Quantum Computing and Hidden Variables

Scott Aaronson*
Institute for Advanced Study, Princeton

This paper initiates the study of hidden variables from a quantum computing perspective. For us, a hidden-variable theory is simply a way to convert a unitary matrix that maps one quantum state to another, into a stochastic matrix that maps the initial probability distribution to the final one in some fixed basis. We list five axioms that we might want such a theory to satisfy, and then investigate which of the axioms can be satisfied simultaneously. Toward this end, we propose a new hidden-variable theory based on network flows. In a second part of the paper, we show that if we could examine the entire history of a hidden variable, then we could efficiently solve problems that are believed to be intractable even for quantum computers. In particular, under any hidden-variable theory satisfying a reasonable axiom, we could solve the Graph Isomorphism problem in polynomial time, and could search an N-item database using $O\left(N^{1/3}\right)$ queries, as opposed to $O\left(N^{1/2}\right)$ queries with Grover's search algorithm. On the other hand, the $N^{1/3}$ bound is optimal, meaning that we could probably *not* solve NP-complete problems in polynomial time. We thus obtain the first good example of a model of computation that appears *slightly* more powerful than the quantum computing model.

PACS numbers: 03.65.Ta, 03.65.Ud, 03.67.Lx, 02.70.-c

I. INTRODUCTION

Quantum mechanics lets us calculate the probability that (say) an electron will be found in an excited state if measured at a particular time. But it is silent about *multiple-time* or *transition* probabilities: that is, what is the probability that the electron will be in an excited state at time t_1, given that it was in its ground state at an earlier time t_0? The usual response is that this question is meaningless, unless of course the electron was *measured* (or otherwise known with probability 1) to be in its ground state at t_0. A different response—pursued by Schrödinger [1], Bohm [2], Bell [3], Nelson [4], Dieks [5], and others—treats the question as provisionally meaningful, and then investigates how one might answer it mathematically. Specific attempts at answers are called "hidden-variable theories."

The appeal of hidden-variable theories is that they provide one possible solution to the measurement problem. For they allow us to apply unitary quantum mechanics to the entire universe (including ourselves), yet still discuss the probability of a future observation conditioned on our current observations. Furthermore, they let us do so without making any assumptions about decoherence or the nature of observers. For example, even if an observer were placed in coherent superposition, that observer would still have a sequence of definite experiences, and the probability of any such sequence could be calculated.

This paper initiates the study of hidden variables from a quantum computing perspective. We restrict our attention to the simplest possible setting: that of discrete time, a finite-dimensional Hilbert space, and a fixed orthogonal basis. Within this setting, we reformulate known hidden-variable theories due to Dieks [5] and Schrödinger [1], and also introduce a new theory based on network flows. However, a more important contribution is the *axiomatic approach* that we use. We propose five axioms for hidden-variable theories in our setting, and then compare theories against each other based on which of the axioms they satisfy. A central question in our approach is which subsets of axioms can be satisfied simultaneously.

In a second part of the paper, we make the connection to quantum computing explicit, by studying the computational complexity of simulating hidden-variable theories. Below we describe our computational results.

A. The Complexity of Sampling Histories

It is often stressed that hidden-variable theories yield exactly the same predictions as ordinary quantum mechanics. On the other hand, these theories describe a different picture of physical reality, with an additional layer of dynamics

*Electronic address: `aaronson@ias.edu`

beyond that of a state vector evolving unitarily. We address a question that, to our knowledge, had never been raised before: *what is the computational complexity of simulating that additional dynamics?* In other words, if we could examine a hidden variable's entire history, then could we solve problems in polynomial time that are intractable even for quantum computers?

We present strong evidence that the answer is yes. The Graph Isomorphism problem asks whether two graphs G and H are isomorphic; while given a basis for a lattice $\mathcal{L} \in \mathbb{R}^n$, the Approximate Shortest Vector problem asks for a nonzero vector in $\mathcal{L}$ within a $\sqrt{n}$ factor of the shortest one. We show that both problems are efficiently solvable by sampling a hidden variable's history, provided the hidden-variable theory satisfies the indifference axiom. By contrast, despite a decade of effort, neither problem is known to lie in BQP, the class of problems solvable in quantum polynomial time with bounded error probability.[40] Thus, if we let DQP (Dynamical Quantum Polynomial-Time) be the class of problems solvable in our new model, then this already provides circumstantial evidence that BQP is strictly contained in DQP.

However, the evidence is stronger than this. For we actually show that DQP contains an entire *class* of problems, of which Graph Isomorphism and Approximate Shortest Vector are special cases. Computer scientists know this class as *Statistical Zero Knowledge*, or SZK. Furthermore, in previous work [6] we showed that "relative to an oracle," SZK is not contained in BQP. This is a technical concept implying that any proof of $\mathsf{SZK} \subseteq \mathsf{BQP}$ would require techniques unlike those that are currently known. Combining our result that $\mathsf{SZK} \subseteq \mathsf{DQP}$ with the oracle separation of [6], we obtain that $\mathsf{BQP} \neq \mathsf{DQP}$ relative to an oracle as well. Given computer scientists' longstanding inability to separate basic complexity classes, this is nearly the best evidence one could hope for that sampling histories yields more power than standard quantum computation.

Besides solving SZK problems, we also show that by sampling histories, one could search an unordered database of N items for a single "marked item" using only $O\left(N^{1/3}\right)$ database queries. By comparison, Grover's quantum search algorithm [7] requires $\Theta\left(N^{1/2}\right)$ queries, while classical algorithms require $\Theta\left(N\right)$ queries.[41] On the other hand, we also show that our $N^{1/3}$ upper bound is the best possible—so even in the histories model, one cannot search an N-item database in $(\log N)^c$ steps for some fixed power c. This implies that $\mathsf{NP} \not\subset \mathsf{DQP}$ relative to an oracle, which in turn suggests that DQP is *still* not powerful enough to solve NP-complete problems in polynomial time. Note that while Graph Isomorphism and Approximate Shortest Vector are in NP, it is strongly believed that they are not NP-complete.

At this point we should address a concern that many readers will have. Once we extend quantum mechanics by positing the "unphysical" ability to sample histories, isn't it completely unsurprising if we can then solve problems that were previously intractable? We believe the answer is no, for three reasons.

First, almost every change that makes the quantum computing model more powerful, seems to make it *so much* more powerful that NP-complete and even harder problems become solvable efficiently. To give some examples, NP-complete problems can be solved in polynomial time using a nonlinear Schrödinger equation, as shown by Abrams and Lloyd [8]; using closed timelike curves, as shown by Brun [9] and Bacon [10] (and conjectured by Deutsch [11]); or using a measurement rule of the form $|\psi|^p$ for any $p \neq 2$, as shown by us [12]. It is also easy to see that we could solve NP-complete problems if, given a quantum state $|\psi\rangle$, we could request a classical description of $|\psi\rangle$, such as a list of amplitudes or a preparation procedure.[42] By contrast, ours is the first independently motivated model we know of that seems more powerful than quantum computing, but only *slightly* so.[43] Moreover, the striking fact that unordered search in our model takes about $N^{1/3}$ steps, as compared to N steps classically and $N^{1/2}$ quantum-mechanically, suggests that DQP somehow "continues a sequence" that begins with P and BQP. It would be interesting to find a model in which search takes $N^{1/4}$ or $N^{1/5}$ steps.

The second reason our results are surprising is that, given a hidden variable, the distribution over its possible values at any *single* time is governed by standard quantum mechanics, and is therefore efficiently samplable on a quantum computer. So if examining the variable's history confers any extra computational power, then it can only be because of *correlations* between the variable's values at different times.

The third reason is our criterion for success. We are not saying merely that one can solve Graph Isomorphism under *some* hidden-variable theory; or even that, under any theory satisfying the indifference axiom, there exists an algorithm to solve it; but rather that there exists a *single* algorithm that solves Graph Isomorphism under any theory satisfying indifference. Thus, we must consider even theories that are specifically designed to thwart such an algorithm.

But what is the motivation for our results? The first motivation is that, within the community of physicists who study hidden-variable theories such as Bohmian mechanics, there is great interest in actually *calculating* the hidden-variable trajectories for specific physical systems [13, 14]. Our results show that, when many interacting particles are involved, this task might be fundamentally intractable, even if a quantum computer were available. The second motivation is that, in classical computer science, studying "unrealistic" models of computation has often led to new insights into realistic ones; and likewise we expect that the DQP model could lead to new results about standard quantum computation. Indeed, in a sense this has already happened. For our result that $\mathsf{SZK} \not\subset \mathsf{BQP}$ relative to an oracle [6] grew out of work on the BQP versus DQP question. Yet the "quantum lower bound for the collision problem" underlying that result provided the first evidence that cryptographic hash functions could be secure against quantum attack, and ruled out a large class of possible quantum algorithms for Graph Isomorphism and related problems.

B. Outline of Paper

Sections II through V B develop our axiomatic approach to hidden variables; then Sections VI through IX study the computational complexity of sampling hidden-variable histories.

Section II formally defines hidden-variable theories in our sense; then Section II A contrasts these theories with related ideas such as Bohmian mechanics and modal interpretations. Section II B addresses the most common objections to our approach: for example, that the implicit dependence on a fixed basis is unacceptable.

In Section III, we introduce five possible axioms for hidden-variable theories. These are indifference to the identity operation; robustness to small perturbations; commutativity with respect to spacelike-separated unitaries; commutativity for the special case of product states; and invariance under decomposition of mixed states into pure states. Ideally, a theory would satisfy all of these axioms. However, we show in Section IV that no theory satisfies both indifference and commutativity; no theory satisfies both indifference and a stronger version of robustness; no theory satisfies indifference, robustness, and decomposition invariance; and no theory satisfies a stronger version of decomposition invariance.

In Section V we shift from negative to positive results. Section V A presents a hidden-variable theory called the *flow theory* or $\mathcal{FT}$, which is based on the Max-Flow-Min-Cut theorem from combinatorial optimization. The idea is to define a network of "pipes" from basis states at an initial time to basis states at a final time, and then route as much probability mass as possible through these pipes. The capacity of each pipe depends on the corresponding entry of the unitary acting from the initial to final time. To find the probability of transitioning from basis state $|i\rangle$ to basis state $|j\rangle$, we then determine how much of the flow originating at $|i\rangle$ is routed along the pipe to $|j\rangle$. Our main results are that $\mathcal{FT}$ is well-defined and that it is robust to small perturbations. Since $\mathcal{FT}$ trivially satisfies the indifference axiom, this implies that the indifference and robustness axioms can be satisfied simultaneously, which was not at all obvious *a priori*.

Section V B presents a second theory that we call the *Schrödinger theory* or $\mathcal{ST}$, since it is based on a pair of integral equations introduced in a 1931 paper of Schrödinger [1]. Schrödinger conjectured, but was unable to prove, the existence and uniqueness of a solution to these equations; the problem was not settled until the work of Nagasawa [15] in the 1980's. In our discrete setting the problem is simpler, and we give a self-contained proof of existence using a matrix scaling technique due to Sinkhorn [16]. The idea is as follows: we want to convert a unitary matrix that maps one quantum state to another, into a nonnegative matrix whose i^{th} column sums to the initial probability of basis state $|i\rangle$, and whose j^{th} row sums to the final probability of basis state $|j\rangle$. To do so, we first replace each entry of the unitary matrix by its absolute value, then normalize each column to sum to the desired initial probability, then normalize each row to sum to the desired final probability. But then the columns are no longer normalized correctly, so we normalize them *again*, then normalize the rows again, and so on. We show that this iterative process converges, from which it follows that $\mathcal{ST}$ is well-defined. We also show that $\mathcal{ST}$ satisfies the indifference and product commutativity axioms, and violates the decomposition invariance axiom. We conjecture that $\mathcal{ST}$ satisfies the robustness axiom; proving that conjecture is one of the main open problems of the paper.

In Section VI we shift our attention to the complexity of sampling histories. We formally define DQP as the class of problems solvable by a classical polynomial-time algorithm with access to a "history oracle." Given a sequence of quantum circuits as input, this oracle returns a sample from a corresponding distribution over histories of a hidden variable, according to some hidden-variable theory $\mathcal{T}$. The oracle can choose $\mathcal{T}$ "adversarially," subject to the constraint that $\mathcal{T}$ satisfies the indifference and robustness axioms. Thus, a key result from Section VI that we rely on is that there *exists* a hidden-variable theory satisfying indifference and robustness.

Section VI A establishes the most basic facts about DQP: for example, that $\mathsf{BQP} \subseteq \mathsf{DQP}$, and that DQP is independent of the choice of gate set. Then Section VII presents the "juggle subroutine," a crucial ingredient in both of our main hidden-variable algorithms. Given a state of the form $(|a\rangle + |b\rangle)/\sqrt{2}$ or $(|a\rangle - |b\rangle)/\sqrt{2}$, the goal of this subroutine is to "juggle" a hidden variable between $|a\rangle$ and $|b\rangle$, so that when we inspect the hidden variable's history, both $|a\rangle$ and $|b\rangle$ are observed with high probability. The difficulty is that this needs to work under *any* indifferent hidden-variable theory.

