Quantum interference as a resource for quantum speedup

Dan Stahlke*
Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA
(Received 16 December 2013; published 1 August 2014)

Quantum states can, in a sense, be thought of as generalizations of classical probability distributions, but are more powerful than probability distributions when used for computation or communication. Quantum speedup therefore requires some feature of quantum states that classical probability distributions lack. One such feature is interference. We quantify interference and show that there can be no quantum speedup due to a small number of operations incapable of generating large amounts of interference (although large numbers of such operations can, in fact, lead to quantum speedup). Low-interference operations include sparse unitaries, Grover reflections, short-time and low-energy Hamiltonian evolutions, and the Haar wavelet transform. Circuits built from such operations can be classically simulated via a Monte Carlo technique making use of a convex combination of two Markov chains. Applications to query complexity, communication complexity, and the Wigner representation are discussed.

DOI: 10.1103/PhysRevA.90.022302 PACS number(s): 03.67.Ac

I. INTRODUCTION

It is well known that certain quantum algorithms, such as Shor's and Grover's, provide a speedup compared to classical algorithms. However, the source of such quantum speedup is still somewhat of a mystery. Insight can be gained by determining necessary resources. Suppose that any quantum circuit not making use of some resource X can be efficiently simulated. Being efficiently simulated, such circuits do not exhibit quantum speedup. One can then conclude that resource X is necessary for quantum speedup. Many such resources have been identified. For circuits on pure states there is no quantum speedup if at all times (i.e., before and after every unitary) the state has a small Schmidt rank [1] or factors into a product state on small subsystems [2]. For qubit circuits there is no quantum speedup if the discord across all bipartite cuts is zero at all times [3]. There is no quantum speedup for circuits that use only Clifford gates [4], or matchgates [5], that have small tree width [6,7], or that use only operations having non-negative Wigner representation [8–10]. For a brief overview of resources identified as important for quantum speedup, see Sec. 9 of [11].

A tempting but naive explanation for quantum speedup is the exponentially large dimensionality of Hilbert space (2^n for n qubits), combined with "quantum parallelism." Shor's algorithm begins by preparing a state $\sqrt{2^{-n}}\sum_x |x\rangle \otimes |f(x)\rangle$, which can be interpreted as simultaneously evaluating f for all 2^n values of x. However, this is not a satisfactory explanation for quantum speedup since classical probability distributions over n bits can also be considered as vectors of dimension 2^n and allow a similar sort of parallelism. We show that the quantum speedup is connected to *interference*, something which classical probability distributions lack. Prior works have mentioned interference as being important for quantum speedup but without offering a quantitative definition [12–15] or have quantified interference without providing a strong connection to speedup [16].

We consider quantum circuits composed of an initial state, followed by several unitary operators, and terminated by measurement of a Hermitian observable. The expectation value of this measurement can be written as a sum of Feynman-like paths in the computational basis, and this sum can be estimated via a Monte Carlo technique that considers an ensemble of paths drawn according to a suitable probability distribution. The required size of the ensemble is lower bounded by the square of the interference, which we define as a sum of absolute values of the path amplitudes (Definition 1). We are not able to reach this lower bound; however, by using a convex combination of a pair of Markov chains we are able to provide a simulation algorithm that runs in time quadratic in the product of the *interference-producing capacities* of each operator in the circuit, defined as the largest amount of interference an operator is capable of producing (Definition 2). This ends up being equal to the largest singular value of the entrywise absolute value of the operator in the computational basis. Briefly, we can estimate expressions of the form $\langle\psi|A\cdots Z|\phi\rangle$, of which quantum circuits $\langle\psi|U^{(1)\dagger}\cdots U^{(T)\dagger}MU^{(T)}\cdots U^{(1)}|\psi\rangle$ are a special case, in time proportional to $\|\bar{A}\|_2^2\cdots\|\bar{Z}\|_2^2$, where $\|\cdot\|_2$ denotes maximum singular value and where a bar over an operator denotes entrywise absolute value in the computational basis. This work was inspired by, and extends, Ref. [15], which provides an efficient simulation when $A,\ldots,Z$ are all sparse.

Operations with small interference-producing capacity include the *efficiently computable sparse* operations as defined in [15] (e.g., permutation matrices and gates acting on a constant number of qubits), as well as the Grover reflection operation, short-time and low-energy Hamiltonian evolutions, and the Haar wavelet transform. Our simulation algorithm will generally be exponentially slow in the length of the circuit, but for the classes of gates listed in the previous sentence it has only polynomial dependence on the number of qubits. An example of a circuit that apparently uses much "quantum magic," but which can nevertheless be simulated in a time polynomial in the number of qubits, is depicted in Fig. 1.

We (of course) cannot efficiently simulate Shor's algorithm. However, replacing the Fourier transform with the Haar transform, which has low interference-producing capacity, yields a circuit that we can simulate (Fig. 2). We show that

*dan@stahlke.org

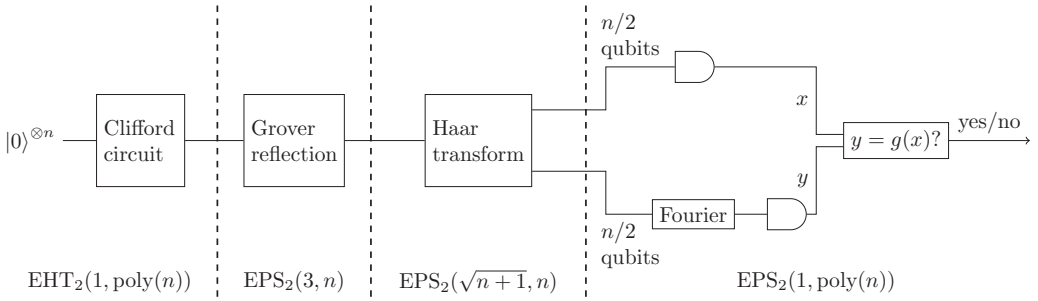

FIG. 1. An example of the type of circuit that can be simulated in poly(n) time using the techniques of this paper. The circuit is divided into four sections: The first section is considered to be the initial state, the middle two sections are unitary matrices, and the last section is a projector. The block labeled $y = g(x)$ represents a classical computation step that outputs "yes" if the first and second measurement operations result in values that are related by an arbitrary [but poly(n) time computable] function g.

there is no quantum advantage for communication protocols that use small interference, although curiously this result does not apply to one-round communication protocols. To our knowledge, interference-producing capacity is the first continuous-valued quantity that has been shown necessary for quantum speedup, escaping the theorem of [17], which shows that a large class of continuous-valued quantities, such as entanglement and discord, are not necessary for quantum speedup.

In Secs. II and III we explain our method for estimating expectation values using a Monte Carlo technique with Markov chains. In Sec. IV we formalize and extend this technique and provide guarantees on runtime. In Sec. V we characterize the types of quantum circuits that our technique can efficiently simulate, and explore a variety of circuits that we cannot efficiently simulate. Section VI discusses further applications, including the Wigner representation and communication complexity. In Sec. VII we formalize our conjecture that interference, rather than interference-producing capacity, is required for quantum speedup. Nontrivial proofs are deferred to Appendices.

II. MONTE CARLO TECHNIQUE

A. Sampling of paths

We make use of the following circuit model. Let ρ be an initial density operator. This state is acted upon by a sequence of unitaries $U^{(1)}, \ldots, U^{(T)}$. Finally, a Hermitian observable (e.g., a projector) M is measured. It is not assumed that the unitary operations or the final observable are local; they can be arbitrary operations potentially involving all qubits or qudits (e.g., a quantum Fourier transform). The expectation value of this final measurement is

$$\mathrm{Tr}\{U^{(1)\dagger} \cdots U^{(T)\dagger} M U^{(T)} \cdots U^{(1)} \rho\}. \tag{1}$$

Our goal is to estimate this expectation value to within small additive error, using a classical computer. We allow the unitaries to be oracle operations (as in Grover's algorithm), in which case we grant the classical computer that runs the simulation access to an equivalent oracle (this is further discussed in Sec. IV C).

This is not the most general type of simulation. In particular, we do not consider the case of a many-outcome measurement (e.g., individual measurements on several qubits or a measurement given by a projective decomposition of the identity) in which the simulation is required to produce individual outcomes according to the same probability distribution with which the quantum circuit produces those outcomes. The ability to estimate the expectation value of a projector to within small multiplicative error would allow simulation of such sampling, as discussed in [18]; however, the algorithm of the present paper only estimates to within additive error.

