $|y\rangle$ themselves. For example, let $|x\rangle = |0\ldots0\rangle$ and $|y\rangle = \sum_{j\neq 0}\alpha_j|j\rangle$, wherein $\{\alpha_j\}$ are arbitrary (e.g., $|y\rangle$ is a Haar-random state). Then it's trivial to distinguish $|x\rangle$ from $|y\rangle$, yet *mapping* $|x\rangle$ to $|y\rangle$ will in general be extremely difficult. Conversely, let $|\psi\rangle = \frac{|x\rangle+|y\rangle}{\sqrt{2}}$ and $|\phi\rangle = \frac{|x\rangle-|y\rangle}{\sqrt{2}}$. Then it's trivial to *map* $|\psi\rangle$ to $|\phi\rangle$ and vice versa. But we know, by Theorem 2, that *distinguishing* the two must in general be extremely difficult.

The general rule is that distinguishability in one basis implies "swappability" in a conjugate basis, and vice versa. (Note that Theorem 2 would also have worked with, e.g., $|\psi\rangle = \frac{|x\rangle+i|y\rangle}{\sqrt{2}}$ and $|\phi\rangle = \frac{|x\rangle-i|y\rangle}{\sqrt{2}}$, rather than $|\psi\rangle = \frac{|x\rangle+|y\rangle}{\sqrt{2}}$ and $|\phi\rangle = \frac{|x\rangle-|y\rangle}{\sqrt{2}}$.)

3 Tightness

A natural question about Theorem 2 is whether a better construction could yield better parameters. For example, given a U such that $U|x\rangle = |y\rangle$ and $U|y\rangle$ is orthogonal to both $|x\rangle$ and $|y\rangle$, could we use it to distinguish $|\psi\rangle = \frac{|x\rangle+|y\rangle}{\sqrt{2}}$ from $|\phi\rangle = \frac{|x\rangle-|y\rangle}{\sqrt{2}}$ *perfectly*, and with the same complexity?

We now prove that, in general, the parameters of Theorem 2 are optimal. Perhaps surprisingly, this optimality result does not depend on any unproved conjectures in complexity theory.

Theorem 3.

(i) For all $0 \leq b \leq a \leq 1$, there exists an n-qubit U implemented by a size-$O(1)$ circuit, as well as states $|x\rangle, |y\rangle$, such that $\langle y|\, U\, |x\rangle = a$ and $\langle x|\, U\, |y\rangle = b$, and yet if $\langle y|\, V\, |x\rangle = a'$ and $\langle x|\, V\, |y\rangle = b'$ where $|a'+b'| \geq |a+b| + \omega(2^{-n/3}\sqrt{\log n})$, then V requires a size-$\omega(2^{n/3})$ circuit.

(ii) For all $\Delta \in [0,1]$, there exist two n-qubit states $|\psi\rangle, |\phi\rangle$ that can be distinguished with bias Δ by a size-$O(1)$ circuit, yet such that distinguishing them with bias $\Delta + \omega(2^{-n/3}\sqrt{\log n})$ requires a size-$\omega(2^{n/3})$ circuit.

We assume here that the size of the universal set of gates is polynomial in the number of qubits.

Proof. By our main result, Theorem 2, we only need to prove part (i). Part (ii) then follows automatically, if we set $|\psi\rangle = \frac{|x\rangle+|y\rangle}{\sqrt{2}}$ and $|\phi\rangle = \frac{|x\rangle-|y\rangle}{\sqrt{2}}$ and $\Delta = \frac{|a+b|}{2}$.

Let $|\eta_0\rangle, \ldots, |\eta_7\rangle$ be n-qubit states, whose pairwise swap complexity is exponential. For example, let them be Haar-random; then by a counting argument, all the pairwise swap complexities will clearly be exponential with overwhelming probability.