Next, Section VIII combines the juggle subroutine with a technique of Valiant and Vazirani [17] to prove that $\mathsf{SZK} \subseteq \mathsf{DQP}$, from which it follows in particular that Graph Isomorphism and Approximate Shortest Vector are in DQP. Then Section IX applies the juggle subroutine to search an N-item database in $O\left(N^{1/3}\right)$ queries, and also proves that this $N^{1/3}$ bound is optimal.

We conclude in Section X with some directions for further research.

II. HIDDEN-VARIABLE THEORIES

Suppose we have an $N \times N$ unitary matrix U, acting on a state

$$|\psi\rangle = \alpha_1 |1\rangle + \cdots + \alpha_N |N\rangle ,$$

where $|1\rangle , \ldots , |N\rangle$ is a standard orthogonal basis. Let

$$U |\psi\rangle = \beta_1 |1\rangle + \cdots + \beta_N |N\rangle .$$

Then can we construct a stochastic matrix S, which maps the vector of probabilities

$$\overrightarrow{p} = \begin{bmatrix} |\alpha_1|^2 \\ \vdots \\ |\alpha_N|^2 \end{bmatrix}$$

induced by measuring $|\psi\rangle$, to the vector

$$\overrightarrow{q} = \begin{bmatrix} |\beta_1|^2 \\ \vdots \\ |\beta_N|^2 \end{bmatrix}$$

induced by measuring $U |\psi\rangle$? Trivially yes. The following matrix maps *any* vector of probabilities to $\overrightarrow{q}$, ignoring the input vector $\overrightarrow{p}$ entirely:

$$S_{\mathcal{PT}} = \begin{bmatrix} |\beta_1|^2 & \cdots & |\beta_1|^2 \\ \vdots & & \vdots \\ |\beta_N|^2 & \cdots & |\beta_N|^2 \end{bmatrix} .$$

Here $\mathcal{PT}$ stands for *product theory*. The product theory corresponds to a strange picture of physical reality, in which memories and records are completely unreliable, there being no causal connection between states of affairs at earlier and later times.

So we would like S to depend on U itself somehow, not just on $|\psi\rangle$ and $U |\psi\rangle$. Indeed, ideally S would be a function *only* of U, and not of $|\psi\rangle$. But this is impossible, as the following example shows. Let U be a $\pi/4$ rotation, and let $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ and $|-\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$. Then $U |+\rangle = |1\rangle$ implies that

$$S(|+\rangle , U) = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} ,$$

whereas $U |-\rangle = |0\rangle$ implies that

$$S(|-\rangle , U) = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} .$$

On the other hand, it is easy to see that, if S can depend on $|\psi\rangle$ as well as U, then there are infinitely many choices for the function $S(|\psi\rangle , U)$. Every choice reproduces the predictions of quantum mechanics perfectly when restricted to single-time probabilities. So how can we possibly choose among them? Our approach in Sections III and V will be to write down axioms that we would like S to satisfy, and then investigate which of the axioms can be satisfied simultaneously.

Formally, a *hidden-variable theory* is a family of functions $\{S_N\}_{N \geq 1}$, where each S_N maps an N-dimensional mixed state ρ and an $N \times N$ unitary matrix U onto a singly stochastic matrix $S_N(\rho, U)$. We will often suppress the dependence on N, ρ, and U, and occasionally use subscripts such as $\mathcal{PT}$ or $\mathcal{FT}$ to indicate the theory in question. Also, if $\rho = |\psi\rangle \langle\psi|$ is a pure state we may write $S(|\psi\rangle , U)$ instead of $S(|\psi\rangle \langle\psi| , U)$.

Let $(M)_{ij}$ denote the entry in the i^{th} column and j^{th} row of matrix M. Then $(S)_{ij}$ is the probability that the hidden variable takes value $|j\rangle$ after U is applied, conditioned on it taking value $|i\rangle$ before U is applied. At a minimum, any theory must satisfy the following marginalization axiom: for all $j \in \{1, \ldots , N\}$,

$$\sum_i (S)_{ij} (\rho)_{ii} = \left(U \rho U^{-1}\right)_{jj}.$$

This says that after U is applied, the hidden variable takes value $|j\rangle$ with probability $\left(U \rho U^{-1}\right)_{jj}$, which is the usual Born probability.

Often it will be convenient to refer, not to S itself, but to the matrix $P(\rho, U)$ of joint probabilities whose (i,j) entry is $(P)_{ij} = (S)_{ij} (\rho)_{ii}$. The i^{th} column of P must sum to $(\rho)_{ii}$, and the j^{th} row must sum to $\left(U \rho U^{-1}\right)_{jj}$. Indeed, we will define the theories $\mathcal{FT}$ and $\mathcal{ST}$ by first specifying the matrix P, and then setting $(S)_{ij} := (P)_{ij} / (\rho)_{ii}$. This approach has the drawback that if $(\rho)_{ii} = 0$, then the i^{th} column of S is undefined. To get around this, we adopt the convention that

$$S(\rho, U) := \lim_{\varepsilon \to 0^+} S(\rho_\varepsilon, U)$$

where $\rho_\varepsilon = (1 - \varepsilon) \rho + \varepsilon I$ and I is the $N \times N$ maximally mixed state. Technically, the limits

$$\lim_{\varepsilon \to 0^+} \frac{(P(\rho_\varepsilon, U))_{ij}}{(\rho_\varepsilon)_{ii}}$$

might not exist, but in the cases of interest to us it will be obvious that they do.

A. Comparison with Previous Work

Before going further, we should contrast our approach with previous approaches to hidden variables, the most famous of which is Bohmian mechanics [2]. Our main criticism of Bohmian mechanics is that it commits itself to a Hilbert space of particle positions and momenta. Furthermore, it is crucial that the positions and momenta be *continuous*, in order for particles to evolve deterministically. To see this, let $|L\rangle$ and $|R\rangle$ be discrete positions, and suppose a particle is in state $|L\rangle$ at time t_0, and state $(|L\rangle + |R\rangle)/\sqrt{2}$ at a later time t_1. Then a hidden variable representing the position would have entropy 0 at t_1, since it is always $|L\rangle$ then; but entropy 1 at t_1, since it is $|L\rangle$ or $|R\rangle$ both with 1/2 probability. Therefore the earlier value cannot determine the later one.[44] It follows that Bohmian mechanics is incompatible with the belief that all physical observables are discrete. But in our view, there are strong reasons to hold that belief, which include black hole entropy bounds; the existence of a natural minimum length scale (10^{-33} cm); results on area quantization in quantum gravity [18]; the fact that many physical quantities once thought to be continuous have turned out to be discrete; the infinities of quantum field theory; the implausibility of analog "hypercomputers"; and conceptual problems raised by the independence of the continuum hypothesis.

Of course there exist stochastic analogues of Bohmian mechanics, among them Nelsonian mechanics [4] and Bohm and Hiley's "stochastic interpretation" [19]. But it is not obvious why we should prefer these to other stochastic hidden-variable theories. From a quantum-information perspective, it is much more natural to take an abstract approach—one that allows arbitrary finite-dimensional Hilbert spaces, and that does not rule out any transition rule *a priori*.

Stochastic hidden variables have also been considered in the context of modal interpretations; see Dickson [20], Bacciagaluppi and Dickson [21], and Dieks [5] for example. However, the central assumptions in that work are extremely different from ours. In modal interpretations, a pure state evolving unitarily poses no problems at all: one simply rotates the hidden-variable basis along with the state, so that the state always represents a "possessed property" of the system in the current basis. Difficulties arise only for mixed states; and there, the goal is to track a whole set of possessed properties. By contrast, our approach is to fix an orthogonal basis, then track a single hidden variable that is an element of that basis. The issues raised by pure states and mixed states are essentially the same.

Finally we should mention the consistent-histories interpretation of Griffiths [22] and Gell-Mann and Hartle [23]. This interpretation assigns probabilities to various histories through a quantum system, so long as the "interference" between those histories is negligible. Loosely speaking, then, the situations where consistent histories make sense are precisely the ones where the question of transition probabilities can be avoided.

B. Objections

Hidden-variable theories, as we define them, are open to several technical objections. For example, we required transition probabilities for only one orthogonal observable. What about other observables? The problem is that, according to the Kochen-Specker theorem, we cannot assign consistent values to all observables at any *single* time, let alone give transition probabilities for those values. This is an issue in any setting, not just ours. The solution we prefer is to postulate a fixed orthogonal basis of "distinguishable experiences," and to interpret a measurement in any other basis as a unitary followed by a measurement in the fixed basis. As mentioned in Section II A, modal interpretations opt for a different solution, which involves sets of bases that change over time with the state itself.

Another objection is that the probability of transitioning from basis state $|i\rangle$ at time t_1 to basis state $|j\rangle$ at time t_2 might depend on how finely we divide the time interval between t_1 and t_2. In other words, for some state $|\psi\rangle$ and unitaries V, W, we might have

$$S(|\psi\rangle, WV) \neq S(V|\psi\rangle, W)\, S(|\psi\rangle, V)$$

(a similar point was made by Gillespie [24]). Indeed, this is true for any hidden-variable theory other than the product theory $\mathcal{PT}$. To see this, observe that for all unitaries U and states $|\psi\rangle$, there exist unitaries V, W such that $U = WV$ and $V|\psi\rangle = |1\rangle$. Then applying V destroys all information in the hidden variable (that is, decreases its entropy to 0); so if we then apply W, then the variable's final value must be uncorrelated with the initial value. In other words, $S(V|\psi\rangle, W)\, S(|\psi\rangle, V)$ must equal $S_{\mathcal{PT}}(|\psi\rangle, U)$. It follows that to any hidden-variable theory we must associate a time scale, or some other rule for deciding when the transitions take place.

In response, let us point out that exactly the same problem arises in *continuous*-time stochastic hidden-variable theories. For if a state $|\psi\rangle$ is governed by the Schrödinger equation $d|\psi\rangle/dt = iH_t|\psi\rangle$, and a hidden variable's probability distribution $\overrightarrow{p}$ is governed by the stochastic equation $d\overrightarrow{p}/d\tau = A_\tau \overrightarrow{p}$, then there is still an arbitrary parameter $d\tau/dt$ on which the dynamics depend.

Finally, it will be objected that we have ignored special relativity. In Section III we will define a *commutativity axiom*, which informally requires that the stochastic matrix S not depend on the temporal order of spacelike separated events. Unfortunately, we will see that when entangled states are involved, commutativity is irreconcilable with another axiom that seems even more basic. The resulting nonlocality has the same character as the nonlocality of Bohmian mechanics—that is, one cannot use it to send superluminal signals in the usual sense, but it is unsettling nonetheless.

III. AXIOMS FOR HIDDEN-VARIABLE THEORIES

We now state five[45] axioms that we might like hidden-variable theories to satisfy.

Indifference. The indifference axiom says that if U is block-diagonal, then S should also be block-diagonal with the same block structure or some refinement thereof. Formally, let a *block* be a subset $B \subseteq \{1, \ldots, N\}$ such that $(U)_{ij} = 0$ for all $i \in B, j \notin B$ and $i \notin B, j \in B$. Then for all blocks B, we should have $(S)_{ij} = 0$ for all $i \in B, j \notin B$ and $i \notin B, j \in B$. In particular, indifference implies that given any state ρ in a tensor product space $\mathcal{H}_A \otimes \mathcal{H}_B$, and any unitary U that acts only on $\mathcal{H}_A$ (that is, never maps a basis state $|i_A\rangle \otimes |i_B\rangle$ to $|j_A\rangle \otimes |j_B\rangle$ where $i_B \neq j_B$), the stochastic matrix $S(\rho, U)$ acts only on $\mathcal{H}_A$ as well.

Robustness. A theory is robust if it is insensitive to small errors in a state or unitary (which, in particular, implies continuity). Suppose we obtain $\widetilde{\rho}$ and $\widetilde{U}$ by perturbing ρ and U respectively. Then for all polynomials p, there should exist a polynomial q such that for all N,

$$\left\| P\left(\widetilde{\rho}, \widetilde{U}\right) - P(\rho, U) \right\|_\infty \leq \frac{1}{p(N)}$$

where $\|M\|_\infty = \max_{ij} \left|(M)_{ij}\right|$, whenever $\|\widetilde{\rho} - \rho\|_\infty \leq 1/q(N)$ and $\left\|\widetilde{U} - U\right\|_\infty \leq 1/q(N)$. Robustness has an important advantage for quantum computing: if a hidden-variable theory is robust then the set of gates used to define the unitaries $U_1, \ldots, U_T$ is irrelevant, since by the Solovay-Kitaev Theorem (see [25, 26]), any universal quantum gate set can simulate any other to a precision ε with $O(\log^c 1/\varepsilon)$ overhead.