Although our primary goal is to estimate expressions of the form (1), we generalize the task by considering products of the

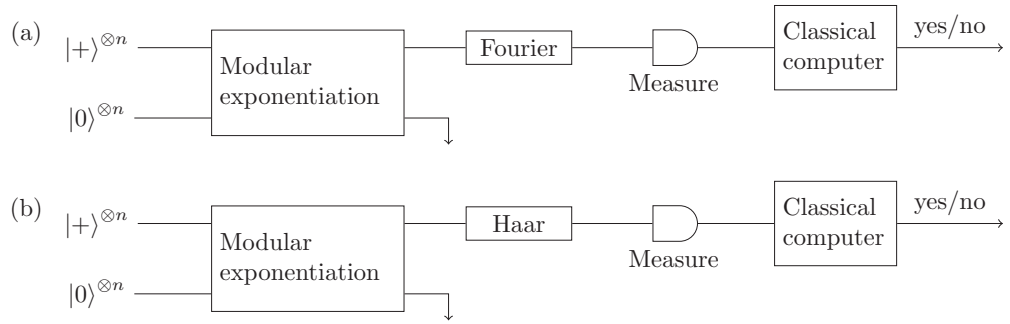

FIG. 2. (a) A depiction of the decisional version of Shor's algorithm, which outputs "yes" if there is a prime factor within some given range. (b) The Haar wavelet transform (Definition 10) plays a similar role as the Fourier transform in classical signal processing. However, substituting the Haar transform for the Fourier transform in Shor's algorithm yields a circuit that can be efficiently simulated on a classical computer. Note that the resulting circuit will not factor numbers and, in fact, probably has no practical use.

form $\mathrm{Tr}\{A^{(1)}\cdots A^{(S)}\sigma\}$, where σ and the $A^{(s)}$ are matrices, not necessarily unitary or Hermitian, and possibly rectangular (we label σ separately from the $A^{(s)}$ in anticipation of the results of the next section). This product can be written as a sum over paths,

$$\mathrm{Tr}\{A^{(1)}\cdots A^{(S)}\sigma\} = \sum_{i_0\cdots i_S} A^{(1)}_{i_0 i_1}\cdots A^{(S)}_{i_{S-1} i_S}\sigma_{i_S i_0}. \tag{2}$$

Or, by defining the tuple index $\pi = (i_0\cdots i_S)$, this can be written as

$$\mathrm{Tr}\{A^{(1)}\cdots A^{(S)}\sigma\} = \sum_{\pi} V(\pi), \tag{3}$$

$$V(\pi) = A^{(1)}_{i_0 i_1}\cdots A^{(S)}_{i_{S-1} i_S}\sigma_{i_S i_0}. \tag{4}$$

Our strategy is to estimate this sum by drawing a reasonably small number of paths π according to a probability distribution, denoted $R(\pi)$. Any probability distribution can be used, although some are more suitable than others. Finding a good $R(\pi)$ will be a central goal of this section and the next. Denote by Π a random variable that takes value π with probability $R(\pi)$. Consider the expectation value of $V(\Pi)/R(\Pi)$:

$$\mathbb{E}\left[\frac{V(\Pi)}{R(\Pi)}\right] = \sum_{\pi}\frac{V(\pi)}{R(\pi)}R(\pi) \tag{5}$$

$$= \sum_{\pi} V(\pi). \tag{6}$$

By the weak law of large numbers, $\sum_\pi V(\pi)$ can be approximated to arbitrary accuracy by computing the mean of sufficiently many samples of $V(\Pi)/R(\Pi)$; however, the efficiency of this strategy hinges on two things. First, it must be possible using a classical computer to efficiently draw random samples according to the probability distribution $R(\pi)$ and to compute the corresponding values $V(\pi)/R(\pi)$. This is an important point to which we return to throughout the paper. Second, the sample mean of $V(\Pi)/R(\Pi)$ must rapidly converge to its expectation value. The Chernoff-Hoeffding bound states that for a random variable whose magnitude is bounded by b, the mean of $O(\epsilon^{-2}b^2)$ samples is very likely to approximate the expectation value to within additive error ϵ. Thus, there is rapid convergence when $\max_\pi\{|V(\pi)|/R(\pi)\}$ is small. Note that this is a sufficient but not necessary condition for rapid convergence; for example, considering the variance of $V(\Pi)/R(\Pi)$ could in some cases reveal that convergence happens more rapidly.

We now present the Chernoff-Hoeffding bound in one of its standard forms, along with a corollary that adapts it to our application.

Theorem 1. Chernoff-Hoeffding bound [19]. Let $X_1,\ldots,X_K$ be independent identically distributed real-valued random variables with expectation value $\mathbb{E}[X]$ and satisfying $|X_k|\leqslant b$. Let $\epsilon>0$. Then

$$\Pr\left\{\left|\frac{1}{K}\sum_{k=1}^{K}X_k - \mathbb{E}[X]\right| > \epsilon\right\} \leqslant 2e^{-K\epsilon^2/2b^2}. \tag{7}$$

Corollary 1. Let $V(\pi)$ be a complex valued function of π and $R(\pi)$ be a probability distribution. Define

$$b_{\max} = \max_{\pi}\left\{\frac{|V(\pi)|}{R(\pi)}\right\}. \tag{8}$$

Let $\epsilon,\delta>0$. Then, with probability less than δ of exceeding the error bound, $\sum_\pi V(\pi)$ can be estimated to within additive error ϵ using $O(\log_2(\delta^{-1})\epsilon^{-2}b^2_{\max})$ draws from the distribution $R(\pi)$ and the same number of evaluations of $V(\pi)/R(\pi)$.

Proof. It can be shown[1] that Theorem 1 can be extended to complex variables at the expense of replacing the right-hand side of (7) by $4e^{-K\epsilon^2/4b^2}$. Define the independent identically distributed random variables $X_k = V(\Pi_k)/R(\Pi_k)$ with $k\in\{1,\ldots,K\}$. Applying the complex valued version of Theorem 1, and noting that $|X_k|\leqslant b_{\max}$ and $\mathbb{E}[V(\Pi)/R(\Pi)] = \sum_\pi V(\pi)$, we get

$$\Pr\left\{\left|\frac{1}{K}\sum_{k=1}^{K}\frac{V(\Pi_k)}{R(\Pi_k)} - \sum_{\pi}V(\pi)\right| > \epsilon\right\} \leqslant 4e^{-K\epsilon^2/4b^2_{\max}}. \tag{9}$$

Setting $K = \ln(4/\delta)4\epsilon^{-2}b^2_{\max} = O(\log_2(\delta^{-1})\epsilon^{-2}b^2_{\max})$ makes the right hand side of (9) equal to δ. ■

Since the number of samples needed depends only logarithmically on δ, it is possible to choose δ to be extremely small (say, 1×10^{-9}) while having only minimal impact on the number of samples needed. With such a small δ, the estimate will be very likely to be within additive error ϵ.

The number of samples needed for an accurate estimate is quadratic in $b_{\max}$, so finding an $R(\pi)$ for which $b_{\max}$ is small is of crucial importance. However, feasibility of the simulation also depends on the difficulty of drawing random paths π according to the distribution $R(\pi)$ and computing the corresponding values $V(\pi)/R(\pi)$. We denote by the letter f the time needed to carry out these operations. Specifically, we require that sampling from $R(\pi)$ and computing $V(\pi)/R(\pi)$ can be carried out in average time $O(f)$, where f is some function of the dimension or number of qubits of a quantum circuit. Since $\sum_\pi V(\pi)$ can be estimated by averaging $O(\log_2(\delta^{-1})\epsilon^{-2}b^2_{\max})$ samples of $V(\Pi)/R(\Pi)$, each of which can be computed in time $O(f)$, the total runtime of the algorithm is $O(\log_2(\delta^{-1})\epsilon^{-2}b^2_{\max}f)$.

Some probability distributions are easier to sample from than others, and this needs to be decided on a case-by-case basis. For example, consider $R(i) = |\psi_i|^2$, where $|\psi\rangle$ is a quantum state. If $|\psi\rangle$ is a computational basis state, then $R(i)$ is rather trivial and can be sampled by simply outputting the sole index i for which $R(i)\neq0$. If $|\psi\rangle$ is a graph state on n qubits, then $R(i)$ is the uniform distribution over the 2^n basis states. This can be sampled in time $O(n)$ by tossing a fair coin n times, once for each qubit, so in this case $f=n$. On the other hand, if $|\psi\rangle$ is defined as being the state just before the final measurement in Shor's algorithm, then it is probably not feasible to sample from $R(i)$ efficiently on a classical computer.

[1]This is shown by applying Theorem 1 separately to the real and imaginary parts and using the fact that the sample mean is within additive error ϵ of the expectation value as long as both the real and the imaginary parts are within $\epsilon/\sqrt{2}$.

For simplicity we assume that all operations can be carried out with perfect computational accuracy, including the degree to which the probability distribution of the generated samples π agrees with an ideal distribution $R(\pi)$, and the precision of the computed $V(\pi)/R(\pi)$ values. Of course, computers can only compute with finite precision. However, since we are concerned only with approximating expectation values to within additive error ϵ, carrying out the computations to finite but high precision is sufficient as long as the total accumulated computational error is small compared to the error tolerance ϵ. This is discussed in more detail in Appendix A of [15].