We add a 3-qubit index register to each $|\eta_k\rangle$, and write the entire state as $\left|\bar{k}\right\rangle \triangleq |k\rangle \otimes |\eta_k\rangle$. Consider the following construction:

$$\begin{aligned}
|x\rangle &= \sqrt{a-b}\left(\frac{|\bar{0}\rangle + |\bar{1}\rangle + |\bar{2}\rangle + |\bar{3}\rangle}{2}\right) + \sqrt{b}\left(\frac{|\bar{4}\rangle + |\bar{5}\rangle}{\sqrt{2}}\right) + \sqrt{c}\,|\bar{6}\rangle \\
|y\rangle &= \sqrt{a-b}\left(\frac{|\bar{0}\rangle + i\,|\bar{1}\rangle - |\bar{2}\rangle - i\,|\bar{3}\rangle}{2}\right) + \sqrt{b}\left(\frac{|\bar{4}\rangle - |\bar{5}\rangle}{\sqrt{2}}\right) + \sqrt{c}\,|\bar{7}\rangle\,. \qquad (1) \\
U &= \left(\sum_{k=0}^{3} i^k \left|\bar{k}\right\rangle\left\langle\bar{k}\right| + |\bar{4}\rangle\langle\bar{4}| - |\bar{5}\rangle\langle\bar{5}| + |\bar{6}\rangle\langle\bar{6}| + |\bar{7}\rangle\langle\bar{7}|\right) \otimes \mathbb{1}_{2^n},
\end{aligned}$$

wherein by normalization, $\langle x|x\rangle = a + c = 1$. To understand the construction, note that U transfers the superposition of the first four states of $|x\rangle$ to the corresponding superposition in $|y\rangle$. In contrast, it transfers the superposition of the first four states in $|y\rangle$ to a superposition orthogonal to both $|x\rangle$ and $|y\rangle$. Next, U applies $|\bar{4}\rangle + |\bar{5}\rangle \longleftrightarrow |\bar{4}\rangle - |\bar{5}\rangle$, and finally, U does not affect the states $|\bar{6}\rangle$ and $|\bar{7}\rangle$.

U can be implemented using $O(1)$ gates since it acts only on the index register. Furthermore, it is easy to verify that $\langle y|\,U\,|x\rangle = a$ and $\langle x|\,U\,|y\rangle = b$.

We will need the following lemma, proved in Appendix A.

Lemma 1. Let $|\eta_0\rangle, |\eta_1\rangle$ be two n-qubit Haar-random states, and let $g = n^{O(1)}$ be the size of a universal set of gates G. Then with $1-\exp(-\exp(n))$ probability over $|\eta_0\rangle, |\eta_1\rangle$, there is no circuit C with $M = O(2^{n/3})$ gates from G, such that $|\langle\eta_0|\,C\,|\eta_1\rangle| \geq \varepsilon$, where $\varepsilon \leq \sqrt{M \log g/N} = O(2^{-n/3}\sqrt{\log n})$.

By Lemma 1, due to the pairwise swap-complexity of the $\{|\eta_k\rangle\}$ states, any unitary $\widetilde{U}$, implemented by $O(2^{n/3})$ gates, with overwhelming probability cannot transform $\left|\bar{k}\right\rangle$ into anything close to $\left|\bar{k}'\right\rangle$, for any $k \neq k'$. Hence, the principal submatrix of $\widetilde{U}$ when removing all columns except those corresponding to the $\left\{\left|\bar{k}\right\rangle\right\}$ states is necessarily in the following form:

$$\widetilde{U}\Big|_{\{|\bar{k}\rangle\}} = \sum_{k=0}^{7} \beta_k e^{i\theta_k} \left|\bar{k}\right\rangle\left\langle\bar{k}\right| + \widetilde{O}(2^{-n/3}),$$

wherein $\beta_k \in [0,1]$, and $\widetilde{O}$ is a big-O notation which ignores logarithmic factors. Cal-

culating $\widetilde{a}+\widetilde{b}$,

$$\widetilde{a} = \langle y| \widetilde{U} |x\rangle = \frac{a-b}{4}\sum_{k=0}^{3}\beta_k(-i)^k e^{i\theta_k} + \frac{b}{2}(\beta_4 e^{i\theta_4} - \beta_5 e^{i\theta_5}) + \widetilde{O}(2^{-n/3}),$$

$$\widetilde{b} = \langle x| \widetilde{U} |y\rangle = \frac{a-b}{4}\sum_{k=0}^{3}\beta_k i^k e^{i\theta_k} + \frac{b}{2}(\beta_4 e^{i\theta_4} - \beta_5 e^{i\theta_5}) + \widetilde{O}(2^{-n/3}).$$