Commutativity. Let ρ_{AB} be a bipartite state, and let U_A and U_B act only on subsystems A and B respectively. Then commutativity means that the order in which U_A and U_B are applied is irrelevant:

$$S\left(U_A \rho_{AB} U_A^{-1}, U_B\right) S(\rho_{AB}, U_A) = S\left(U_B \rho_{AB} U_B^{-1}, U_A\right) S(\rho_{AB}, U_B).$$

Product Commutativity. A theory is product commutative if it satisfies commutativity for all separable pure states $|\psi\rangle = |\psi_A\rangle \otimes |\psi_B\rangle$.

Decomposition Invariance. A theory is decomposition invariant if

$$S(\rho, U) = \sum_{i=1}^{N} p_i S(|\psi_i\rangle\langle\psi_i|, U)$$

for every decomposition

$$\rho = \sum_{i=1}^{N} p_i \left|\psi_i\right\rangle \left\langle\psi_i\right|$$

of ρ into pure states. Theorem 2, part (ii) will show that the analogous axiom for $P(\rho, U)$ is unsatisfiable.

A. Comparing Theories

To fix ideas, let us compare some hidden-variable theories with respect to the above axioms. We have already seen the product theory $\mathcal{PT}$ in Section II. It is easy to show that $\mathcal{PT}$ satisfies robustness, commutativity, and decomposition invariance. However, we consider $\mathcal{PT}$ unsatisfactory because it violates indifference: even if a unitary U acts only on the first of two qubits, $S_{\mathcal{PT}}(\rho, U)$ will readily produce transitions involving the second qubit.

Recognizing this problem, Dieks [5] proposed an alternative theory that in our setting corresponds to the following.[46] First partition the set of basis states into minimal blocks $B_1, \ldots, B_m$ between which U never sends amplitude. Then apply the product theory separately to each block; that is, if i and j belong to the same block B_k then set

$$(S)_{ij} = \frac{\left(U\rho U^{-1}\right)_{jj}}{\sum_{\widehat{j} \in B_k} \left(U\rho U^{-1}\right)_{\widehat{j}\widehat{j}}},$$

and otherwise set $(S)_{ij} = 0$. The resulting *Dieks theory*, $\mathcal{DT}$, satisfies indifference by construction. However, it does not satisfy robustness (or even continuity), since the set of blocks can change if we replace '0' entries in U by arbitrarily small nonzero entries.

In Section V we will introduce two other hidden-variable theories, the flow theory $\mathcal{FT}$ and the Schrödinger theory $\mathcal{ST}$. The following table lists which axioms the four theories satisfy.

	$\mathcal{PT}$ (Product)	$\mathcal{DT}$ (Dieks)	$\mathcal{FT}$ (Flow)	$\mathcal{ST}$ (Schrödinger)
Indifference	No	Yes	Yes	Yes
Robustness	Yes	No	Yes	?
Commutativity	Yes	No	No	No
Product Commutativity	Yes	Yes	No	Yes
Decomposition Invariance	Yes	Yes	No	No

If we could prove that $\mathcal{ST}$ satisfies robustness, then the above table together with the impossibility results of Section IV would completely characterize which of the axioms can be satisfied simultaneously.

IV. IMPOSSIBILITY RESULTS

This section shows that certain sets of axioms cannot be satisfied by any hidden-variable theory. We first show that the failure of $\mathcal{DT}$, $\mathcal{FT}$, and $\mathcal{ST}$ to satisfy commutativity is inherent, and not a fixable technical problem.

Theorem 1 *No hidden-variable theory satisfies both indifference and commutativity.*

Proof. Assume indifference holds, and let our initial state be $|\psi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$. Suppose U_A applies a $\pi/8$ rotation to the first qubit, and U_B applies a $-\pi/8$ rotation to the second qubit. Then

$$U_A|\psi\rangle = U_B|\psi\rangle = \frac{1}{\sqrt{2}}\left(\cos\frac{\pi}{8}|00\rangle - \sin\frac{\pi}{8}|01\rangle + \sin\frac{\pi}{8}|10\rangle + \cos\frac{\pi}{8}|11\rangle\right),$$

$$U_A U_B|\psi\rangle = U_B U_A|\psi\rangle = \frac{1}{2}\left(|00\rangle - |01\rangle + |10\rangle + |11\rangle\right).$$

Let v_t be the value of the hidden variable after t unitaries have been applied. Let E be the event that $v_0 = |00\rangle$ initially, and $v_2 = |10\rangle$ at the end. If U_A is applied before U_B, then the unique 'path' from v_0 to v_2 consistent with indifference sets $v_1 = |10\rangle$. So

$$\Pr[E] \le \Pr[v_1 = |10\rangle] = \frac{1}{2}\sin^2\frac{\pi}{8}.$$

But if U_B is applied before U_A, then the probability that $v_0 = |11\rangle$ and $v_2 = |10\rangle$ is at most $\frac{1}{2}\sin^2\frac{\pi}{8}$, by the same reasoning. Thus, since v_2 must equal $|10\rangle$ with probability 1/4, and since the only possibilities for v_0 are $|00\rangle$ and $|11\rangle$,

$$\Pr[E] \geq \frac{1}{4} - \frac{1}{2}\sin^2\frac{\pi}{8} > \frac{1}{2}\sin^2\frac{\pi}{8}.$$

We conclude that commutativity is violated. ■

Let us remark on the relationship between Theorem 1 and Bell's Theorem. Any hidden-variable theory that is "local" in Bell's sense would immediately satisfy both indifference and commutativity. However, the converse is not obvious, since there might be nonlocal information in the states $U_A|\psi\rangle$ or $U_B|\psi\rangle$, which an indifferent commutative theory could exploit but a local one could not. Theorem 1 rules out this possibility, and in that sense is a strengthening of Bell's Theorem.

The next result places limits on decomposition invariance.

Theorem 2

(i) No theory satisfies indifference, robustness, and decomposition invariance.

(ii) No theory has the property that

$$P(\rho, U) = \sum_{i=1}^{N} p_i P(|\psi_i\rangle\langle\psi_i|, U)$$

for every decomposition $\sum_{i=1}^{N} p_i |\psi_i\rangle\langle\psi_i|$ *of* ρ.

Proof.

(i) Suppose the contrary. Let

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix},$$
$$|\varphi_\theta\rangle = \cos\theta|0\rangle + \sin\theta|1\rangle.$$

Then for every θ not a multiple of $\pi/2$, we must have

$$S(|\varphi_{-\theta}\rangle, R_\theta) = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix},$$
$$S(|\varphi_{\pi/2-\theta}\rangle, R_\theta) = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}.$$

So by decomposition invariance, letting $I = (|0\rangle\langle 0| + |1\rangle\langle 1|)/2$ denote the maximally mixed state,

$$S(I, R_\theta) = S\left(\frac{|\varphi_{-\theta}\rangle\langle\varphi_{-\theta}| + |\varphi_{\pi/2-\theta}\rangle\langle\varphi_{\pi/2-\theta}|}{2}, R_\theta\right) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

and therefore

$$P(I, R_\theta) = \begin{bmatrix} \frac{(\rho)_{00}}{2} & \frac{(\rho)_{11}}{2} \\ \frac{(\rho)_{00}}{2} & \frac{(\rho)_{11}}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix}.$$

By robustness, this holds for $\theta = 0$ as well. But this is a contradiction, since by indifference $P(I, R_0)$ must be half the identity.

(ii) Suppose the contrary; then

$$P(I, R_{\pi/8}) = \frac{P(|0\rangle, R_{\pi/8}) + P(|1\rangle, R_{\pi/8})}{2}.$$

So considering transitions from $|0\rangle$ to $|1\rangle$,

$$\left(P\left(I,R_{\pi/8}\right)\right)_{01}=\frac{\left(P\left(|0\rangle,R_{\pi/8}\right)\right)_{11}+0}{2}=\frac{1}{2}\sin^2\frac{\pi}{8}.$$

But

$$P\left(I,R_{\pi/8}\right)=\frac{P\left(\left|\varphi_{\pi/8}\right\rangle,R_{\pi/8}\right)+P\left(\left|\varphi_{5\pi/8}\right\rangle,R_{\pi/8}\right)}{2}$$

also. Since $R_{\pi/8}\left|\varphi_{\pi/8}\right\rangle=\left|\varphi_{\pi/4}\right\rangle$, we have

$$\begin{aligned}\left(P\left(I,R_{\pi/8}\right)\right)_{01}&\geq\frac{1}{2}\left(P\left(\left|\varphi_{\pi/8}\right\rangle,R_{\pi/8}\right)\right)_{01}\\&\geq\frac{1}{2}\left(\frac{1}{2}-\left(P\left(\left|\varphi_{\pi/8}\right\rangle,R_{\pi/8}\right)\right)_{11}\right)\\&\geq\frac{1}{2}\left(\frac{1}{2}-\sin^2\frac{\pi}{8}\right)\\&>\frac{1}{2}\sin^2\frac{\pi}{8}\end{aligned}$$

which is a contradiction.

■

Notice that all three conditions in Theorem 2, part (i) were essential—for $\mathcal{PT}$ satisfies robustness and decomposition invariance, $\mathcal{DT}$ satisfies indifference and decomposition invariance, and $\mathcal{FT}$ satisfies indifference and robustness.

Our last impossibility result says that no hidden-variable theory satisfies both indifference and "strong continuity," in the sense that for all $\varepsilon>0$ there exists $\delta>0$ such that $\|\widetilde{\rho}-\rho\|\leq\delta$ implies $\|S(\widetilde{\rho},U)-S(\rho,U)\|\leq\varepsilon$. To see this, let

$$\begin{aligned}U&=\begin{bmatrix}1&0&0\\0&\frac{1}{\sqrt{2}}&-\frac{1}{\sqrt{2}}\\0&\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\end{bmatrix},\\ \rho&=\sqrt{1-2\delta^2}\,|0\rangle+\delta\,|1\rangle+\delta\,|2\rangle,\\ \widetilde{\rho}&=\sqrt{1-2\delta^2}\,|0\rangle+\delta\,|1\rangle-\delta\,|2\rangle.\end{aligned}$$

Then by indifference,

$$S(\rho,U)=\begin{bmatrix}1&0&0\\0&0&0\\0&1&1\end{bmatrix},\qquad S(\widetilde{\rho},U)=\begin{bmatrix}1&0&0\\0&1&1\\0&0&0\end{bmatrix}.$$

This is the reason why we defined robustness in terms of the joint probabilities matrix P rather than the stochastic matrix S. On the other hand, note that by giving up indifference, we *can* satisfy strong continuity, as is shown by $\mathcal{PT}$.

V. SPECIFIC THEORIES

This section presents two nontrivial examples of hidden-variable theories: the flow theory in Section V A, and the Schrödinger theory in Section V B.

A. Flow Theory

The idea of the flow theory is to convert a unitary matrix into a weighted directed graph, and then route probability mass through that graph like oil through pipes. Given a unitary U, let

$$\begin{bmatrix}\beta_1\\\vdots\\\beta_N\end{bmatrix}=\begin{bmatrix}(U)_{11}&\cdots&(U)_{N1}\\\vdots&&\vdots\\(U)_{1N}&\cdots&(U)_{NN}\end{bmatrix}\begin{bmatrix}\alpha_1\\\vdots\\\alpha_N\end{bmatrix},$$

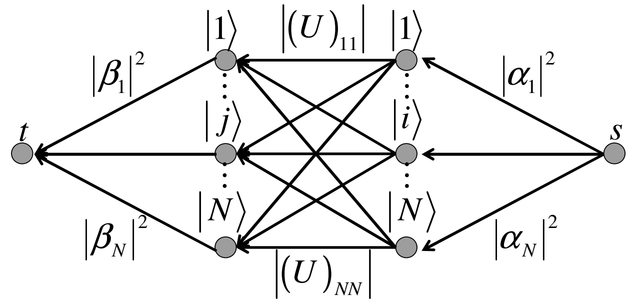

FIG. 1: A network (weighted directed graph with source and sink) corresponding to the unitary U and state $|\psi\rangle$

where for the time being

$$\begin{aligned} |\psi\rangle &= \alpha_1 |1\rangle + \cdots + \alpha_N |N\rangle , \\ U|\psi\rangle &= \beta_1 |1\rangle + \cdots + \beta_N |N\rangle \end{aligned}$$

are pure states. Then consider the network G shown in Figure 1. We have a source vertex s, a sink vertex t, and N input and N output vertices labeled by basis states $|1\rangle, \ldots, |N\rangle$. Each edge of the form $(s, |i\rangle)$ has capacity $|\alpha_i|^2$, each edge $(|i\rangle, |j\rangle)$ has capacity $\left|(U)_{ij}\right|$, and each edge $(|j\rangle, t)$ has capacity $|\beta_j|^2$. A natural question is how much probability mass can flow from s to t without violating the capacity constraints. Rather surprisingly, we show that one unit of mass (that is, all of it) can. Interestingly, this result would be false if edge $(|i\rangle, |j\rangle)$ had capacity $\left|(U)_{ij}\right|^2$ (or even $\left|(U)_{ij}\right|^{1+\varepsilon}$) instead of $\left|(U)_{ij}\right|$. We also show that there exists a mapping from networks to maximal flows in those networks, that is *robust* in the sense that a small change in edge capacities produces only a small change in the amount of flow through any edge.