B. Interference

An efficient simulation requires choosing a probability distribution $R(\pi)$ for which $b_{\max}$ of (8) is not large. A tempting choice is

$$R_{\text{opt}}(\pi) := \frac{|V(\pi)|}{\sum_{\pi'} |V(\pi')|}. \tag{10}$$

It can be shown[2] that this is the unique distribution yielding the minimum possible value of $b_{\max}$,

$$b_{\text{opt}} = \textstyle\sum_{\pi} |V(\pi)|. \tag{11}$$

Being the lowest possible value of $b_{\max}$, (11) represents a lower bound on the number of samples needed as guaranteed by the Chernoff-Hoeffding bound, although a more careful analysis of variances, for instance, could show that the algorithm actually produces a faster-than-expected convergence.

An efficient algorithm requires both that $b_{\max}$ be small and that $R(\pi)$ can be sampled from efficiently. We do not know of a way to efficiently sample from the probability distribution (10) in general, so this is not useful for computing the expectation value. Nevertheless, it is worthwhile to discuss for a moment the case where the one condition is met (small $b_{\max}$) even if the other condition is not met (ability to efficiently draw samples). For concreteness, consider a simple quantum circuit with only one unitary, $\text{Tr}\{U^\dagger M U \rho\}$. This can be written as a sum over paths,

$$\text{Tr}\{U^\dagger M U \rho\} = \sum_{\pi} V(\pi), \tag{12}$$

with $\pi = (i,j,k,l)$ and $V(i,j,k,l) = U^\dagger_{ij} M_{jk} U_{kl} \rho_{li}$. Plugging this into (11) gives

$$b_{\text{opt}} = \text{Tr}\{\bar{U}^\dagger \bar{M} \bar{U} \bar{\rho}\}, \tag{13}$$

where a bar over a vector or matrix denotes entrywise absolute value in the computational basis, a notation that is used throughout this paper. This generalizes in the obvious way for circuits with more than one unitary.

Comparing (11) and (12), both are sums over paths but the latter involves an absolute value for each path. The sum (12) has magnitude bounded by 1 if the observable M has eigenvalues bounded in magnitude by 1. The sum (11), on the other hand, can take a much larger value than (12) when the terms in the latter sum exhibit cancellations due to destructive interference. For example, consider the case $|\psi\rangle = N^{-1/2} \sum_i |i\rangle$, U the Fourier transform, and M the identity, giving $b_{\text{opt}} = \sqrt{N}$.

It may be enlightening to consider a physical example. To this end, we introduce a simple toy-model version of Young's double-slit experiment. Let states $|0\rangle$ and $|1\rangle$ represent a particle immediately exiting the upper and lower slits, respectively, and let $|x\rangle$ represent a particle impacting the detector at position x. The transfer operator representing passage of the particle from the slits to the detector will be some unitary U satisfying $U(\alpha|0\rangle + \beta|1\rangle) = \int_x (\alpha\psi_x + \beta\phi_x)|x\rangle dx$. A particle passing through the upper slit will impact the detector at position x with probability density $|\psi_x|^2$; for a particle passing through the lower slit the probability density is $|\phi_x|^2$. A particle in a superposition of passing through upper and lower slits, in state $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$, will impact the screen at x with probability density

$$\left|\frac{1}{\sqrt{2}}\psi_x + \frac{1}{\sqrt{2}}\phi_x\right|^2 = \frac{1}{2}|\psi_x|^2 + \frac{1}{2}|\phi_x|^2 + \text{Re}(\psi_x^*\phi_x). \tag{14}$$

The first two terms on the right-hand side represent the probability that would be expected if the particle were in a classical stochastic mixture of passing through one slit or the other. The third is the interference term. Integrating this term over x yields zero, as it must in order for the probabilities to sum to 1. The total amount of interference can be quantified by instead integrating the absolute value of this term. Similarly, if we were interested in only part of the detector, say $x \in [0,1]$, the interference associated with that region could be defined by integrating only over this range. It turns out to be more mathematically convenient to include all three terms in the definition of interference; for one thing, the resulting quantity will be multiplicative when considering a system composed of noninteracting subsystems. The $|\psi_x|^2/2 + |\phi_x|^2/2$ terms contribute at most 1 (exactly 1 if integrating over the entire range). In summary, we may define the interference associated with the $x \in [0,1]$ region of the detector as

$$\mathcal{I} = \int_{x\in[0,1]} \left(\frac{1}{2}|\psi_x|^2 + \frac{1}{2}|\phi_x|^2 + |\psi_x^*\phi_x|\right) dx. \tag{15}$$

This is essentially what is done in (13). Specifically, setting $\rho = |+\rangle\langle+|$ and $M = \int_{x\in[0,1]} |x\rangle\langle x| dx$ in (13) yields

$$b_{\text{opt}} = \int_{x\in[0,1]} \left(\frac{1}{\sqrt{2}}|\psi_x| + \frac{1}{\sqrt{2}}|\phi_x|\right)^2 dx \tag{16}$$

$$= \int_{x\in[0,1]} \left(\frac{1}{2}|\psi_x|^2 + \frac{1}{2}|\phi_x|^2 + |\psi_x^*\phi_x|\right) dx. \tag{17}$$

Note that (13) depends upon the choice of basis since the entrywise absolute value is basis dependent. Typically, one has some canonical basis in mind; for example, when one says that the double-slit experiment exhibits interference, this is relative to the position basis. For quantum circuits there is the computational basis, although in the interest of efficient simulation one may choose to use some other basis.

For a more complicated apparatus, such as a network of beam splitters, similar arguments apply: We quantify interference by computing a sum over paths, summing the

[2]Let $R(\pi)$ be any probability distribution that differs from $R_{\text{opt}}(\pi)$ of (10). Then there must be a π' such that $R(\pi') < R_{\text{opt}}(\pi')$. It follows that $\max_\pi \{|V(\pi)|/R(\pi)\} > |V(\pi')|/R_{\text{opt}}(\pi') = \sum_\pi |V(\pi)|$.

absolute value of each path contribution. This definition depends upon a choice of course graining. For instance, a box which simply passes a photon from input to output undisturbed could be said to contribute no interference. On the other hand, if one were to take a more detailed view of this box—suppose, for example, that it contains a perfectly balanced Mach-Zehnder interferometer—then one could conclude that there is, in fact, interference. The same applies to simulation of quantum circuits. Although our simulation technique has difficulty simulating the Fourier transform, a Fourier transform followed by its inverse presents no difficulty if one course grains the circuit by replacing $F^\dagger F$ with the identity.

The above considerations lead to the following definition.

Definition 1. The *interference* of a quantum circuit with initial state ρ, unitary operators $U^{(1)}, \ldots, U^{(T)}$, and measurement M is

$$\mathcal{I}(U^{(1)\dagger}, \ldots, U^{(T)\dagger}, M, U^{(T)}, \ldots, U^{(1)}, \rho) = \mathrm{Tr}\{\bar{U}^{(1)\dagger} \cdots \bar{U}^{(T)\dagger} \bar{M} \bar{U}^{(T)} \cdots \bar{U}^{(1)} \bar{\rho}\}. \quad (18)$$

More generally, the interference of an arbitrary expression of the form $\mathrm{Tr}\{A^{(1)} \cdots A^{(S)} \sigma\}$ is

$$\mathcal{I}(A^{(1)}, \ldots, A^{(S)}, \sigma) = \mathrm{Tr}\{\bar{A}^{(1)} \cdots \bar{A}^{(S)} \bar{\sigma}\}. \quad (19)$$

This definition depends on the choice of basis. Unless otherwise specified, the standard (aka computational) basis is used.

With this definition, we have that $b_{\max} \geqslant \mathcal{I}(U^\dagger, M, U, \rho)$ in (8) for any choice of probability distribution, with equality when the distribution (10) is used. Since the number of samples needed to estimate the expectation value using our technique is proportional to $b_{\max}^2$, any quantum circuit with very large interference could never feasibly be simulated with our technique, no matter the choice of $R(\pi)$.

While we do not know how to efficiently sample from the optimal probability distribution (10), we conjecture that there is still some way to efficiently estimate the expectation value of a quantum circuit in cases where the interference is low. The precise statement of this conjecture is a delicate matter taken up in Sec. VII. We, however, show, by the end of the next section, that it is possible to simulate circuits in which each unitary as well as the final observable has a low *interference-producing capacity* (Definition 2).

A connection between $\mathcal{I}$ and the decoherence functional of Gell-Mann and Hartle is discussed in Sec. VI D.