By carefully choosing θ_k, and taking $\beta_k = 1$, we upper-bound the absolute sum:

$$\left|\widetilde{a}+\widetilde{b}\right| = \left|\frac{a-b}{2}\left(\beta_0 e^{\theta_0} - \beta_2 e^{i\theta_2}\right) + b\left(\beta_4 e^{i\theta_4} - \beta_5 e^{i\theta_5}\right)\right| + \widetilde{O}(2^{-n/3}) \leq |a+b| + \widetilde{O}(2^{-n/3}),$$

which proves the theorem.

□

Theorem 3 implies, in particular, that there exist cases where we can efficiently map $|x\rangle$ to $|y\rangle$, but only via a circuit that also maps $|y\rangle$ to $-|x\rangle$, and where eliminating the -1 factor requires an exponentially larger circuit. In these cases, and *only* in these cases, we get efficient mapping between $|x\rangle$ and $|y\rangle$, without any corresponding ability to distinguish $\frac{|x\rangle+|y\rangle}{\sqrt{2}}$ from $\frac{|x\rangle-|y\rangle}{\sqrt{2}}$ efficiently.

Having said that, in the *specific* case of Schrödinger's cat, if we have the ability to map $|\text{Alive}\rangle$ to $|\text{Dead}\rangle$ and $|\text{Dead}\rangle$ to $-|\text{Alive}\rangle$, then we also have the ability to swap the $|\text{Alive}\rangle$ and $|\text{Dead}\rangle$ states without the -1 relative phase. The reason is that it's easy enough to *distinguish* a live cat from a dead one, so we could simply correct the phase after applying the swap, conditional on being in the $|\text{Alive}\rangle$ state. We thus see that, in the proof Theorem 3, it was crucial to consider pairs of states that are exponentially hard not only to swap, but *also* to distinguish from each other. (We thank Ed Witten for this observation.)

4 Relative state complexity vs. swap complexity

The following corollary summarizes the relation between the circuit complexity of swapping two states, their relative complexity and their absolute state complexity.

Corollary 1. *Consider two orthogonal states $|x\rangle, |y\rangle$, and let $|\psi\rangle = \frac{|x\rangle+|y\rangle}{2}$ and $|\phi\rangle = \frac{|x\rangle-|y\rangle}{2}$. Then,*

$$\mathcal{C}(|x\rangle, |y\rangle) \leq \mathcal{S}(|x\rangle, |y\rangle) \leq \min\left[\mathcal{C}(|\psi\rangle, |0\ldots0\rangle), \mathcal{C}(|\phi\rangle, \ |0\ldots0\rangle)\right]$$

(ignoring constant factors and ε the subscripts of $\mathcal{C}, \mathcal{S}$). The separation of the inequalities can be exponential.

Proof. The first inequality is trivial, since swapping is at least as hard as mapping. For the second inequality, recall by Theorem 2 that $\mathcal{S}(|x\rangle, |y\rangle)$ is at most the complexity

of distinguishing $|\psi\rangle$ from $|\phi\rangle$. Let A be a minimal circuit to prepare $|\psi\rangle$ from $|0^n\rangle$. Then to distinguish $|\psi\rangle$ from $|\phi\rangle$, we simply apply $A^\dagger$ to $|\psi\rangle$, and check whether we got back to $|0^n\rangle$ (and similarly given a minimal circuit to prepare $|\phi\rangle$).