The proofs of these theorems use classical results from the theory of network flows (see [27] for an introduction). In particular, let a *cut* be a set of edges that separates s from t; the *value* of a cut is the sum of the capacities of its edges. Then a fundamental result called the *Max-Flow-Min-Cut Theorem* [28] says that the maximum possible amount of flow from s to t equals the minimum value of any cut. Using that result we can show the following.

Theorem 3 *One unit of flow can be routed from s to t in G.*

Proof. By the above, it suffices to show that any cut C in G has value at least 1. Let A be the set of $i \in \{1, \ldots, N\}$ such that $(s, |i\rangle) \notin C$, and let B be the set of j such that $(|j\rangle, t) \notin C$. Then C must contain every edge $(|i\rangle, |j\rangle)$ such that $i \in A$ and $j \in B$, and we can assume without loss of generality that C contains no other edges. So the value of C is

$$\sum_{i \notin A} |\alpha_i|^2 + \sum_{j \notin B} |\beta_j|^2 + \sum_{i \in A,\ j \in B} \left|(U)_{ij}\right| .$$

Therefore we need to prove the matrix inequality

$$\left(1 - \sum_{i \in A} |\alpha_i|^2\right) + \left(1 - \sum_{j \in B} |\beta_j|^2\right) + \sum_{i \in A,\ j \in B} \left|(U)_{ij}\right| \geq 1,$$

or

$$1 + \sum_{i \in A,\ j \in B} \left|(U)_{ij}\right| \geq \sum_{i \in A} |\alpha_i|^2 + \sum_{j \in B} |\beta_j|^2 . \tag{1}$$

Let U be fixed, and consider the maximum of the right-hand side of equation (1) over all $|\psi\rangle$. Since

$$\beta_j = \sum_i (U)_{ij}\, \alpha_i,$$

this maximum is equal to the largest eigenvalue λ of the positive semidefinite matrix

$$\sum_{i\in A} |i\rangle\langle i| + \sum_{j\in B} |u_j\rangle\langle u_j|$$

where for each j,

$$|u_j\rangle = (U)_{1j}|1\rangle + \cdots + (U)_{Nj}|N\rangle .$$

Let H_A be the subspace of states spanned by $\{|i\rangle : i \in A\}$, and let H_B be the subspace spanned by $\{|u_j\rangle : j \in B\}$. Also, let $L_A(|\psi\rangle)$ be the length of the projection of $|\psi\rangle$ onto H_A, and let $L_B(|\psi\rangle)$ be the length of the projection of $|\psi\rangle$ onto H_B. Then since the $|i\rangle$'s and $|u_j\rangle$'s form orthogonal bases for H_A and H_B respectively, we have

$$\begin{aligned}\lambda &= \max_{|\psi\rangle}\left(\sum_{i\in A}|\langle i|\psi\rangle|^2 + \sum_{j\in B}|\langle u_j|\psi\rangle|^2\right)\\ &= \max_{|\psi\rangle}\left(L_A(|\psi\rangle)^2 + L_B(|\psi\rangle)^2\right).\end{aligned}$$

So letting θ be the angle between H_A and H_B,

$$\begin{aligned}\lambda &= 2\cos^2\frac{\theta}{2}\\ &= 1+\cos\theta\\ &\le 1+\max_{|a\rangle\in H_A,\ |b\rangle\in H_B}|\langle a|b\rangle|\\ &= 1+\max_{\substack{|\gamma_1|^2+\cdots+|\gamma_N|^2=1\\ |\delta_1|^2+\cdots+|\delta_N|^2=1}}\left|\left(\sum_{i\in A}\gamma_i\langle i|\right)\left(\sum_{j\in B}\delta_j|u_j\rangle\right)\right|\\ &\le 1+\sum_{i\in A,\ j\in B}\left|(U)_{ij}\right|\end{aligned}$$

which completes the theorem. ■

Observe that Theorem 3 still holds if U acts on a mixed state ρ, since we can write ρ as a convex combination of pure states $|\psi\rangle\langle\psi|$, construct a flow for each $|\psi\rangle$ separately, and then take a convex combination of the flows.

Using Theorem 3, we now define the flow theory $\mathcal{FT}$. Let $F(\rho,U)$ be the set of maximal flows for ρ, U—representable by $N\times N$ arrays of real numbers f_{ij} such that $0 \le f_{ij} \le \left|(U)_{ij}\right|$ for all i,j, and also

$$\sum_j f_{ij} = (\rho)_{ii}, \quad \sum_i f_{ij} = \left(U\rho U^{-1}\right)_{jj}.$$

Clearly $F(\rho,U)$ is a convex polytope, which Theorem 3 asserts is nonempty. Form a maximal flow $f^*(\rho,U) \in F(\rho,U)$ as follows: first let f^*_{11} be the maximum of f_{11} over all $f \in F(\rho,U)$. Then let f^*_{12} be the maximum of f_{12} over all $f \in F(\rho,U)$ such that $f_{11} = f^*_{11}$. Continue to loop through all i,j pairs in lexicographic order, setting each f^*_{ij} to its maximum possible value consistent with the $(i-1)N+j-1$ previous values. Finally, let $(P)_{ij} = f^*_{ij}$ for all i,j. As discussed in Section II, given P we can easily obtain the stochastic matrix S by dividing the i^{th} column by $(\rho)_{ii}$, or taking a limit in case $(\rho)_{ii} = 0$.

It is easy to check that $\mathcal{FT}$ so defined satisfies the indifference axiom. Showing that $\mathcal{FT}$ satisfies robustness is harder. Our proof is based on the Ford-Fulkerson algorithm [28], a classic algorithm for computing maximal flows that works by finding a sequence of "augmenting paths," each of which increases the flow from s to t by some positive amount.

Theorem 4 *$\mathcal{FT}$ satisfies robustness.*

Proof. Let G be an arbitrary flow network with source s, sink t, and directed edges $e_1,\ldots,e_m$, where each e_i has capacity c_i and leads from v_i to w_i. It will be convenient to introduce a fictitious edge e_0 from t to s with unlimited capacity; then maximizing the flow through G is equivalent to maximizing the flow through e_0. Suppose we produce a new network $\widetilde{G}$ by increasing a single capacity c_{i^*} by some $\varepsilon > 0$. Let f^* be the optimal flow for G, obtained by first maximizing the flow f_0 through e_0, then maximizing the flow f_1 through e_1 holding f_0 fixed, and so on up to f_m. Let $\widetilde{f}^*$ be the maximal flow for $\widetilde{G}$ produced in the same way. We claim that for all $i \in \{0, \ldots, m\}$,

$$\left| \widetilde{f}_i^* - f_i^* \right| \leq \varepsilon.$$

To see that the theorem follows from this claim: first, if f^* is robust under adding ε to c_{i^*}, then it must also be robust under subtracting ε from c_{i^*}. Second, if we change ρ, U to $\widetilde{\rho}, \widetilde{U}$ such that $\|\widetilde{\rho} - \rho\|_\infty \leq 1/q(N)$ and $\left\|\widetilde{U} - U\right\|_\infty \leq 1/q(N)$, then we can imagine the $N^2 + 2N$ edge capacities are changed one by one, so that

$$\begin{aligned} \left\| f^*\left(\widetilde{\rho}, \widetilde{U}\right) - f^*(\rho, U) \right\|_\infty &\leq \sum_{ij} \left| \left| \left(\widetilde{U}\right)_{ij} \right| - \left| (U)_{ij} \right| \right| + \sum_i \left| (\widetilde{\rho})_{ii} - (\rho)_{ii} \right| \\ &\quad + \sum_j \left| \left(\widetilde{U}\widetilde{\rho}\widetilde{U}^{-1}\right)_{jj} - \left(U\rho U^{-1}\right)_{jj} \right| \\ &\leq \frac{4N^2}{q(N)}. \end{aligned}$$

(Here we have made no attempt to optimize the bound.)

We now prove the claim. To do so we describe an iterative algorithm for computing f^*. First maximize the flow f_0 through e_0, by using the Ford-Fulkerson algorithm to find a maximal flow from s to t. Let $f^{(0)}$ be the resulting flow, and let $G^{(1)}$ be the residual network that corresponds to $f^{(0)}$. For each i, that is, $G^{(1)}$ has an edge $e_i = (v_i, w_i)$ of capacity $c_i^{(1)} = c_i - f_i^{(0)}$, and an edge $\overline{e}_i = (w_i, v_i)$ of capacity $\overline{c}_i^{(1)} = f_i^{(0)}$. Next maximize f_1 subject to f_0 by using the Ford-Fulkerson algorithm to find "augmenting cycles" from w_1 to v_1 and back to w_1 in $G^{(1)} \setminus \{e_0, \overline{e}_0\}$. Continue in this manner until each of $f_1, \ldots, f_m$ has been maximized subject to the previous f_i's. Finally set $f^* = f^{(m)}$.

Now, one way to compute $\widetilde{f}^*$ is to start with f^*, then repeatedly "correct" it by applying the same iterative algorithm to maximize $\widetilde{f}_0$, then $\widetilde{f}_1$, and so on. Let $\varepsilon_i = \left| \widetilde{f}_i^* - f_i^* \right|$; then we need to show that $\varepsilon_i \leq \varepsilon$ for all $i \in \{0, \ldots, m\}$. The proof is by induction on i. Clearly $\varepsilon_0 \leq \varepsilon$, since increasing c_{i^*} by ε can increase the value of the minimum cut from s to t by at most ε. Likewise, after we maximize $\widetilde{f}_0$, the value of the minimum cut from w_1 to v_1 can increase by at most $\varepsilon - \varepsilon_0 + \varepsilon_0 = \varepsilon$. For of the at most ε new units of flow from w_1 to v_1 that increasing c_{i^*} made available, ε_0 of them were "taken up" in maximizing $\widetilde{f}_0$, but the process of maximizing $\widetilde{f}_0$ could have again increased the minimum cut from w_1 to v_1 by up to ε_0. Continuing in this way,

$$\varepsilon_2 \leq \varepsilon - \varepsilon_0 + \varepsilon_0 - \varepsilon_1 + \varepsilon_1 = \varepsilon,$$

and so on up to ε_m. This completes the proof. ■

That $\mathcal{FT}$ violates decomposition invariance now follows from Theorem 2, part (i). One can also show that $\mathcal{FT}$ violates product commutativity, by considering the following example: let $|\psi\rangle = \left|\varphi_{\pi/4}\right\rangle \otimes \left|\varphi_{-\pi/8}\right\rangle$ be a 2-qubit initial state, and let $R_{\pi/4}^A$ and $R_{\pi/4}^B$ be $\pi/4$ rotations applied to the first and second qubits respectively. Then

$$S\left(R_{\pi/4}^A |\psi\rangle, R_{\pi/4}^B\right) S\left(|\psi\rangle, R_{\pi/4}^A\right) \neq S\left(R_{\pi/4}^B |\psi\rangle, R_{\pi/4}^A\right) S\left(|\psi\rangle, R_{\pi/4}^B\right).$$

We omit a proof for brevity.

B. Schrödinger Theory

Our final hidden-variable theory, which we call the *Schrödinger theory* or $\mathcal{ST}$, is the most interesting one mathematically. The idea—to make a matrix into a stochastic matrix via row and column rescaling—is natural enough that we came upon it independently, only later learning that it originated in a 1931 paper of Schrödinger [1]. The idea was subsequently developed by Fortet [29], Beurling [30], Nagasawa [15], and others. Our goal is to give what (to our knowledge) is the first self-contained, reasonably accessible presentation of the main result in this area; and to interpret that result in what we think is the correct way: as providing one example of a hidden-variable theory, whose strengths and weaknesses should be directly compared to those of other theories.

Most of the technical difficulties in [1, 15, 29, 30] arise because the stochastic process being constructed involves continuous time and particle positions. Here we eliminate those difficulties by restricting attention to discrete time and finite-dimensional Hilbert spaces. We thereby obtain a generalized version[47] of a problem that computer scientists know as (r,c)-*scaling of matrices* [16, 31, 32].

As in the case of the flow theory, given a unitary U acting on a state ρ, the first step is to replace each entry of U by its absolute value, obtaining a nonnegative matrix $U^{(0)}$ defined by $\left(U^{(0)}\right)_{ij} := \left|(U)_{ij}\right|$. We then wish to find nonnegative column multipliers $\alpha_1,\ldots,\alpha_N$ and row multipliers $\beta_1,\ldots,\beta_N$ such that for all i,j,

$$\alpha_i\beta_1\left(U^{(0)}\right)_{i1}+\cdots+\alpha_i\beta_N\left(U^{(0)}\right)_{iN}=(\rho)_{ii}\,, \tag{2}$$

$$\alpha_1\beta_j\left(U^{(0)}\right)_{1j}+\cdots+\alpha_N\beta_j\left(U^{(0)}\right)_{Nj}=\left(U\rho U^{-1}\right)_{jj}\,. \tag{3}$$

If we like, we can interpret the α_i's and β_j's as dynamical variables that reach equilibrium precisely when equations (2) and (3) are satisfied. Admittedly, it might be thought physically implausible that such a complicated dynamical process should take place at every instant of time. On the other hand, it is hard to imagine a more "benign" way to convert $U^{(0)}$ into a joint probabilities matrix, than by simply rescaling its rows and columns.