III. MARKOV CHAINS

A. Introduction

The problem with the probability distribution (10) is that there is no obvious way to efficiently sample from it using a classical computer. So while only $O(\log_2(\delta^{-1})\epsilon^{-2}\mathcal{I}^2)$ samples are needed (with $\mathcal{I}$ given by Definition 1), each sample may be very complicated to evaluate. The essence of the difficulty is that this distribution treats the circuit holistically, so drawing samples apparently requires an understanding of how all the factors of (2) interact with each other. In order to avoid this problem, we instead use a probability distribution defined in terms of a time-inhomogeneous Markov chain with a transition corresponding to each operator in (2). More precisely, we take the convex combination of two (unrelated) Markov chains, one proceeding left to right and the other proceeding right to left. This way, it is only necessary to understand each individual operator, not the interactions between operators. The computation time of this simulation will end up being related not to the interference $\mathcal{I}$ but rather the product of the interference-producing capacities of each factor (a term that is defined at the end of this section).

The end result of this section is an algorithm for estimating products of the form $\mathrm{Tr}\{A^{(1)} \cdots A^{(S)} \sigma\}$, where σ and the $A^{(t)}$ are matrices, not necessarily unitary or Hermitian and possibly rectangular. This includes as a special case quantum circuits of the form $\mathrm{Tr}\{U^{(1)\dagger} \cdots U^{(T)\dagger} M U^{(T)} \cdots U^{(1)} \rho\}$. We build the algorithm step by step, considering first an example that demonstrates why a convex combination of probability distributions is needed, considering second an example that explains how the Markov chains are built, and using finally a convex combination of Markov chains. The exposition in this section is meant to be instructive; formal theorems are taken up in Sec. IV.

B. Inner product

Consider the task of estimating the inner product $\langle\psi|\phi\rangle = \sum_i \psi_i^* \phi_i$, where the two vectors satisfy the property $\|\psi\|_p = \|\phi\|_q = 1$ with $1/p + 1/q = 1$.[3] In the context of quantum circuits $p = q = 2$ is the natural choice; however, we allow general ℓ^p norms because the case $p = 1$, $q = \infty$ is also important and because the general case may be of independent interest. Here, as in the more general case that follows, the key is to find a probability distribution $R(i)$ that will be suitable for application of Corollary 1 . It is needed that

$$b_{\max} = \max_i \left\{ \frac{|V(i)|}{R(i)} \right\} = \max_i \left\{ \frac{|\psi_i^* \phi_i|}{R(i)} \right\} \quad (20)$$

is not large. There are two obvious choices for the probability distribution: $P(i) = |\psi_i|^p$ and $Q(i) = |\phi_i|^q$. Unfortunately, neither of these will guarantee a small $b_{\max}$. However, for each i at least one of the distributions $P(i)$ or $Q(i)$ will work well. The solution is to take a convex combination of these two distributions,

$$R(i) = \frac{1}{p} P(i) + \frac{1}{q} Q(i). \quad (21)$$

The algorithm that follows is an adaptation of one that appears in [15] (they used $p = q = 2$ and a slightly different technique). We present it as a formal theorem, in order to demonstrate how to carefully track the algorithm's time complexity.

Example 1. Let $1 \leqslant p \leqslant \infty$ and $1/p + 1/q = 1$. Let $|\psi\rangle$ and $|\phi\rangle$ be vectors with $\|\psi\|_p = \|\phi\|_q = 1$. Suppose that it is possible to sample from the probability distributions $P(i) = |\psi_i|^p$ and $Q(i) = |\phi_i|^q$ and to compute entries ψ_i and ϕ_i in average time $O(f)$ for some f. It is possible, with probability less than $\delta > 0$ of exceeding the error bound, to

[3] The ℓ^p norm, $\|\cdot\|_p$, is defined as $\|\psi\|_p = (\sum_i |\psi_i|^p)^{1/p}$ when $1 \leqslant p < \infty$ and $\|\psi\|_p = \max_i |\psi_i|$ when $p = \infty$. When $1/p + 1/q = 1$, the norms $\|\cdot\|_p$ and $\|\cdot\|_q$ are dual to each other.

estimate $\langle\psi|\phi\rangle$ to within additive error $\epsilon > 0$ in average time $O(\log_2(\delta^{-1})\epsilon^{-2} f)$.

Proof. Let $V(i) = \psi_i^* \phi_i$ and $R(i) = P(i)/p + Q(i)/q$. To apply Corollary 1 we need to bound $b_{\max} = \max_i\{|V(i)|/R(i)\}$. Making use of the (weighted) inequality of arithmetic and geometric means[4]

$$b_{\max} = \max_i\{|V(i)|/R(i)\} \tag{22}$$

$$= \max_i\{|\psi_i^*\phi_i|/[P(i)/p + Q(i)/q]\} \tag{23}$$

$$\leqslant \max_i\{|\psi_i^*\phi_i|/[P(i)^{1/p} Q(i)^{1/q}]\} \tag{24}$$

$$= 1. \tag{25}$$

By Corollary 1, $\langle\psi|\phi\rangle = \sum_i V(i)$ can be estimated at the cost of drawing $O(\log_2(\delta^{-1})\epsilon^{-2})$ samples i according to $R(i)$ and computing the corresponding $V(i)/R(i)$ values. Sampling from $R(i)$ can be accomplished as follows: Flip a biased coin that lands heads up with probability $1/p$. If it lands heads up, then draw i from $P(i)$, otherwise draw i from $Q(i)$. By assumption, this takes an average time $O(f)$. Next, $V(i)/R(i)$ can be computed directly from ψ_i and ϕ_i, each of which can, in turn, be computed in average time $O(f)$. The $O(\log_2(\delta^{-1})\epsilon^{-2})$ samples (as well as their mean) can therefore be computed in average time $O(\log_2(\delta^{-1})\epsilon^{-2} f)$. ∎

C. Nearly stochastic matrices

We now move to a more general case, the estimation of $\langle\psi|A^{(1)}\cdots A^{(S)}|\phi\rangle$. For the sake of simplicity, suppose that there are only two operators (i.e., $S = 2$) so that the goal is to estimate $\langle\psi|AB|\phi\rangle$. This can be written as a sum over paths as in (2),

$$\langle\psi|AB|\phi\rangle = \sum_{ijk} \psi_i^* A_{ij} B_{jk} \phi_k. \tag{26}$$

To apply Corollary 1 to this problem, set $\pi = (i,j,k)$ and $V(i,j,k) = \psi_i^* A_{ij} B_{jk} \phi_k$. For efficient simulation it suffices to find a probability distribution $P(i,j,k)$ from which we can efficiently draw samples using a classical computer, for which $V(i,j,k)/P(i,j,k)$ can be efficiently computed and for which

$$b_{\max} = \max_{ijk}\left\{\frac{|\psi_i^* A_{ij} B_{jk}\phi_k|}{P(i,j,k)}\right\} \tag{27}$$

is small enough that the estimation will converge reasonably fast. As discussed in the previous section, a tempting choice for the probability distribution is given by (10); however, it is not clear how one would efficiently draw samples from this since doing so apparently requires an understanding of how $\langle\psi|$, A, B, and $|\phi\rangle$ interact with each other. To avoid this problem, we define $P(i,j,k)$ in terms of a time-inhomogeneous Markov chain,

$$P(i,j,k) = P_\psi(i) P_A(j|i) P_B(k|j), \tag{28}$$

with each transition depending on only one of the components of $\langle\psi|AB|\phi\rangle$. Plugging this into (27) gives

$$b_{\max} = \max_{ijk}\left\{\frac{|\psi_i^* A_{ij} B_{jk}\phi_k|}{P_\psi(i) P_A(j|i) P_B(k|j)}\right\} \tag{29}$$

$$= \max_{ijk}\left\{\frac{|\psi_i^*|}{P_\psi(i)}\frac{|A_{ij}|}{P_A(j|i)}\frac{|B_{jk}|}{P_B(k|j)}|\phi_k|\right\} \tag{30}$$

$$\leqslant \max_i\left\{\frac{|\psi_i^*|}{P_\psi(i)}\right\}\max_{ij}\left\{\frac{|A_{ij}|}{P_A(j|i)}\right\} \times \max_{jk}\left\{\frac{|B_{jk}|}{P_B(k|j)}\right\}\max_k\{|\phi_k|\}. \tag{31}$$

The goal is then to find $P_\psi(i)$, $P_A(j|i)$, and $P_B(k|j)$ that minimize the terms of (31). Consider first the case where $\langle\psi|$ is a probability distribution, the matrices A and B are right-stochastic matrices,[5] and $|\phi\rangle$ has small entries (say, $\|\phi\|_\infty \leqslant 1$). We can set $P_\psi(i) = \psi_i$, $P_A(j|i) = A_{ij}$, and $P_B(k|j) = B_{jk}$, with the result that each factor in (31) is bounded by 1. If $|\phi\rangle$ is not a probability distribution, we can turn it into one by defining $P_\psi(i) = |\psi_i|/\|\psi\|_1$; similarly, if A is not a right-stochastic matrix, we can set $P_A(j|i) = |A_{ij}|/\sum_{j'}|A_{ij'}|$ (and likewise for B). Then (31) becomes

$$b_{\max} \leqslant \|\psi\|_1 \max_i\left\{\sum_{j'}|A_{ij'}|\right\}\max_j\left\{\sum_{k'}|B_{jk'}|\right\}\|\phi\|_\infty \tag{32}$$

$$= \|\psi\|_1\|\bar{A}\|_\infty\|\bar{B}\|_\infty\|\phi\|_\infty. \tag{33}$$

Here, as in the rest of the paper, we use the induced norm for operators: $\|A\|_p = \max_{\boldsymbol{u}}\|A\boldsymbol{u}\|_p/\|\boldsymbol{u}\|_p$ (we do not use the entrywise or Schatten norms). Under this notation, $\|M\|_2$ is the largest singular value of M, $\|M\|_1$ is the maximum absolute column sum, and $\|M\|_\infty$ is the maximum absolute row sum. By Corollary 1, the value of $\langle\psi|AB|\phi\rangle$ can be estimated by drawing

$$O\big(\log_2(\delta^{-1})\epsilon^{-2} b_{\max}^2\big) \leqslant O\big(\log_2(\delta^{-1})\epsilon^{-2}\|\psi\|_1^2\|\bar{A}\|_\infty^2\|\bar{B}\|_\infty^2\|\phi\|_\infty^2\big) \tag{34}$$

samples (i,j,k) from $P(i,j,k)$ and averaging the corresponding $V(i,j,k)/P(i,j,k)$.