For an exponential separation of the first inequality, consider Equation 1 with $b = c = 0$. The unitary U transfers $|x\rangle$ to $|y\rangle$ efficiently, hence $\mathcal{C}(|x, y\rangle) = O(1)$. On the other hand, $\mathcal{S}(|x, y\rangle)$ can be exponential due to the tightness theorem (Theorem 3). For an exponential separation of the second inequality, consider $|x\rangle = |0\rangle |\eta\rangle$ and $|y\rangle = |1\rangle |\eta\rangle$ where $|\eta\rangle$ is a Haar-random state. Then $\mathcal{S}(|x\rangle, |y\rangle) = O(1)$, but preparing either $|x\rangle$ or $|y\rangle$ requires an exponentially large complexity with overwhelming probability. □

Interestingly, while relative state complexity is a metric, swap complexity is not. Swap complexity is a "semimetric": it's symmetric and reflexive, but does not satisfy the triangle inequality, as shown by the following counterexample.

Consider the following 3 states:

$$|x\rangle = |000\rangle \qquad |y\rangle = |1--\rangle \qquad |z\rangle = |011\rangle, \tag{2}$$

and the universal set of gates: Hadamard, NOT, CNOT and a phase gate $R_\phi = |0\rangle\langle 0| + e^{i\phi}|1\rangle\langle 1|$, with $\phi \ll 1$. It is easy to see that

$$\mathcal{S}_0(|x\rangle, |z\rangle) = 2 \qquad \mathcal{S}_0(|z\rangle, |y\rangle) = 3, \tag{3}$$

where $\mathcal{S}_0$ is the complexity of swapping $|x\rangle$ and $|y\rangle$ with zero error. Our exhaustive search found that the smallest circuit for perfectly swapping x and y is of size 7 (see Figure 3).

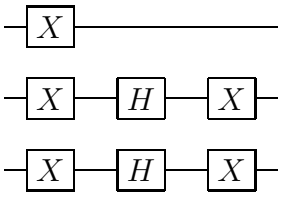

Figure 3: A 7-gate circuit to swap $|000\rangle$ and $|1--\rangle$.

Hence,

$$\mathcal{S}_0(|x\rangle, |y\rangle) > \mathcal{S}_0(|x\rangle, |z\rangle) + \mathcal{S}_0(|z\rangle, |y\rangle) \tag{4}$$

Note that swap complexity does satisfy the triangle inequality in the special case where $|x\rangle, |y\rangle, |z\rangle$ are all computational basis states.

5 Discussion

By using quantum circuit complexity, we were able to formalize a folklore observation in the foundations of quantum mechanics: namely, that the ability to measure the coherence in Schrödinger's cat is somehow related to the ability to bring a dead cat back

to life. We were also able to articulate in precisely what circumstances that folklore observation would become false. Our results inspired a more general investigation of *swap complexity* of pairs of quantum states, which is related to their relative complexity but can be exponentially greater, and which might be independent interest.

Our equivalence theorem has some interesting implications for physics. For example, if we have a superposition of a state $|x\rangle$ of polynomial complexity and a state $|y\rangle$ of exponential complexity, then no polynomial-time experiment can ever detect the relative phase between $|x\rangle$ and $|y\rangle$. (For otherwise, we could efficiently *map* $|x\rangle$ to $|y\rangle$!)

In a previous work [10], Aaronson and Susskind proved that evolving the state

$$|\psi_0\rangle = \frac{1}{\sqrt{2^n}} \sum_{j \in \{0,1\}^n} |j\rangle \otimes |j\rangle$$

by a "generic" (computationally universal) Hamiltonian H for exponential time yields a state with superpolynomial circuit complexity unless PSPACE $\subset$ PP/POLY. Combining that result with our Theorem 2 and Corollary 1 means that unless PSPACE $\subset$ PP/POLY, there can be no feasible experiment, in general, to measure the phase between a state and same state after being evolved for exponential time. Even if mapping one state to the other is merely "thermodynamics-hard," in the sense that it's hard to unscramble an egg, still, distinguishing the superposition from the incoherent mixture with any non-negligible bias would be thermodynamics-hard as well.

One might wonder about the apparent symmetry of our results in the case of Schödinger's cat, since reviving a cat seems so much harder than taking its life. However note that in this work, both $|\text{Alive}\rangle$ and $|\text{Dead}\rangle$ are taken to determine the exact states of every atom of the cat. If we accounted for other possible "alive" and "dead" states, then of course we expect many more configurations of dead cats than alive cats, so thermodynamics suffices to explain why killing a cat is so much easier than reviving one.