We will show that multipliers satisfying (2) and (3) always exist. The intuition of a dynamical process reaching equilibrium turns out to be key to the proof. For all $t\geq 0$, let

$$\left(U^{(2t+1)}\right)_{ij}=\frac{(\rho)_{ii}}{\sum_k\left(U^{(2t)}\right)_{ik}}\left(U^{(2t)}\right)_{ij},$$

$$\left(U^{(2t+2)}\right)_{ij}=\frac{\left(U\rho U^{-1}\right)_{jj}}{\sum_k\left(U^{(2t+1)}\right)_{kj}}\left(U^{(2t+1)}\right)_{ij}.$$

In words, we obtain $U^{(2t+1)}$ by normalizing each column i of $U^{(2t)}$ to sum to $(\rho)_{ii}$; likewise we obtain $U^{(2t+2)}$ by normalizing each row j of $U^{(2t+1)}$ to sum to $\left(U\rho U^{-1}\right)_{jj}$. The crucial fact is that the above process always converges to some $P(\rho,U)=\lim_{t\rightarrow\infty}U^{(t)}$. We can therefore take

$$\alpha_i=\prod_{t=0}^{\infty}\frac{(\rho)_{ii}}{\sum_k\left(U^{(2t)}\right)_{ik}},$$

$$\beta_j=\prod_{t=0}^{\infty}\frac{\left(U\rho U^{-1}\right)_{jj}}{\sum_k\left(U^{(2t+1)}\right)_{kj}}$$

for all i,j. Although we will not prove it here, it turns out that this yields the *unique* solution to equations (2) and (3), up to a global rescaling of the form $\alpha_i\rightarrow\alpha_i c$ for all i and $\beta_j\rightarrow\beta_j/c$ for all j [15].

Our convergence proof will reuse a result about network flows from Section V A, in order to define a nondecreasing "progress measure" based on Kullback-Leibler distance.

Theorem 5 *The limit $P(\rho,U)=\lim_{t\rightarrow\infty}U^{(t)}$ exists.*

Proof. A consequence of Theorem 3 is that for every ρ,U, there exists an $N\times N$ array of nonnegative real numbers f_{ij} such that

(1) $f_{ij}=0$ whenever $\left|(U)_{ij}\right|=0$,

(2) $f_{i1}+\cdots+f_{iN}=(\rho)_{ii}$ for all i, and

(3) $f_{1j}+\cdots+f_{Nj}=\left(U\rho U^{-1}\right)_{jj}$ for all j.

Given any such array, define a progress measure

$$Z^{(t)}=\prod_{ij}\left(U^{(t)}\right)_{ij}^{f_{ij}},$$

where we adopt the convention $0^0=1$. We claim that $Z^{(t+1)}\geq Z^{(t)}$ for all $t\geq 1$. To see this, assume without loss of generality that we are on an odd step $2t+1$, and let $C_i^{(2t)}=\sum_j\left(U^{(2t)}\right)_{ij}$ be the i^{th} column sum before we normalize it. Then

$$\begin{aligned}
Z^{(2t+1)} &= \prod_{ij} \left(U^{(2t+1)}\right)_{ij}^{f_{ij}} \\
&= \prod_{ij} \left(\frac{(\rho)_{ii}}{C_i^{(2t)}} \left(U^{(2t)}\right)_{ij} \right)^{f_{ij}} \\
&= \left(\prod_{ij} \left(U^{(2t)}\right)_{ij}^{f_{ij}} \right) \left(\prod_i \left(\frac{(\rho)_{ii}}{C_i^{(2t)}} \right)^{f_{i1}+\cdots+f_{iN}} \right) \\
&= Z^{(2t)} \cdot \prod_i \left(\frac{(\rho)_{ii}}{C_i^{(2t)}} \right)^{(\rho)_{ii}}.
\end{aligned}$$

As a result of the $2t^{th}$ normalization step, we had $\sum_i C_i^{(2t)} = 1$. Subject to that constraint, the maximum of

$$\prod_i \left(C_i^{(2t)}\right)^{(\rho)_{ii}}$$

over the $C_i^{(2t)}$'s occurs when $C_i^{(2t)} = (\rho)_{ii}$ for all i—a simple calculus fact that follows from the nonnegativity of Kullback-Leibler distance. This implies that $Z^{(2t+1)} \geq Z^{(2t)}$. Similarly, normalizing rows leads to $Z^{(2t+2)} \geq Z^{(2t+1)}$.

It follows that the limit $P(\rho, U) = \lim_{t\to\infty} U^{(t)}$ exists. For suppose not; then some $C_i^{(t)}$ is bounded away from $(\rho)_{ii}$, so there exists an $\varepsilon > 0$ such that $Z^{(t+1)} \geq (1+\varepsilon) Z^{(t)}$ for all even t. But this is a contradiction, since $Z^{(0)} > 0$ and $Z^{(t)} \leq 1$ for all t. ■

Besides showing that $P(\rho, U)$ is well-defined, Theorem 5 also yields a procedure to *calculate* $P(\rho, U)$ (as well as the α_i's and β_j's). It can be shown that this procedure converges to within entrywise error ε after a number steps polynomial in N and $1/\varepsilon$. Also, once we have $P(\rho, U)$, the stochastic matrix $S(\rho, U)$ is readily obtained by normalizing each column of $P(\rho, U)$ to sum to 1. This completes the definition of the Schrödinger theory $\mathcal{ST}$.

It is immediate that $\mathcal{ST}$ satisfies indifference. Let us show that it satisfies product commutativity as well.

Proposition 6 $\mathcal{ST}$ *satisfies product commutativity.*

Proof. Given a state $|\psi\rangle = |\psi_A\rangle \otimes |\psi_B\rangle$, let $U_A \otimes I$ act only on $|\psi_A\rangle$ and let $I \otimes U_B$ act only on $|\psi_B\rangle$. Then we claim that

$$S(|\psi\rangle, U_A \otimes I) = S(|\psi_A\rangle, U_A) \otimes I.$$

The reason is simply that multiplying all amplitudes in $|\psi_A\rangle$ and $U_A|\psi_A\rangle$ by a constant factor α_x, as we do for each basis state $|x\rangle$ of $|\psi_B\rangle$, has no effect on the scaling procedure that produces $S(|\psi_A\rangle, U_A)$. Similarly

$$S(|\psi\rangle, I \otimes U_B) = I \otimes S(|\psi_B\rangle, U_B).$$

It follows that

$$\begin{aligned}
S(|\psi_A\rangle, U_A) \otimes S(|\psi_B\rangle, U_B) &= S(U_A|\psi_A\rangle \otimes |\psi_B\rangle, I \otimes U_B)\, S(|\psi\rangle, U_A \otimes I) \\
&= S(|\psi_A\rangle \otimes U_B|\psi_B\rangle, U_A \otimes I)\, S(|\psi\rangle, I \otimes U_B).
\end{aligned}$$

■

On the other hand, numerical simulations readily show that $\mathcal{ST}$ violates decomposition invariance, even when $N = 2$ (we omit a concrete example for brevity).

VI. THE COMPUTATIONAL MODEL

We now explain our model of computation, building our way up to the complexity class DQP. From now on, the states ρ that we consider will always be pure states of $\ell = \log_2 N$ qubits. That is, $\rho = |\psi\rangle\langle\psi|$ where

$$|\psi\rangle = \sum_{x \in \{0,1\}^\ell} \alpha_x |x\rangle.$$

Our algorithms will work under *any* hidden-variable theory that satisfies the indifference axiom. On the other hand, if we take into account that even in theory (let alone in practice), a generic unitary cannot be represented exactly with a finite universal gate set, only approximated arbitrarily well, then we also need the robustness axiom. Thus, it is reassuring that there *exists* a hidden-variable theory (namely $\mathcal{FT}$) that satisfies both indifference and robustness.

Let a quantum computer have the initial state $|0\rangle^{\otimes \ell}$, and suppose we apply a sequence $\mathcal{U} = (U_1, \ldots, U_T)$ of unitary operations, each of which is implemented by a polynomial-size quantum circuit. Then a *history* of a hidden variable through the computation is a sequence $H = (v_0, \ldots, v_T)$ of basis states, where v_t is the variable's value immediately after U_t is applied (thus $v_0 = |0\rangle^{\otimes \ell}$). Given any hidden-variable theory $\mathcal{T}$, we can obtain a probability distribution $\Omega(\mathcal{U}, \mathcal{T})$ over histories by just applying $\mathcal{T}$ repeatedly, once for each U_t, to obtain the stochastic matrices

$$S\left(|0\rangle^{\otimes \ell}, U_1\right), \quad S\left(U_1 |0\rangle^{\otimes \ell}, U_2\right), \quad \ldots \quad S\left(U_{T-1} \cdots U_1 |0\rangle^{\otimes \ell}, U_T\right).$$

Note that $\Omega(\mathcal{U}, \mathcal{T})$ is a Markov distribution; that is, each v_t is independent of the other v_i's conditioned on v_{t-1} and v_{t+1}. Admittedly, $\Omega(\mathcal{U}, \mathcal{T})$ could depend on the precise way in which the combined circuit $U_T \cdots U_1$ is "sliced" into component circuits $U_1, \ldots, U_T$. But as we showed in Section II B, such dependence on the granularity of unitaries is unavoidable in any hidden-variable theory other than $\mathcal{PT}$.

Given a hidden-variable theory $\mathcal{T}$, let $\mathcal{O}(\mathcal{T})$ be an oracle that takes as input a positive integer ℓ, and a sequence of quantum circuits $\mathcal{U} = (U_1, \ldots, U_T)$ that act on ℓ qubits. Here each U_t is specified by a sequence $\left(g_{t,1}, \ldots, g_{t,m(t)}\right)$ of gates chosen from some finite universal gate set $\mathcal{G}$. The oracle $\mathcal{O}(\mathcal{T})$ returns as output a sample $(v_0, \ldots, v_T)$ from the history distribution $\Omega(\mathcal{U}, \mathcal{T})$ defined previously. Now let A be a deterministic classical Turing machine that is given oracle access to $\mathcal{O}(\mathcal{T})$. The machine A receives an input x, makes a single oracle query to $\mathcal{O}(\mathcal{T})$, then produces an output based on the response. We say a set of strings L is in DQP if there exists an A such that for all sufficiently large n and inputs $x \in \{0,1\}^n$, and all theories $\mathcal{T}$ satisfying the indifference and robustness axioms, A correctly decides whether $x \in L$ with probability at least $2/3$, in time polynomial in n.

Let us make some remarks about the above definition. There is no real significance in our requirement that A be deterministic and classical, and that it be allowed only one query to $\mathcal{O}(\mathcal{T})$. We made this choice only because it suffices for our upper bounds; it might be interesting to consider the effects of other choices. However, other aspects of the definition are not arbitrary. The order of quantifiers matters; we want a single A that works for *any* hidden-variable theory satisfying indifference and robustness. Also, we require A to succeed only for sufficiently large n since by choosing a large enough polynomial $q(N)$ in the statement of the robustness axiom, an adversary might easily make A incorrect on a finite number of instances.

A. Basic Results

Having defined the complexity class DQP, let us establish its most basic properties. First of all, it is immediate that $\mathsf{BQP} \subseteq \mathsf{DQP}$; that is, sampling histories is at least as powerful as standard quantum computation. For v_1, the first hidden-variable value returned by $\mathcal{O}(\mathcal{T})$, can be seen as simply the result of applying a polynomial-size quantum circuit U_1 to the initial state $|0\rangle^{\otimes \ell}$ and then measuring in the standard basis. A key further observation is the following.

Theorem 7 *Any universal gate set yields the same complexity class* DQP*. By universal, we mean that any unitary matrix (real or complex) can be approximated, without the need for ancilla qubits.*

Proof. Let $\mathcal{G}$ and $\mathcal{G}'$ be universal gate sets. Also, let $\mathcal{U} = (U_1, \ldots, U_T)$ be a sequence of ℓ-qubit unitaries, each specified by a polynomial-size quantum circuit over $\mathcal{G}$. We have $T, \ell = O(\mathrm{poly}(n))$ where n is the input length. We can also assume without loss of generality that $\ell \geq n$, since otherwise we simply insert $n - \ell$ dummy qubits that are never acted on (by the indifference axiom, this will not affect the results). We want to approximate $\mathcal{U}$ by another sequence of ℓ-qubit unitaries, $\mathcal{U}' = (U_1', \ldots, U_T')$, where each U_t' is specified by a quantum circuit over $\mathcal{G}'$. In particular, for all t we want $\|U_t' - U_t\|_\infty \leq 2^{-\ell^2 T}$. By the Solovay-Kitaev Theorem [25, 26], we can achieve this using $\mathrm{poly}\left(n, \ell^2 T\right) = \mathrm{poly}(n)$ gates from $\mathcal{G}'$; moreover, the circuit for U_t' can be constructed in polynomial time given the circuit for U_t.