D. General p,q

In the case of quantum circuits, it is the ℓ^2- norm that is relevant. Instead of $b_{\max} \leqslant \|\psi\|_1\|\bar{A}\|_\infty\|\bar{B}\|_\infty\|\phi\|_\infty$ from the previous example, we want $b_{\max} \leqslant \|\psi\|_2\|\bar{A}\|_2\|\bar{B}\|_2\|\phi\|_2$. For the sake of generality, we allow arbitrary p,q satisfying $1/p + 1/q = 1$. The goal is to find a probability distribution that yields $b_{\max} \leqslant \|\psi\|_p\|\bar{A}\|_q\|\bar{B}\|_q\|\phi\|_q$. As in Sec. III B, the way to proceed is by taking a convex combination of two probability distributions, $R(i,j,k) = P(i,j,k)/p + Q(i,j,k)$.

[4]The weighted inequality of arithmetic and geometric means is a generalization of the more familiar inequality $x/2 + y/2 \geqslant \sqrt{xy}$. If $1 \leqslant p \leqslant \infty$ and $1/p + 1/q = 1$, then $x/p + y/q \geqslant x^{1/p}y^{1/q}$.

[5]A right-stochastic matrix is a non-negative matrix with each row summing to 1; a left-stochastic matrix has each column summing to 1. We do not require stochastic matrices to be square.

Here $P(i,j,k)$ will be a time-inhomogeneous Markov chain proceeding in the $i \to j \to k$ direction and $Q(i,j,k)$ a different Markov chain proceeding in the $k \to j \to i$ direction. Again the inequality of arithmetic and geometric means plays a crucial role, giving

$$R(i,j,k) = P(i,j,k)/p + Q(i,j,k)/q \tag{35}$$
$$\geqslant P(i,j,k)^{1/p} Q(i,j,k)^{1/q} \tag{36}$$
$$= [P_\psi(i) P_A(j|i) P_B(k|j)]^{1/p} \times [Q_A(i|j) Q_B(j|k) Q_\phi(k)]^{1/q}. \tag{37}$$

With this we have

$$b_{\max} = \max_{ijk} \left\{ \frac{|\psi_i^* A_{ij} B_{jk} \phi_k|}{R(i,j,k)} \right\} \tag{38}$$
$$\leqslant \max_{ijk} \left\{ \frac{|\psi_i^* A_{ij} B_{jk} \phi_k|}{P(i,j,k)^{1/p} Q(i,j,k)^{1/q}} \right\} \tag{39}$$
$$= \max_{ijk} \left\{ \frac{|\psi_i^*|}{P_\psi(i)^{1/p}} \frac{|A_{ij}|}{P_A(j|i)^{1/p} Q_A(i|j)^{1/q}} \times \frac{|B_{jk}|}{P_B(k|j)^{1/p} Q_B(j|k)^{1/q}} \frac{|\phi_k|}{Q_\phi(k)^{1/q}} \right\} \tag{40}$$
$$\leqslant \max_i \left\{ \frac{|\psi_i^*|}{P_\psi(i)^{1/p}} \right\} \max_{ij} \left\{ \frac{|A_{ij}|}{P_A(j|i)^{1/p} Q_A(i|j)^{1/q}} \right\} \times \max_{jk} \left\{ \frac{|B_{jk}|}{P_B(k|j)^{1/p} Q_B(j|k)^{1/q}} \right\} \max_k \left\{ \frac{|\phi_k|}{Q_\phi(k)^{1/q}} \right\} \tag{41}$$
$$= b_\psi b_A b_B b_\phi, \tag{42}$$

where b_ψ, b_A, b_B, and b_ϕ label the four factors of (41). By Corollary 1, the number of samples needed in order to estimate $\langle\psi|AB|\phi\rangle$ is $O(\log_2(\delta^{-1})\epsilon^{-2} b_\psi^2 b_A^2 b_B^2 b_\phi^2)$. The quantities b_ψ, b_A, b_B, and b_ϕ are therefore identified as being the simulation cost due to each of the components of $\langle\psi|AB|\phi\rangle$. We show in Appendix A (Theorem 9) that for any choice of probability distribution $b_A \geqslant \|\bar{A}\|_q$ and that there are optimal probability distributions achieving $b_A = \|\bar{A}\|_q$ (and similarly for B, ψ, and ϕ). Using these gives

$$b_{\max} \leqslant \|\psi\|_p \|\bar{A}\|_q \|\bar{B}\|_q \|\phi\|_q. \tag{43}$$

Whether these optimal probability distributions can be efficiently sampled from is a matter that needs to be considered on a case-by-case basis; however, we show in Sec. V that this is indeed the case for a wide range of matrices, both unitary and Hermitian. Additionally, in terms of query complexity rather than time complexity these efficient sampling requirements can, for the most part, be ignored, as we discuss further in Sec. IV C.

E. Dyads and density operators

It is possible to further generalize to expressions of the form $\mathrm{Tr}\{AB\sigma\}$. The special case $\langle\psi|AB|\phi\rangle$ is obtained by setting $\sigma = |\phi\rangle\langle\psi|$. The above derivation is easily adapted by writing σ_{ki}, $P_\sigma(i)$, and $Q_\sigma(k)$ instead of $\phi_k \psi_i^*$, $P_\psi(i)$ and $Q_\phi(k)$. With these substitutions, (38)–(42) become

$$b_{\max} = \max_{ijk} \left\{ \frac{|A_{ij} B_{jk} \sigma_{ki}|}{R(i,j,k)} \right\} \tag{44}$$
$$\leqslant \max_{ijk} \left\{ \frac{|A_{ij}|}{P_A(j|i)^{1/p} Q_A(i|j)^{1/q}} \frac{|B_{jk}|}{P_B(k|j)^{1/p} Q_B(j|k)^{1/q}} \times \frac{|\sigma_{ki}|}{P_\sigma(i)^{1/p} Q_\sigma(k)^{1/q}} \right\} \tag{45}$$
$$\leqslant \max_{ij} \left\{ \frac{|A_{ij}|}{P_A(j|i)^{1/p} Q_A(i|j)^{1/q}} \right\} \times \max_{jk} \left\{ \frac{|B_{jk}|}{P_B(k|j)^{1/p} Q_B(j|k)^{1/q}} \right\} \times \max_{ki} \left\{ \frac{|\sigma_{ki}|}{P_\sigma(i)^{1/p} Q_\sigma(k)^{1/q}} \right\} \tag{46}$$
$$= b_A b_B b_\sigma. \tag{47}$$

The b_σ factor differs from the other two in that the probability distributions are not conditional. This stems from the fact that σ represents the starting point of the Markov chains. If $\sigma = |\phi\rangle\langle\psi|$ then taking the probability distributions $P_\sigma(i) = |\psi_i|^p / \|\psi\|_p$ and $Q_\sigma(k) = |\phi_k|^q / \|\phi\|_q$ gives $b_\sigma = \|\psi\|_p \|\phi\|_q$ as in (43). If $p = q = 2$ and if σ is a density operator (positive semidefinite and trace 1), then taking the probability distributions $P_\sigma(i) = Q_\sigma(i) = \sigma_{ii}$ gives $b_\sigma = 1$ due to the inequality $|\sigma_{ki}| \leqslant \sqrt{\sigma_{kk}\sigma_{ii}}$, which is satisfied by positive semidefinite matrices.

F. Interference-producing capacity

In Sec. II B we interpreted the lowest possible $b_{\max}$ value, obtained by using the holistic probability distribution (10), as being the interference of a quantum circuit. Although this probability distribution achieves the lowest $b_{\max}$, there is no clear way to draw samples efficiently and for this reason the Markov chain technique of this section was developed. The result was a strategy that depends only on properties of the individual operators rather than on the expression as a whole. The $b_{\max}$ value for this strategy is upper bounded by (47).