6 Acknowledgments

We thank Edward Witten for the comment at the end of Section 3, Daniel Gottesman for the note in the proof of Theorem 1, Yosi Avron for pointing us to the question of symmetry of swap complexity and relative state complexity in the discussion, and Henry Yuen for catching some errors in an earlier draft.

References

[1] Kyriakos Papadodimas and Suvrat Raju. Black hole interior in the holographic correspondence and the information paradox. *Physical Review Letters*, 112(5):051301, 2014.

[2] Daniel Harlow. Aspects of the papadodimas-raju proposal for the black hole interior. *Journal of High Energy Physics*, 2014(11):55, 2014.

[3] Adam R Brown and Leonard Susskind. Second law of quantum complexity. *Physical Review D*, 97(8):086015, 2018.

[4] Leonard Susskind. Three lectures on complexity and black holes. *arXiv preprint arXiv:1810.11563*, 2018.

[5] Yakir Aharonov and Daniel Rohrlich. *Quantum paradoxes: quantum theory for the perplexed*. John Wiley & Sons, 2008.

[6] Eugene P Wigner. Remarks on the mind-body question. In *Philosophical reflections and syntheses*, pages 247–260. Springer, 1995.

[7] Ludvik Bass. The mind of wigner's friend. *Hermathena*, pages 52–68, 1971.

[8] Yosi Atia and Dorit Aharonov. Fast-forwarding of hamiltonians and exponentially precise measurements. *Nature communications*, 8(1):1572, 2017.

[9] Alexei Yu Kitaev, Alexander Shen, and Mikhail N Vyalyi. *Classical and quantum computation*. Number 47. American Mathematical Soc., 2002.

[10] Scott Aaronson. The complexity of quantum states and transformations: From quantum money to black holes. *arXiv preprint arXiv:1607.05256*, 2016.

[11] Scott Aaronson and Greg Kuperberg. Quantum versus classical proofs and advice. In *Twenty-Second Annual IEEE Conference on Computational Complexity (CCC'07)*, pages 115–128. IEEE, 2007.

A Appendix: Proof of Lemma 1

Lemma 1. Let $|\eta_0\rangle, |\eta_1\rangle$ be two n-qubit Haar-random states, and let $g = n^{O(1)}$ be the size of a universal set of gates G. Then with $1-\exp(-\exp(n))$ probability over $|\eta_0\rangle, |\eta_1\rangle$, there is no circuit C with $M = O(2^{n/3})$ gates from G such that $|\langle\eta_0| C |\eta_1\rangle| \geq \varepsilon$, where $\varepsilon \leq \sqrt{M \log g/N} = O(2^{-n/3}\sqrt{\log n})$.

Proof. We use a simple counting argument. Starting at $|\eta_0\rangle$, a circuit with M gates taken from a universal set of gates of size $g = n^{O(1)}$ reaches at most $O(g^M)$ different states $\{|\gamma_j\rangle\}$. The following fact yields the probability of $|\langle\eta_0|\gamma_j\rangle| \geq \varepsilon$ for a specific j.

Fact 1 (see Lemma 3.6 in [11])**.** *Let $|\psi\rangle$ be a Haar-random state of dimension N. Then for any $\varepsilon > 0$,*

$$\Pr\left(|\langle\psi|0\ldots0\rangle| \geq \varepsilon\right) = (1-\varepsilon^2)^{N-1}. \tag{5}$$

By the union bound, the probability that $|\eta_1\rangle$ has an overlap at least ε with any of the states $\{|\gamma_j\rangle\}$ is at most

$$g^M(1-\varepsilon^2)^{N-1} \leq g^M e^{-\varepsilon^2(N-1)} \leq 2^{M\log(g)-\log(e)\cdot M\log(g)(1-1/N)} \leq 2^{M\log(g)\frac{1-\log(e)}{2}}, \tag{6}$$

where the first inequality comes from $(1+x) \leq e^x$. The exponent in the last expression is of order $-2^{n/3}\log n$. Hence, the probability of any $|\gamma_j\rangle$ to have at least ε overlap with $|\eta_1\rangle$ is doubly-exponentially small. □