Let $|\psi_t\rangle = U_t \cdots U_1 |0\rangle^{\otimes \ell}$ and $|\psi_t'\rangle = U_t' \cdots U_1' |0\rangle^{\otimes \ell}$. Notice that for all $t \in \{1, \ldots, T\}$,

$$\begin{aligned} \left\| |\psi_t'\rangle - |\psi_t\rangle \right\|_\infty &\leq 2^\ell \left(\left\| |\psi_{t-1}'\rangle - |\psi_{t-1}\rangle \right\|_\infty + 2^{-\ell^2 T} \right) \\ &\leq T 2^{\ell T} \left(2^{-\ell^2 T} \right) = T 2^{-\ell(\ell-1)T}, \end{aligned}$$

since $\| |\psi_0'\rangle - |\psi_0\rangle \|_\infty = 0$. Here $\| \ \|_\infty$ denotes the maximum entrywise difference between two vectors in $\mathbb{C}^{2^\ell}$. Also, given a theory $\mathcal{T}$, let P_t and P_t' be the joint probabilities matrices corresponding to U_t and U_t' respectively. Then by the robustness axiom, there exists a polynomial q such that if $\|U_t' - U_t\|_\infty \leq 1/q\left(2^\ell\right)$ and $\left\| |\psi_{t-1}'\rangle - |\psi_{t-1}\rangle \right\|_\infty \leq 1/q\left(2^\ell\right)$, then $\|P_t - P_t'\|_\infty \leq 2^{-3\ell}$. For all such polynomials q, we have $2^{-\ell^2 T} \leq 1/q\left(2^\ell\right)$ and $T 2^{-\ell(\ell-1)T} \leq 1/q\left(2^\ell\right)$ for sufficiently large $n \leq \ell$. Therefore $\|P_t - P_t'\|_\infty \leq 2^{-3\ell}$ for all t and sufficiently large n.

Now assume n is sufficiently large, and consider the distributions $\Omega\left(\mathcal{U},\mathcal{T}\right)$ and $\Omega\left(\mathcal{U}',\mathcal{T}\right)$ over classical histories $H = (v_0,\ldots,v_T)$. For all $t \in \{1,\ldots,T\}$ and $x \in \{0,1\}^\ell$, we have

$$\left| \Pr_{\Omega(\mathcal{U},\mathcal{T})}\left[v_t = |x\rangle\right] - \Pr_{\Omega(\mathcal{U}',\mathcal{T})}\left[v_t = |x\rangle\right] \right| \leq 2^\ell \left(2^{-3\ell}\right) = 2^{-2\ell}.$$

It follows by the union bound that the variation distance $\|\Omega\left(\mathcal{U}',\mathcal{T}\right) - \Omega\left(\mathcal{U},\mathcal{T}\right)\|$ is at most

$$T 2^\ell \left(2^{-2\ell}\right) = \frac{T}{2^\ell} \leq \frac{T}{2^n}.$$

In other words, $\Omega\left(\mathcal{U}',\mathcal{T}\right)$ can be distinguished from $\Omega\left(\mathcal{U},\mathcal{T}\right)$ with bias at most $T/2^n$, which is exponentially small. So any classical postprocessing algorithm that succeeds with high probability given $H \in \Omega\left(\mathcal{U},\mathcal{T}\right)$, also succeeds with high probability given $H \in \Omega\left(\mathcal{U}',\mathcal{T}\right)$. This completes the theorem. ■

Unfortunately, the best upper bound on DQP we have been able to show is DQP $\subseteq$ EXP; that is, any problem in DQP is solvable in deterministic exponential time. The proof is trivial: let $\mathcal{T}$ be the flow theory $\mathcal{FT}$, with the slight modification that we omit the step from Section V A of symmetrizing over all permutations of basis states. Then by using the Ford-Fulkerson algorithm, we can clearly construct the requisite maximum flows in time polynomial in 2^ℓ (hence exponential in n), and thereby calculate the probability of each possible history $(v_1,\ldots,v_T)$ to suitable precision.

VII. THE JUGGLE SUBROUTINE

This section presents a crucial subroutine that will be used in both algorithms of this paper: the algorithm for simulating statistical zero knowledge in Section VIII, and the algorithm for search in $N^{1/3}$ queries in Section IX. Given an ℓ-qubit state $\left(|a\rangle + |b\rangle\right)/\sqrt{2}$, where $|a\rangle$ and $|b\rangle$ are unknown basis states, the goal of the juggle subroutine is to learn both a and b. The name arises because our strategy will be to "juggle" a hidden variable, so that if it starts out at $|a\rangle$ then with non-negligible probability it transitions to $|b\rangle$, and vice versa. Inspecting the entire history of the hidden variable will then reveal both a and b, as desired.

To produce this behavior, we will exploit a basic feature of quantum mechanics: that observable information in one basis can become unobservable phase information in a different basis. We will apply a sequence of unitaries that hide all information about a and b in phases, thereby forcing the hidden variable to "forget" whether it started at $|a\rangle$ or $|b\rangle$. We will then invert those unitaries to return the state to $\left(|a\rangle + |b\rangle\right)/\sqrt{2}$, at which point the hidden variable, having "forgotten" its initial value, must be unequal to that value with probability $1/2$.

We now give the subroutine. Let $|\psi\rangle = \left(|a\rangle + |b\rangle\right)/\sqrt{2}$ be the initial state. The first unitary, U_1, consists of Hadamard gates on $\ell - 1$ qubits chosen uniformly at random, and the identity operation on the remaining qubit, i. Next U_2 consists of a Hadamard gate on qubit i. Finally U_3 consists of Hadamard gates on all ℓ qubits. Let $a = a_1 \ldots a_\ell$ and $b = b_1 \ldots b_\ell$. Then since $a \neq b$, we have $a_i \neq b_i$ with probability at least $1/\ell$. Assuming that occurs, the state

$$U_1 |\psi\rangle = \frac{1}{2^{\ell/2}} \left(\sum_{z \in \{0,1\}^\ell \,:\, z_i = a_i} (-1)^{a \cdot z - a_i z_i} |z\rangle + \sum_{z \in \{0,1\}^\ell \,:\, z_i = b_i} (-1)^{b \cdot z - b_i z_i} |z\rangle \right)$$

assigns nonzero amplitude to all 2^ℓ basis states. Then $U_2 U_1 |\psi\rangle$ assigns nonzero amplitude to $2^{\ell-1}$ basis states $|z\rangle$, namely those for which $a \cdot z \equiv b \cdot z \pmod 2$. Finally $U_3 U_2 U_1 |\psi\rangle = |\psi\rangle$.

Let v_t be the value of the hidden variable after U_t is applied. Then assuming $a_i \neq b_i$, we claim that v_3 is independent of v_0. So in particular, if $v_0 = |a\rangle$ then $v_3 = |b\rangle$ with $1/2$ probability, and if $v_0 = |b\rangle$ then $v_3 = |a\rangle$ with $1/2$ probability. To see this, observe that when U_1 is applied, there is no interference between basis states $|z\rangle$ such that $z_i = a_i$, and those such that $z_i = b_i$. So by the indifference axiom, the probability mass at $|a\rangle$ must spread out evenly among all $2^{\ell-1}$ basis states that agree with a on the i^{th} bit, and similarly for the probability mass at $|b\rangle$. Then after U_2 is applied, v_2 can differ from v_1 only on the i^{th} bit, again by the indifference axiom. So each basis state of $U_2 U_1 |\psi\rangle$ must receive an equal contribution from probability mass originating at $|a\rangle$, and probability mass originating at $|b\rangle$. Therefore v_2 is independent of v_0, from which it follows that v_3 is independent of v_0 as well.

Unfortunately, the juggle subroutine only works with probability $1/(2\ell)$—for it requires that $a_i \neq b_i$, and even then, inspecting the history $(v_0, v_1, \ldots)$ only reveals both $|a\rangle$ and $|b\rangle$ with probability $1/2$. Furthermore, the definition of DQP does not allow more than one call to the history oracle. However, all we need to do is pack multiple subroutine calls into a single oracle call. That is, choose U_4 similarly to U_1 (except with a different value of i), and set $U_5 = U_2$ and $U_6 = U_3$. Do the same with U_7, U_8, and U_9, and so on. Since $U_3, U_6, U_9, \ldots$ all return the quantum state to $|\psi\rangle$, the effect is that of multiple independent juggle attempts. With $2\ell^2$ attempts, we can make the failure probability at most $(1 - 1/(2\ell))^{2\ell^2} < e^{-\ell}$.

As a final remark, it is easy to see that the juggle subroutine works equally well with states of the form $|\psi\rangle = (|a\rangle - |b\rangle)/\sqrt{2}$. This will prove useful in Section IX.

VIII. SIMULATING SZK

Our goal is to show that $\mathsf{SZK} \subseteq \mathsf{DQP}$. Here SZK, or Statistical Zero Knowledge, was originally defined as the class of all problems that possess a certain kind of "zero-knowledge proof protocol"—that is, a protocol between an omniscient prover and a verifier, by which the verifier becomes convinced of the answer to a problem, yet without learning anything else about the problem. However, for our purposes this cryptographic definition of SZK is irrelevant. For Sahai and Vadhan [33] have given an alternate and much simpler characterization: a problem is in SZK if and only if it can be reduced to a problem called Statistical Difference, which involves deciding whether two probability distributions are close or far.

More formally, let P_0 and P_1 be functions that map n-bit strings to $q(n)$-bit strings for some polynomial q, and that are specified by classical polynomial-time algorithms. Let Λ_0 and Λ_1 be the probability distributions over $P_0(x)$ and $P_1(x)$ respectively, if $x \in \{0,1\}^n$ is chosen uniformly at random. Then the problem is to decide whether $\|\Lambda_0 - \Lambda_1\|$ is less than $1/3$ or greater than $2/3$, given that one of these is the case. Here

$$\|\Lambda_0 - \Lambda_1\| = \frac{1}{2} \sum_{y \in \{0,1\}^{q(n)}} \left| \Pr_{x \in \{0,1\}^n} [P_0(x) = y] - \Pr_{x \in \{0,1\}^n} [P_1(x) = y] \right|$$

is the variation distance between Λ_0 and Λ_1.

To illustrate, let us show that Graph Isomorphism is in SZK. Given two graphs G_0 and G_1, take Λ_0 to be the uniform distribution over all permutations of G_0, and Λ_1 to be uniform over all permutations of G_1. This way, if G_0 and G_1 are isomorphic, then Λ_0 and Λ_1 will be identical, so $\|\Lambda_0 - \Lambda_1\| = 0$. On the other hand, if G_0 and G_1 are non-isomorphic, then Λ_0 and Λ_1 will be perfectly distinguishable, so $\|\Lambda_0 - \Lambda_1\| = 1$. Since Λ_0 and Λ_1 are clearly samplable by polynomial-time algorithms, it follows that any instance of Graph Isomorphism can be expressed as an instance of Statistical Difference. For a proof that Approximate Shortest Vector is in SZK, we refer the reader to Goldreich and Goldwasser [34] (see also Aharonov and Ta-Shma [35]).

Our proof will use the following "amplification lemma" from [33]:[48]

Lemma 8 (Sahai and Vadhan) *Given efficiently-samplable distributions Λ_0 and Λ_1, we can construct new efficiently-samplable distributions Λ_0' and Λ_1', such that if $\|\Lambda_0 - \Lambda_1\| \leq 1/3$ then $\|\Lambda_0' - \Lambda_1'\| \leq 2^{-n}$, while if $\|\Lambda_0 - \Lambda_1\| \geq 2/3$ then $\|\Lambda_0' - \Lambda_1'\| \geq 1 - 2^{-n}$.*

In particular, Lemma 8 means we can assume without loss of generality that either $\|\Lambda_0 - \Lambda_1\| \leq 2^{-n^c}$ or $\|\Lambda_0 - \Lambda_1\| \geq 1 - 2^{-n^c}$ for some constant $c > 0$.

Having covered the necessary facts about SZK, we can now proceed to the main result.