Consider now the minimum possible value of one of the factors in (47), for instance b_A. In Appendix A (Theorem 9) we show that the best possible choice of $P_A(j|i)$ and $Q_A(i|j)$ yields $b_A = \|\bar{A}\|_q$. In the case of quantum circuits the relevant norm is $p = q = 2$, so this becomes[6]

$$b_A = \|\bar{A}\|_2. \tag{48}$$

This can be interpreted in terms of interference: It is the largest possible contribution A can make to the interference $\mathcal{I}$ of Definition 1 . Specifically, since $\|\cdot\|_2$ gives the maximum singular value of its argument, we have

$$\mathcal{I}(A^{(1)}, \ldots, A^{(S)}, |\phi\rangle\langle\psi|) \leqslant \|\bar{A}^{(1)}\|_2 \cdots \|\bar{A}^{(S)}\|_2 \|\phi\|_2 \|\psi\|_2. \tag{49}$$

[6]We focus here on the case $p = 2$ of relevance to quantum circuits, although the entire section could easily be generalized to $p \neq 2$.

TABLE I. The $\mathcal{I}_{\max}$ value for various matrices. Operators with larger $\mathcal{I}_{\max}$ value are harder to simulate using our technique. Proofs for the nontrivial cases are presented in Appendix C .

Matrix	$\mathcal{I}_{\max}$
Fourier or Hadamard transform on n qubits	$2^{n/2}$
Arbitrary gate on n qudits	No more than $d^{n/2}$
Haar wavelet transform on n qubits	$\sqrt{1+n}$
k-sparse unitary	No more than $\sqrt{k}$
Grover reflection	$\mathcal{I}_{\max} \to 3$ as $n \to \infty$
Permutation in computational basis	1
Pauli matrices	1
Rank 1 projector	1

Furthermore, for any operator A we have

$$\max_{\|\psi\rangle_2=\|\phi\rangle_2=1} \mathcal{I}(A,|\phi\rangle\langle\psi|) = \|\bar{A}\|_2. \tag{50}$$

For this reason, we interpret $\|\bar{A}\|_2$ as being the interference producing capacity of A.[7]

Definition 2. The *interference-producing capacity* of a matrix A is

$$\mathcal{I}_{\max}(A) = \|\bar{A}\|_2. \tag{51}$$

This definition, like Definition 1, is basis dependent. Here the basis dependence arises from the entrywise absolute value. Unless otherwise specified, we work in the computational basis. In the next sections we show the product of the $\mathcal{I}_{\max}$ values for the operations and final measurement of a circuit to be a necessary resource for quantum speedup: If this quantity is low, then a circuit can be classically simulated. The same claim applies also for other bases and even for more exotic representations (as we show in Sec. VI A). The situation is not so much different from, for instance, Gottesman-Knill theorem which claims that stabilizer circuits may be efficiently simulated [4]. Although a circuit may at first not appear to be a stabilizer circuit, it may be so after a change of basis (i.e., after conjugating the initial state, all unitary operations, and all measurements by some unitary).

The $\mathcal{I}_{\max}$ values for various operators are listed in Table I. As shown informally in this section, and more formally in the next section, it is possible to efficiently simulate quantum circuits when the product of the $\mathcal{I}_{\max}$ values of all operators is not large. So, one may interpret a small $\mathcal{I}_{\max}$ value to mean that a unitary operator contributes only minimally to quantum speedup. On the high end of the table are the Fourier and Hadamard transforms, having the maximum possible value of $\mathcal{I}_{\max}$; these are difficult for us to simulate (at least in the computational basis). On the low end are the Pauli and the permutation matrices, having $\mathcal{I}_{\max} = 1$; these contribute nothing to quantum speedup (relative to our simulation scheme). Among unitaries, the only operators with $\mathcal{I}_{\max} = 1$ are permutations with phases, $U = \sum_j e^{i\theta_j}|\sigma(j)\rangle\langle j|$.

[7]Our measure of interference is different from, and seemingly unrelated to, the one defined in [16], which in the case of unitary matrices reduces to $N - \sum_{ij} |U_{ij}|^4$.

IV. EPS AND EHT OPERATORS

A. Definitions

We now present two definitions codifying the requirements operators must meet in order that products of the form $\mathrm{Tr}\{A^{(1)} \cdots A^{(S)}\sigma\}$ can be estimated using the techniques of the previous section. In the previous section, using a pair of Markov chains yielded a simulation strategy in which each component of $\mathrm{Tr}(AB\sigma)$ can be treated independently, with A, B, and σ contributing costs b_A, b_B, and b_σ to the total number of samples needed as per (47). Each sample requires drawing a random path according to the distribution $R(i,j,k)$ and then computing $V(i,j,k)/R(i,j,k)$. Drawing the random path requires considering only one operator at a time since $R(i,j,k)$ is defined in terms of Markov chains. Similarly, computing $V(i,j,k)/R(i,j,k)$ can be done considering one operator at a time since

$$\frac{V(i,j,k)}{R(i,j,k)} = \frac{A_{ij}B_{jk}\sigma_{ki}}{P(i,j,k)/p + Q(i,j,k)/q} \tag{52}$$

$$= \left\{ \frac{1}{p}\frac{P(i,j,k)}{A_{ij}B_{jk}\sigma_{ki}} + \frac{1}{q}\frac{Q(i,j,k)}{A_{ij}B_{jk}\sigma_{ki}} \right\}^{-1} \tag{53}$$

$$= \left\{ \frac{1}{p}\frac{P_A(j|i)}{A_{ij}}\frac{P_B(k|j)}{B_{jk}}\frac{P_\sigma(i)}{\sigma_{ki}} + \frac{1}{q}\frac{Q_A(i|j)}{A_{ij}}\frac{Q_B(j|k)}{B_{jk}}\frac{Q_\sigma(k)}{\sigma_{ki}} \right\}^{-1}. \tag{54}$$

Focusing on a single component, say A, conditions for efficient simulation can be identified (note that σ requires slightly different conditions, which we deal with later). First, the quantity b_A of (47) should be small in order that the number of samples required be small. Second, it must be possible to efficiently sample from the probability distributions $P_A(j|i)$ and $Q_A(i|j)$ and to compute the contributions due to A in (54), namely, $P_A(j|i)/A_{ij}$ and $Q_A(i|j)/A_{ij}$. We express these conditions as a definition. However, it will be useful to generalize by allowing an extra index k in the definition below (not related to the k that appears above). If k takes only a single value (say, $k = 0$) the definition below exactly encompasses the conditions outlined above. The extra freedom granted by k will allow, as we show shortly, treatment of sums, products, and exponentials of matrices (Theorem 4). In the case $p = 1$, $q = \infty$ it was the matrices resembling stochastic matrices that could be efficiently simulated. For this reason, for general p,q we give the name *efficient pseudostochastic* (EPS) to matrices that we can efficiently simulate.

Definition 3. EPS. Let $1 \leqslant p \leqslant \infty$, $1/p + 1/q = 1$, and $b < \infty$. An $M \times N$ matrix A is $\mathrm{EPS}_p(b,f)$ if there is a finite or countable set K, values $\alpha_{mnk} \in \mathbb{C}$, and conditional probability distributions $P(n,k|m)$ and $Q(m,k|n)$, with $m \in \{1,\ldots,M\}$, $n \in \{1,\ldots,N\}$, and $k \in K$, satisfying the following conditions:

(a) $\sum_{k\in K} \alpha_{mnk} = A_{mn}$;[8]

[8]We show in Appendix B (Lemma 4) that this series converges absolutely, so there is no ambiguity regarding the way that an infinite K is enumerated.

(b)

$$\max_{mnk}\left\{\frac{|\alpha_{mnk}|}{P(n,k|m)^{1/p}Q(m,k|n)^{1/q}}\right\}\leqslant b, \tag{55}$$

with the convention that $0/0=0$;

(c) given any m, it is possible in average time $O(f)$ on a classical computer to sample n,k from the probability distribution $P(n,k|m)$ and then compute $\alpha_{mnk}/P(n,k|m)$ and $\alpha_{mnk}/Q(m,k|n)$;

(d) given any n, it is possible in average time $O(f)$ on a classical computer to sample m,k from the probability distribution $Q(m,k|n)$ and then compute $\alpha_{mnk}/P(n,k|m)$ and $\alpha_{mnk}/Q(m,k|n)$.

This definition is related to interference-producing capacity in the following way. It is always possible to satisfy conditions (a) and (b) with $b=\|\bar{A}\|_q$, and it is impossible to do better. This is proved in Appendix A . So, for the case $p=q=2$ the optimal value of b is equal to the interference-producing capacity of A. Since b (multiplied for all operators in a circuit) determines how many samples will be required for our simulation technique, this connects interference-producing capacity to difficulty of simulation.

Although conditions (a) and (b) can always be satisfied with $b=\|\bar{A}\|_q$ for some α_{mnk}, $P(n,k|m)$, and $Q(m,k|n)$, it could be the case that these do not satisfy (c) and (d). In other words, it may be time consuming to sample from these probability distributions. An example would be a permutation matrix $A|x\rangle=|g(x)\rangle$. Such a matrix has $\|\bar{A}\|_q=1$, so it has no interference-producing capacity. Nevertheless, it would be difficult to simulate if the function g were difficult to calculate. In some sense (c) and (d) constitute a requirement that the matrix A be well understood from a computational perspective. In practice, (c) and (d) have not presented an obstacle for any of the operators that we have considered. If one is concerned with query complexity rather than time complexity, then (c) and (d) can mostly be ignored. This will be explored in Sec. IV C.

There is a subtlety in conditions (c) and (d) that deserves discussion. It is required that the operations be carried out in *average* time $O(f)$. It is allowed that $\alpha_{mnk}/P(n,k|m)$ and $\alpha_{mnk}/Q(m,k|n)$ be difficult to compute for some m,n,k triples as long as those occur rarely when sampling from $P(n,k|m)$ or $Q(m,k|n)$. In our implementation of exponentials of operators [Theorem 4(c)], the time required is proportional to k, and so is unbounded since $k\in\{0,1,\dots\}$; however, $P(n,k|m)$ and $Q(m,k|n)$ decay exponentially in k so the average time is small.

We now present a definition that embodies the conditions that σ must satisfy in order to yield an efficient simulation. Looking to (46) and (54), the difference between the factors relating to σ and those relating to A are that the latter involve conditional probability distributions. This stems from the fact that the Markov chains begin at σ and so have no index to condition upon. With this difference in mind, we provide a definition analogous to Definition 3 but with nonconditional probability distributions. Since the Markov chains begin and end at σ, we name the suitable matrices *efficient head/tail* (EHT) matrices.

Definition 4. EHT. Let $1\leqslant p\leqslant\infty$, $1/p+1/q=1$, and $b<\infty$. An $M\times N$ matrix σ is $\mathrm{EHT}_p(b,f)$ if there is a finite or countable set K, values $\alpha_{mnk}\in\mathbb{C}$, and probability distributions $P(n,k)$ and $Q(m,k)$ with $m\in\{1,\dots,M\}$, $n\in\{1,\dots,N\}$, and $k\in K$, satisfying the following conditions:

(a) $\sum_{k\in K}\alpha_{mnk}=\sigma_{mn}$;

(b)

$$\max_{mnk}\left\{\frac{|\alpha_{mnk}|}{P(n,k)^{1/p}Q(m,k)^{1/q}}\right\}\leqslant b, \tag{56}$$

with the convention that $0/0=0$;

(c) it is possible in average time $O(f)$ on a classical computer to sample n,k from the probability distribution $P(n,k)$ and then, given any $m\in\{1,\dots,M\}$, to compute $\alpha_{mnk}/P(n,k)$ and $\alpha_{mnk}/Q(m,k)$;

(d) it is possible in average time $O(f)$ on a classical computer to sample m,k from the probability distribution $Q(m,k)$ and then, given any $n\in\{1,\dots,N\}$, to compute $\alpha_{mnk}/P(n,k)$ and $\alpha_{mnk}/Q(m,k)$.

This definition does not relate to interference. For the case of quantum circuits we can assume σ to be a density operator. In Sec. IV D we show that for density operators it is always possible to achieve $b=1$ in the above definition as long as one can simulate measurements in the computational basis and compute individual matrix entries in average time $O(f)$.

The definition of EHT is more strict than that of EPS: Any EHT operator can be seen to also be EPS by using the probability distributions $P(n,k|m)=P(n,k)$ and $Q(m,k|n)=Q(m,k)$. Therefore, since it is not possible to have $b<\|\bar{A}\|_q$ for EPS operators, it is also not possible to have $b<\|\bar{\sigma}\|_q$ for EHT operators. As mentioned above, in the case of EPS it is always possible to satisfy conditions (a) and (b) with $b=\|\bar{A}\|_q$; however, since EHT is more strict, there are operators σ for which it is not possible to have $b=\|\bar{\sigma}\|_q$. Theorem 9(d) in Appendix A gives that $b=\|\bar{\sigma}\|_{\mathrm{Tr}}$ is possible when $p=q=2$, where $\|\cdot\|_{\mathrm{Tr}}$ is the trace norm (and a generalization is provided for $p\neq 2$).

In Sec. V we consider the case $p=q=2$, which is the norm relevant to quantum circuits, and give several examples of states that are $\mathrm{EHT}_2(b,f)$ and operators that are $\mathrm{EPS}_2(b,f)$, where b is small and f is polynomial in the number of qubits (or $\mathrm{polylog}_2$ in the dimension of the system). Expectation values of circuits built from such states and operators can be efficiently simulated. Specifically, we have the following theorem, the central theorem of this paper, whose proof is deferred until after Lemma 1.

Theorem 2. Efficient simulation. Let σ be $\mathrm{EHT}_p(b_\sigma,f_\sigma)$ and for $t\in\{1,\dots,S\}$ let $A^{(t)}$ be $\mathrm{EPS}_p(b_t,f_t)$. Then, with probability less than $\delta>0$ of exceeding the error bound, $\mathrm{Tr}\{A^{(1)}\cdots A^{(S)}\sigma\}$ can be estimated to within additive error $\epsilon>0$ in average time $O(\log_2(\delta^{-1})\epsilon^{-2}b^2f)$, where $b=b_\sigma\prod_t b_t$ and $f=f_\sigma+\sum_t f_t$.

B. Operations that preserve EPS/EHT properties

We now discuss mathematical operations that preserve the EPS and EHT properties. These include scaling, transpose, adjoint, multiplication, addition, and exponentiation (Theorems 3 and 4). The first three follow immediately from the definitions, so the following theorem is presented without proof.

Theorem 3. Let A be $\mathrm{EPS}_p(b,f)$ and σ be $\mathrm{EHT}_p(b,f)$. Let $s \in \mathbb{C}$ be a scalar. Then

(a) σ is $\mathrm{EPS}_p(b,f)$;
(b) sA is $\mathrm{EPS}_p(|s|b,f)$;
(c) $s\sigma$ is $\mathrm{EHT}_p(|s|b,f)$;
(d) $A^\top$ and $A^\dagger$ are $\mathrm{EPS}_p(b,f)$;
(e) $\sigma^\top$ and $\sigma^\dagger$ are $\mathrm{EHT}_p(b,f)$.

The presence of the k index in Definition 3 allows treatment of sums and products of operators. Consider, for instance, the product AB. The two factors of (45) relating to A and B can be combined to match the conditions of Definition 3 as follows. Begin by relabeling the indices of (45) from i,j,k to m,k,n and proceed as follows:

$$b_{\max} \leqslant \max_{mnk}\left\{\frac{|\sigma_{nm}|}{P_\sigma(m)^{1/p}Q_\sigma(n)^{1/q}}\frac{|A_{mk}|}{P_A(k|m)^{1/p}Q_A(m|k)^{1/q}}\frac{|B_{kn}|}{P_B(n|k)^{1/p}Q_B(k|n)^{1/q}}\right\} \tag{57}$$

$$\leqslant \max_{mn}\left\{\frac{|\sigma_{nm}|}{P_\sigma(m)^{1/p}Q_\sigma(n)^{1/q}}\right\}\max_{mnk}\left\{\frac{|A_{mk}|}{P_A(k|m)^{1/p}Q_A(m|k)^{1/q}}\frac{|B_{kn}|}{P_B(n|k)^{1/p}Q_B(k|n)^{1/q}}\right\} \tag{58}$$

$$\leqslant \max_{mn}\left\{\frac{|\sigma_{nm}|}{P_\sigma(m)^{1/p}Q_\sigma(n)^{1/q}}\right\}\max_{mnk}\left\{\frac{|A_{mk}B_{kn}|}{[P_A(k|m)P_B(n|k)]^{1/p}[Q_A(m|k)Q_B(k|n)]^{1/q}}\right\} \tag{59}$$

$$= b_\sigma b_{AB}. \tag{60}$$

Defining $P_{AB}(n,k|m) = P_A(k|m)P_B(n|k)$, $Q_{AB}(m,k|n) = Q_B(k|n)Q_A(m|k)$, and $\alpha_{mnk} = A_{mk}B_{kn}$, the b_{AB} factor reduces to

$$b_{AB} = \max_{mnk}\left\{\frac{|\alpha_{mnk}|}{P_{AB}(n,k|m)^{1/p}Q_{AB}(m,k|n)^{1/q}}\right\}. \tag{61}$$

This resembles the factors involving A or B that appear in (46) but with the addition of an extra index k appearing in both the numerator and in the probability distributions. Allowing such an extra index enables treatment of AB in the same manner as the individual factors A and B. This is formalized by Theorem 4(b) below, which states that the product of EPS matrices is EPS. In the general case, this procedure is slightly complicated by the fact that A and B may in turn have their own extra indices k' and k'', which must be inherited by the product AB.