Theorem 9 *$\mathsf{SZK} \subseteq \mathsf{DQP}$.*

Proof. We show how to solve Statistical Difference by using a history oracle. For simplicity, we start with the special case where P_0 and P_1 are both one-to-one functions. In this case, the circuit sequence $\mathcal{U}$ given to the history oracle does the following: it first prepares the state

$$\frac{1}{2^{(n+1)/2}} \sum_{b \in \{0,1\}, x \in \{0,1\}^n} |b\rangle |x\rangle |P_b(x)\rangle .$$

It then applies the juggle subroutine to the joint state of the $|b\rangle$ and $|x\rangle$ registers, taking $\ell = n+1$. Notice that by the indifference axiom, the hidden variable will never transition from one value of $P_b(x)$ to another—exactly as if we had *measured* the third register in the standard basis. All that matters is the reduced state $|\psi\rangle$ of the first two registers, which has the form $(|0\rangle|x_0\rangle + |1\rangle|x_1\rangle)/\sqrt{2}$ for some x_0, x_1 if $\|\Lambda_0 - \Lambda_1\| = 0$, and $|b\rangle|x\rangle$ for some b, x if $\|\Lambda_0 - \Lambda_1\| = 1$. We have already seen that the juggle subroutine can distinguish these two cases: when the hidden-variable history is inspected, it will contain two values of the $|b\rangle$ register in the former case, and only one value in the latter case. Also, clearly the case $\|\Lambda_0 - \Lambda_1\| \leq 2^{-n^c}$ is statistically indistinguishable from $\|\Lambda_0 - \Lambda_1\| = 0$ with respect to the subroutine, and likewise $\|\Lambda_0 - \Lambda_1\| \geq 1 - 2^{-n^c}$ is indistinguishable from $\|\Lambda_0 - \Lambda_1\| = 1$.

We now consider the general case, where P_0 and P_1 need not be one-to-one. Our strategy is to reduce to the one-to-one case, by using a well-known hashing technique of Valiant and Vazirani [17]. Let $\mathcal{D}_{n,k}$ be the uniform distribution over all affine functions mapping $\{0,1\}^n$ to $\{0,1\}^k$, where we identify those sets with the finite fields $\mathbb{F}_2^n$ and $\mathbb{F}_2^k$ respectively. What Valiant and Vazirani showed is that, for all subsets $A \subseteq \{0,1\}^n$ such that $2^{k-2} \leq |A| \leq 2^{k-1}$, and all $s \in \{0,1\}^k$,

$$\Pr_{h \in \mathcal{D}_{n,k}} \left[\left| A \cap h^{-1}(s) \right| = 1 \right] \geq \frac{1}{8}.$$

As a corollary, the expectation over $h \in \mathcal{D}_{n,k}$ of

$$\left| \left\{ s \in \{0,1\}^k : \left| A \cap h^{-1}(s) \right| = 1 \right\} \right|$$

is at least $2^k/8$. It follows that, if x is drawn uniformly at random from A, then

$$\Pr_{h,x} \left[\left| A \cap h^{-1}(h(x)) \right| = 1 \right] \geq \frac{2^k/8}{|A|} \geq \frac{1}{4}.$$

This immediately suggests the following algorithm for the many-to-one case. Draw k uniformly at random from $\{2, \ldots, n+1\}$; then draw $h_0, h_1 \in \mathcal{D}_{n,k}$. Have $\mathcal{U}$ prepare the state

$$\frac{1}{2^{(n+1)/2}} \sum_{b \in \{0,1\}, x \in \{0,1\}^n} |b\rangle |x\rangle |P_b(x)\rangle |h_b(x)\rangle,$$

and then apply the juggle subroutine to the joint state of the $|b\rangle$ and $|x\rangle$ registers, ignoring the $|P_b(x)\rangle$ and $|h_b(x)\rangle$ registers as before.

Suppose $\|\Lambda_0 - \Lambda_1\| = 0$. Also, given $x \in \{0,1\}^n$ and $i \in \{0,1\}$, let $A_i = P_i^{-1}(P_i(x))$ and $H_i = h_i^{-1}(h_i(x))$, and suppose $2^{k-2} \leq |A_0| = |A_1| \leq 2^{k-1}$. Then

$$\Pr_{s,h_0,h_1} \left[|A_0 \cap H_0| = 1 \wedge |A_1 \cap H_1| = 1 \right] \geq \left(\frac{1}{4} \right)^2,$$

since the events $|A_0 \cap H_0| = 1$ and $|A_1 \cap H_1| = 1$ are independent of each other conditioned on x. Assuming both events occur, as before the juggle subroutine will reveal both $|0\rangle|x_0\rangle$ and $|1\rangle|x_1\rangle$ with high probability, where x_0 and x_1 are the unique elements of $A_0 \cap H_0$ and $A_1 \cap H_1$ respectively. By contrast, if $\|\Lambda_0 - \Lambda_1\| = 1$ then only one value of the $|b\rangle$ register will ever be observed. Again, replacing $\|\Lambda_0 - \Lambda_1\| = 0$ by $\|\Lambda_0 - \Lambda_1\| \leq 2^{-n^c}$, and $\|\Lambda_0 - \Lambda_1\| = 1$ by $\|\Lambda_0 - \Lambda_1\| \geq 1 - 2^{-n^c}$, can have only a negligible effect on the history distribution.

Of course, the probability that the correct value of k is chosen, and that $A_0 \cap H_0$ and $A_1 \cap H_1$ both have a unique element, could be as low as $1/(16n)$. To deal with this, we simply increase the number of calls to the juggle subroutine by an $O(n)$ factor, drawing new values of k, h_0, h_1 for each call. We pack multiple subroutine calls into a single oracle call as described in Section VII, except that now we uncompute the entire state (returning it to $|0 \cdots 0\rangle$) and then recompute it between subroutine calls. A final remark: since the algorithm that calls the history oracle is deterministic, we "draw" new values of k, h_0, h_1 by having $\mathcal{U}$ prepare a uniform superposition over all possible values. The indifference axiom justifies this procedure, by guaranteeing that within each call to the juggle subroutine, the hidden-variable values of k, h_0, and h_1 remain constant. ■

Let us end this section with some brief remarks about the oracle result of [6]. Given a function $g : \{0,1\}^n \to \{0,1\}^n$, the *collision problem* is to decide whether g is one-to-one or two-to-one, given that one of these is the case. The question is, how many queries to g are needed to solve this problem (where a query just returns $g(x)$ given x)? It is not hard to see that $\Theta\left(2^{n/2}\right)$ queries are necessary and sufficient for classical randomized algorithms. What we showed in [6] is that $\Omega\left(2^{n/5}\right)$ queries are needed by any *quantum* algorithm as well. Subsequently Shi [36] managed to improve the quantum lower bound to $\Omega\left(2^{n/3}\right)$ queries, thereby matching an upper bound of Brassard, Høyer, and Tapp [37]. On the other hand, the collision problem is easily reducible to the Statistical Difference problem, and is therefore solvable in polynomial time by sampling histories. This is the essence of the statement that $\mathsf{BQP} \neq \mathsf{DQP}$ relative to an oracle.

IX. SEARCH IN $N^{1/3}$ QUERIES

Given a Boolean function $f:\{0,1\}^n \rightarrow \{0,1\}$, the database search problem is simply to find a string x such that $f(x)=1$. We can assume without loss of generality that this "marked item" x is unique.[49] We want to find it using as few queries to f as possible, where a query returns $f(y)$ given y.

Let $N=2^n$. Then classically, of course, $\Theta(N)$ queries are necessary and sufficient. By querying f in superposition, Grover's algorithm [7] finds x using $O\left(N^{1/2}\right)$ queries, together with $\widetilde{O}\left(N^{1/2}\right)$ auxiliary computation steps (here the $\widetilde{O}$ hides a factor of the form $(\log N)^c$). Bennett et al. [38] showed that any quantum algorithm needs $\Omega\left(N^{1/2}\right)$ queries.

In this section, we show how to find the marked item by sampling histories, using only $O\left(N^{1/3}\right)$ queries and $\widetilde{O}\left(N^{1/3}\right)$ computation steps. Formally, the model is as follows. Each of the quantum circuits $U_1,\ldots,U_T$ that algorithm A gives to the history oracle $\mathcal{O}(\mathcal{T})$ is now able to query f. Suppose U_t makes q_t queries to f; then the total number of queries made by A is defined to be $Q=q_1+\cdots+q_T$. The total number of *computation* steps is at least the number of steps required to write down $U_1,\ldots,U_T$, but could be greater.

Theorem 10 *In the* DQP *model, we can search a database of N items for a unique marked item using $O\left(N^{1/3}\right)$ queries and $\widetilde{O}\left(N^{1/3}\right)$ computation steps.*

Proof. Assume without loss of generality that $N=2^n$ with $n|3$, and that each database item is labeled by an n-bit string. Let $x\in\{0,1\}^n$ be the label of the unique marked item. Then the sequence of quantum circuits $\mathcal{U}$ does the following: it first runs $O\left(2^{n/3}\right)$ iterations of Grover's algorithm, in order to produce the n-qubit state $\alpha\left|x\right\rangle+\beta\sum_{y\in\{0,1\}^n}\left|y\right\rangle$, where

$$\alpha=\sqrt{\frac{1}{2^{n/3}+2^{-n/3+1}+1}},$$
$$\beta=2^{-n/3}\alpha$$

(one can check that this state is normalized). Next $\mathcal{U}$ applies Hadamard gates to the first $n/3$ qubits. This yields the state

$$2^{-n/6}\alpha\sum_{y\in\{0,1\}^{n/3}}(-1)^{x_A\cdot y}\left|y\right\rangle\left|x_B\right\rangle+2^{n/6}\beta\sum_{z\in\{0,1\}^{2n/3}}\left|0\right\rangle^{\otimes n/3}\left|z\right\rangle,$$

where x_A consists of the first $n/3$ bits of x, and x_B consists of the remaining $2n/3$ bits. Let Y be the set of $2^{n/3}$ basis states of the form $\left|y\right\rangle\left|x_B\right\rangle$, and Z be the set of $2^{2n/3}$ basis states of the form $\left|0\right\rangle^{\otimes n/3}\left|z\right\rangle$.

Notice that $2^{-n/6}\alpha=2^{n/6}\beta$. So with the sole exception of $\left|0\right\rangle^{\otimes n/3}\left|x_B\right\rangle$ (which belongs to both Y and Z), the "marked" basis states in Y have the same amplitude as the "unmarked" basis states in Z. This is what we wanted. Notice also that, if we manage to find any $\left|y\right\rangle\left|x_B\right\rangle\in Y$, then we can find x itself using $2^{n/3}$ further classical queries: simply test all possible strings that end in x_B. Thus, the goal of our algorithm will be to cause the hidden variable to visit an element of Y, so that inspecting the variable's history reveals that element.

As in Theorem 9, the tools that we need are the juggle subroutine, and a way of reducing many basis states to two. Let s be drawn uniformly at random from $\{0,1\}^{n/3}$. Then $\mathcal{U}$ appends a third register to the state, and sets it equal to $\left|z\right\rangle$ if the first two registers have the form $\left|0\right\rangle^{\otimes n/3}\left|z\right\rangle$, or to $\left|s,y\right\rangle$ if they have the form $\left|y\right\rangle\left|x_B\right\rangle$. Disregarding the basis state $\left|0\right\rangle^{\otimes n/3}\left|x_B\right\rangle$ for convenience, the result is

$$2^{-n/6}\alpha\left(\sum_{y\in\{0,1\}^{n/3}}(-1)^{x_A\cdot y}\left|y\right\rangle\left|x_B\right\rangle\left|s,y\right\rangle+\sum_{z\in\{0,1\}^{2n/3}}\left|0\right\rangle^{\otimes n/3}\left|z\right\rangle\left|z\right\rangle\right).$$

Next $\mathcal{U}$ applies the juggle subroutine to the joint state of the first two registers. Suppose the hidden-variable value has the form $\left|0\right\rangle^{\otimes n/3}\left|z\right\rangle\left|z\right\rangle$ (that is, lies outside Y). Then with probability $2^{-n/3}$ over s, the first $n/3$ bits of z are equal to s. Suppose this event occurs. Then conditioned on the third register being $\left|z\right\rangle$, the reduced state of the first two registers is

$$\frac{(-1)^{x_A\cdot z_B}\left|z_B\right\rangle\left|x_B\right\rangle+\left|0\right\rangle^{\otimes n/3}\left|z\right\rangle}{\sqrt{2}},$$

where z_B consists of the last $n/3$ bits of z. So it follows from Section VII that with probability $\Omega(1/n)$, the juggle subroutine will cause the hidden variable to transition from $|0\rangle^{\otimes n/3}|z\rangle$ to $|z_B\rangle|x_B\rangle$, and hence from Z to Y.

The algorithm calls the juggle subroutine $\Theta\left(2^{n/3}n\right)=\Theta\left(N^{1/3}\log N\right)$ times, drawing a new value of s and recomputing the third register after each call. Each call moves the hidden variable from Z to Y with independent probability $\Omega\left(2^{-n/3}/n\right)$; therefore with high probability *some* call does so. Note that this juggling phase does not involve any database queries. Also, as in Theorem 9, "drawing" s really means preparing a uniform superposition over all possible s. Finally, the probability that the hidden variable ever visits the basis state $|0\rangle^{\otimes n/3}|x_B\rangle$ is exponentially small (by the union bound), which justifies our having disregarded it. ■

A curious feature of Theorem 10 is the tradeoff between queries and computation steps. Suppose we had run Q iterations of Grover's algorithm, or in other words made Q queries to f. Then provided $Q\leq\sqrt{N}$, the marked state $|x\rangle$ would have occurred with probability $\Omega\left(Q^2/N\right)$, meaning that $\widetilde{O}\left(N/Q^2\right)$ calls to the juggle subroutine would have been sufficient to find x. Of course, the choice of Q that minimizes $\max\left\{Q,N/Q^2\right\}$ is $Q=N^{1/3}$. On the other hand, had we been willing to spend $\widetilde{O}(N)$ computation steps, we could have found x with only a *single* query![50] Thus, one might wonder whether some other algorithm could push the number of queries below $N^{1/3}$, without simultaneously increasing the number of computation steps. The following theorem rules out that possibility.

Theorem 11 *In the* DQP *model,* $\Omega\left(N^{1/3}\right)$ *computation steps are needed to search an* N*-item database for a unique marked item. As a consequence, there exists an oracle relative to which* NP $\not\subset$ DQP*; that is,* NP*-complete problems are not efficiently solvable by sampling histories.*

Proof. Let $N=2^n$ and $f:\{0,1\}^n\rightarrow\{0,1\}$. Given a sequence of quantum circuits $\mathcal{U}=(U_1,\ldots,U_T)$ that query f, and assuming that $x\in\{0,1\}^n$ is the unique string such that $f(x)=1$, let $|\psi_t(x)\rangle$ be the quantum state after U_t is applied but before U_{t+1} is. Then the "hybrid argument" of Bennett et al. [38] implies that, by simply changing the location of the marked item from x to x^*, we can ensure that

$$\left\||\psi_t(x)\rangle-|\psi_t(x^*)\rangle\right\|=O\left(\frac{Q_t^2}{N}\right)$$

where $\|\ \|$ represents trace distance, and Q_t is the total number of queries made to f by $U_1,\ldots,U_t$. Therefore $O\left(Q_t^2/N\right)$ provides an upper bound on the probability of noticing the $x\rightarrow x^*$ change by monitoring v_t, the value of the hidden variable after U_t is applied. So by the union bound, the probability of noticing the change by monitoring the entire history $(v_1,\ldots,v_T)$ is at most of order

$$\sum_{t=1}^{T}\frac{Q_t^2}{N}\leq\frac{TQ_T^2}{N}.$$

This cannot be $\Omega(1)$ unless $T=\Omega\left(N^{1/3}\right)$ or $Q_T=\Omega\left(N^{1/3}\right)$, either of which implies an $\Omega\left(N^{1/3}\right)$ lower bound on the total number of steps.

To obtain an oracle relative to which NP $\not\subset$ DQP, we can now use a standard and well-known "diagonalization method" due to Baker, Gill, and Solovay [39] to construct an infinite sequence of exponentially hard search problems, such that any DQP machine fails on at least one of the problems, whereas there exists an NP machine that succeeds on all of them. We omit the details. ■

X. DISCUSSION

The idea that certain observables in quantum mechanics might have trajectories governed by dynamical laws has reappeared many times: in Schrödinger's 1931 stochastic approach [1], Bohmian mechanics [2], modal interpretations [5, 20, 21], and elsewhere. Yet because all of these proposals yield the same predictions for single-time probabilities, if we are to decide between them it must be on the basis of internal mathematical considerations. One message of this paper has been that such considerations can actually get us quite far.

To focus attention on the core issues, we restricted attention to the simplest possible setting: discrete time, a finite-dimensional Hilbert space, and a single orthogonal basis. Within this setting, we proposed what seem like reasonable axioms that any hidden-variable theory should satisfy: for example, indifference to the identity operation, robustness to small perturbations, and independence of the temporal order of spacelike-separated events. We then showed that not all of these axioms can be satisfied simultaneously. But perhaps more surprisingly, we also showed that certain subsets of axioms *can* be satisfied for quite nontrivial reasons. In showing that the indifference and robustness axioms can be simultaneously satisfied, Section V revealed an unexpected connection between unitary matrices and the classical theory of network flows.

As mentioned previously, an important open problem is to show that the Schrödinger theory satisfies robustness. Currently, we can only show that the matrix $P_{\mathcal{ST}}(\rho, U)$ is robust to *exponentially* small perturbations, not polynomially small ones. The problem is that if any row or column sum in the $U^{(t)}$ matrix is extremely small, then the (r, c)-scaling process will magnify tiny errors in the entries. Intuitively, though, this effect should be washed out by later scaling steps.

A second open problem is whether there exists a theory that satisfies indifference, as well as commutativity for all separable *mixed* states (not just separable pure states). A third problem is to investigate other notions of robustness—for example, robustness to small *multiplicative* rather than additive errors.

On the complexity side, perhaps the most interesting problem left open by this paper is the computational complexity of simulating Bohmian mechanics. We strongly conjecture that this problem, like the hidden-variable problems we have seen, is strictly harder than simulating an ordinary quantum computer. The trouble is that Bohmian mechanics does not quite fit in our framework: as discussed in Section II B, we cannot have deterministic hidden-variable trajectories for discrete degrees of freedom such as qubits. Even worse, Bohmian mechanics violates the continuous analogue of the indifference axiom. On the other hand, this means that by trying to implement (say) the juggle subroutine with Bohmian trajectories, one might learn not only about Bohmian mechanics and its relation to quantum computation, but also about how essential the indifference axiom really is for our implementation.

Another key open problem is to show better upper bounds on DQP. Recall that we were only able to show $\mathsf{DQP} \subseteq \mathsf{EXP}$, by giving a classical exponential-time algorithm to simulate the flow theory $\mathcal{FT}$. Can we improve this to (say) $\mathsf{DQP} \subseteq \mathsf{PSPACE}$? Clearly it would suffice to give a PSPACE algorithm that computes the transition probabilities for some theory $\mathcal{T}$ satisfying the indifference and robustness axioms. On the other hand, this might not be *necessary*—that is, there might be an indirect simulation method that does not work by computing (or even sampling from) the distribution over histories. It would also be nice to pin down the complexities of simulating specific hidden-variable theories, such as $\mathcal{FT}$ and $\mathcal{ST}$.

Acknowledgments

I thank Umesh Vazirani, Ronald de Wolf, and two anonymous reviewers for comments on earlier versions of this paper; Dorit Aharonov, Guido Bacciagaluppi, John Preskill, Rob Spekkens, Antony Valentini, and Avi Wigderson for helpful discussions; Andris Ambainis for correcting an ambiguity in the definition of DQP; Pioter Drubetskoy for pointing out a mistake in Section VIII; and Dennis Dieks for correspondence. This work was done while the author was a PhD student at UC Berkeley, supported by an NSF Graduate Fellowship.

[1] E. Schrödinger, Sitzungsber. Preuss. Akad. Wissen. Phys. Math. Kl. pp. 144–153 (1931).
[2] D. Bohm, Phys. Rev. **85**, 166 (1952).
[3] J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics* (Cambridge, 1987).
[4] E. Nelson, *Quantum Fluctuations* (Princeton, 1985).
[5] D. Dieks, Phys. Rev. A **49**, 2290 (1994).
[6] S. Aaronson, in *Proc. ACM STOC* (2002), pp. 635–642, quant-ph/0111102.
[7] L. K. Grover, in *Proc. ACM STOC* (1996), pp. 212–219, quant-ph/9605043.
[8] D. S. Abrams and S. Lloyd, Phys. Rev. Lett. **81**, 3992 (1998), quant-ph/9801041.
[9] T. Brun, Foundations of Physics Letters **16**, 245 (2003), gr-qc/0209061.
[10] D. Bacon (2003), quant-ph/0309189.
[11] D. Deutsch, Phys. Rev. D **44**, 3197 (1991).
[12] S. Aaronson, in *Proceedings of the Växjö Conference "Quantum Theory: Reconsideration of Foundations"*, edited by A. Khrennikov (2004), quant-ph/0401062.
[13] C. Philippidis, C. Dewdney, and B. J. Hiley, Nuovo Cimento **52B**, 15 (1979).
[14] E. Guay and L. Marchildon, J. Phys. A.: Math. Gen. **36**, 5617 (2003), quant-ph/0302085.
[15] M. Nagasawa, Prob. Theory and Related Fields **82**, 109 (1989).
[16] R. Sinkhorn, Ann. Math. Statist. **35**, 876 (1964).
[17] L. G. Valiant and V. V. Vazirani, Theoretical Comput. Sci. **47**, 85 (1986).
[18] C. Rovelli and L. Smolin, Nuclear Physics **B442**, 593 (1995), erratum in Vol. B456, p. 753. gr-qc/9411005.
[19] D. Bohm and B. Hiley, *The Undivided Universe* (Routledge, 1993).
[20] M. Dickson, in *Stanford Encyclopedia of Philosophy* (Stanford University, 2002), at http://plato.stanford.edu/entries/qm-modal/.

[21] G. Bacciagaluppi and M. Dickson, Found. Phys. **29**, 1165 (1999), quant-ph/9711048.
[22] R. B. Griffiths, Phys. Rev. A **57**, 1604 (1998), quant-ph/9708028.
[23] M. Gell-Mann and J. Hartle, in *Complexity, Entropy, and the Physics of Information*, edited by W. H. Zurek (Addison-Wesley, 1990).
[24] D. T. Gillespie, Phys. Rev. A **49**, 1607 (1994).
[25] A. Kitaev, Russian Math. Surveys **52**, 1191 (1997).
[26] M. Nielsen and I. Chuang, *Quantum Computation and Quantum Information* (Cambridge, 2000).
[27] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms (2nd edition)* (MIT Press, 2001).
[28] L. R. Ford and D. R. Fulkerson, *Flows in Networks* (Princeton, 1962).
[29] R. Fortet, J. Math Pures et. Appl. **9**, 83 (1940).
[30] A. Beurling, Ann. Math. **72**, 189 (1960).
[31] J. Franklin and J. Lorenz, Linear Algebra Appl. **114/115**, 717 (1989).
[32] N. Linial, A. Samorodnitsky, and A. Wigderson, Combinatorica **20**, 545 (2000).
[33] A. Sahai and S. Vadhan, J. ACM **50**, 196 (2003), eCCC TR00-084. Earlier version in IEEE FOCS 1997.
[34] O. Goldreich and S. Goldwasser, in *Proc. ACM STOC* (1998), pp. 1–9.
[35] D. Aharonov and A. Ta-Shma, in *Proc. ACM STOC* (2003), pp. 20–29, quant-ph/0301023.
[36] Y. Shi, in *Proc. IEEE FOCS* (2002), pp. 513–519, quant-ph/0112086.
[37] G. Brassard, P. Høyer, and A. Tapp, ACM SIGACT News **28**, 14 (1997), quant-ph/9705002.
[38] C. Bennett, E. Bernstein, G. Brassard, and U. Vazirani, SIAM J. Comput. **26**, 1510 (1997), quant-ph/9701001.
[39] T. Baker, J. Gill, and R. Solovay, SIAM J. Comput. **4**, 431 (1975).
[40] See www.complexityzoo.com for more information about the complexity classes mentioned in this paper.
[41] For readers unfamiliar with asymptotic notation: $O(f(N))$ means "at most order $f(N)$," $\Omega(f(N))$ means "at least order $f(N)$," and $\Theta(f(N))$ means "exactly order $f(N)$."
[42] For as Abrams and Lloyd [8] observed, we can so arrange things that $|\psi\rangle = |0\rangle$ if an NP-complete instance of interest to us has no solution, but $|\psi\rangle = \sqrt{1-\varepsilon}\,|0\rangle + \sqrt{\varepsilon}\,|1\rangle$ for some tiny ε if it has a solution.
[43] One can define other, less motivated, models with the same property by allowing "non-collapsing measurements" of quantum states, but these models are very closely related to ours. Indeed, a key ingredient of our results will be to show that certain kinds of non-collapsing measurements can be *simulated* using histories.
[44] Put differently, Bohm's conservation of probability result breaks down because the "wavefunctions" at t_0 and t_1 are degenerate, with all amplitude concentrated on finitely many points. But in a discrete Hilbert space, *every* wavefunction is degenerate in this sense!
[45] In an earlier version of this paper, there were two more axioms: symmetry under relabeling of basis states and a weaker version of robustness. We have omitted these axioms because they are largely irrelevant for our results.
[46] Dieks (personal communication) says he would no longer defend this theory.
[47] In (r,c)-scaling, we are given an invertible real matrix, and the goal is to rescale all rows and columns to sum to 1. The generalized version is to rescale the rows and columns to given values (not necessarily 1).
[48] Note that in this lemma, the constants $1/3$ and $2/3$ are not arbitrary; it is important for technical reasons that $(2/3)^2 > 1/3$.
[49] For if there are multiple marked items, then we can reduce to the unique marked item case by using the Valiant-Vazirani hashing technique described in Theorem 9.
[50] One should not make too much of this fact; one way to interpret it is simply that the "number of queries" should be redefined as $Q + T$ rather than Q.