Sums are handled in a similar way. An expression such as $\mathrm{Tr}[(A+B)\sigma]$ is estimated by using A for a fraction of the samples and B for the remainder. This works since $\mathrm{Tr}[(A+B)\sigma]$ is twice the average of $\mathrm{Tr}(A\sigma)$ and $\mathrm{Tr}(B\sigma)$. The k index is used to randomly choose between A or B for each sample. Exponentials are treated by applying these sum and product rules to $e^A = \sum_{j=0}^\infty A^j/j!$.

Theorem 4. Operations on EPS. Let A be a matrix that is $\mathrm{EPS}_p(b_A,f_A)$ and let B be a matrix that is $\mathrm{EPS}_p(b_B,f_B)$. Then, assuming in each case that A and B have a compatible number of rows and columns, the following hold:

(a) $A+B$ is $\mathrm{EPS}_p(b_A+b_B, \max\{f_A,f_B\})$;
(b) AB is $\mathrm{EPS}_p(b_Ab_B, f_A+f_B)$;
(c) e^A is $\mathrm{EPS}_p(e^b, bf)$.

Proof. The proofs are in Appendix B . Rule (a) is a special case of Theorem 13, which treats finite or infinite linear combinations. ■

Since the value b in Definition 3 (with $p=q=2$) is lower bounded by interference producing capacity $\mathcal{I}_{\max}$, Theorem 4 has the following interpretation. By (a), $\mathcal{I}_{\max}$ is convex. By (b), it is submultiplicative. By (c), the interference-producing capacity of a Hamiltonian evolution e^{iHt} is at most exponential in $t\mathcal{I}_{\max}(H)$.

We now prove Theorem 2, regarding estimation of $\mathrm{Tr}\{A^{(1)}\cdots A^{(S)}\sigma\}$. While this can be proved directly using Markov chains, as was done in Sec. III, this would be notationally tedious. It is much easier to first repeatedly apply the product rule, Theorem 4(b), to show that $A = A^{(1)}\cdots A^{(S)}$ is $\mathrm{EPS}_p(\prod_t b_t, \sum_t f_t)$. It then suffices to show that $\mathrm{Tr}(A\sigma)$ can be estimated. Although this may seem like a slightly nonconstructive proof, this strategy arose due to object-oriented techniques (C++) used during actual implementation of the algorithm. Unrolling the proof of the product theorem, as well as the proof of the theorem that follows, gives an argument very similar to that presented in Sec. III.

Lemma 1. Let σ be an $N\times M$ matrix that is $\mathrm{EHT}_p(b_\sigma,f_\sigma)$. Let A be an $M\times N$ matrix that is $\mathrm{EPS}_p(b_A,f_A)$. It is possible to estimate $\mathrm{Tr}(A\sigma)$ to within additive error $\epsilon>0$, with probability less than $\delta>0$ of exceeding the error bound, in average time $O(\log_2(\delta^{-1})\epsilon^{-2}b_\sigma^2b_A^2(f_\sigma+f_A))$.

Proof. The proof is in Appendix B and follows along the lines of the techniques developed in Sec. III. ■

Proof of Theorem 2 . By iterated application of Theorem 4(b), $A = A^{(1)}\cdots A^{(S)}$ is $\mathrm{EPS}_p(\prod_t b_t, \sum_t f_t)$. By Lemma 1 the value of $\mathrm{Tr}(A\sigma)$ can be estimated in time $O(\log_2(\delta^{-1})\epsilon^{-2}b^2f)$, where $b = b_\sigma\prod_t b_t$ and $f = f_\sigma + \sum_t f_t$. ■

C. Query complexity

The simulation algorithm of this paper involves sampling a number of paths via Markov chains, each path evaluation in turn requiring certain operations to be performed. Definitions 3 and 4 each consist of two pairs of conditions, (a) and (b), relating to the number of paths that need to be evaluated (quantified by b), and (c) and (d), concerning tasks that need to be performed for each path (quantified by f). In Appendix A we show (Theorem 9) that there are always α_{mnk}, $P(n,k|m)$, and $Q(m,k|n)$ satisfying conditions (a) and (b) with $b = \|\bar{A}\|_q$ (and, in fact, smaller b is not possible). However, these probability distributions may not satisfy (c) and (d), which require that the distributions can be sampled from efficiently. It is difficult to make any general statement

regarding satisfaction of (c) and (d), since time complexity of computation is, in general, a difficult problem; satisfaction of these two conditions needs to be considered on a case-by-case basis. However, when considering query complexity rather than time complexity, (c) and (d) can for the most part be ignored as we now explain. Note that communication complexity (discussed in Sec. VI B) offers another context in which (c) and (d) can be ignored, since there, too, computation time is free.

Consider the situation where an algorithm is required to answer some question about an oracle, which is to be thought of as a black box provided to the algorithm (Grover's algorithm is a prominent example). For a classical (i.e., nonquantum) algorithm the oracle can be any function between two finite sets, say $g : X \to Y$. It is convenient to consider sets of integers, $X = \{0,1,\ldots,|X|-1\}$ and $Y = \{0,1,\ldots,|Y|-1\}$. The algorithm can query the oracle by providing it a value $x \in X$, and the oracle responds with $g(x)$. This is the only allowed way to gain information about g. The query complexity of the algorithm is defined to be the number of times it queries the oracle. In particular, the query complexity is not affected by the amount of time spent performing computations between queries; computation, even lengthy computation, is not charged for.

Quantum circuits are provided access to an oracle in the form of a unitary operator,[9]

$$\mathcal{O}_g = \sum_{x\in X, y\in Y} |x\rangle\langle x| \otimes |y + g(x)\rangle\langle y|, \tag{62}$$

where $|x\rangle \otimes |y\rangle \in \mathbb{C}^{|X|} \otimes \mathbb{C}^{|Y|}$ are computational basis vectors and where the addition $y + g(x)$ is modulo $|Y|$. The query complexity of a quantum circuit is defined to be the number of times $\mathcal{O}_g$ appears in the circuit. For example, Grover's algorithm has query complexity $O(\sqrt{N})$.

Computational complexity classes can be analyzed by comparing how two classes perform when given access to equivalent oracles. For example, oracles have been constructed relative to which quantum computers perform exponentially more efficiently than classical computers (e.g., Simon's problem [20]), whereas proving that quantum computers are faster than classical computers in the absence of an oracle is an extremely difficult open problem.

Considering query complexity rather than time complexity simplifies the analysis of the present paper. Suppose we wish to simulate a quantum circuit containing at least one instance of an oracle $\mathcal{O}_g$ (e.g., Grover's algorithm) on a classical computer that also has oracle access to g. Simulation of the quantum circuit on the classical computer will require making queries to g and we can ask how many queries are needed, ignoring the amount of computational time used. We do this by modifying conditions (c) and (d) of Definitions 3 and 4 to require that the sampling and computation tasks be completed using $O(f)$ queries to g, rather than requiring $O(f)$ time (time now being a resource that is not charged for). We refer to such modified definitions by invoking the phrase "in terms of query complexity."

We now show that, in terms of query complexity, $\mathcal{O}_g$ is $\mathrm{EPS}_p(1,1)$. Since this unitary operates on two subsystems, $\mathbb{C}^{|X|} \otimes \mathbb{C}^{|Y|}$, the indices m and n in Definition 3 are tuple valued. We write $m = (x,y) \in X \times Y$ and $n = (x',y') \in X \times Y$. Take K to be the singleton set $\{0\}$ and define

$$\alpha_{(x,y)(x',y')k} := P((x',y'),k|(x,y)) \tag{63}$$

$$:= Q((x,y),k|(x',y')) \tag{64}$$

$$:= \langle xy|\mathcal{O}_g|x'y'\rangle \tag{65}$$

$$= \delta(x,x')\delta(y + g(x),y'), \tag{66}$$

where δ is the Kronecker δ. It is easy to see that these satisfy conditions (a) and (b) of Definition 3 with $b = 1$. Sampling from these probability distributions and computing the values of any of these quantities can be done with a single query of g (note that the conditional probability distributions are deterministic); therefore, conditions (c) and (d) are satisfied with $f = 1$.

On the other hand, for matrices that are not defined in terms of the oracle g, such as the $I - 2|+\rangle\langle+|$ reflection operators in Grover's algorithm, the operations required by conditions (c) and (d) can be carried out using zero queries. Therefore, conditions (c) and (d) can be completely ignored, and we can take $f = 0$. We are then free to focus on determining the probability distributions, giving the smallest possible value of b in conditions (a) and (b) without regard to whether these can be efficiently sampled from (since we are charging for queries only and time is free). It is desirable to make b as small as possible, since this determines the number of paths that need to be sampled. The number of paths sampled matters, because each will require evaluating the entire Markov chain, which involves every operator. At least one of these operators involves the oracle, so at least one query needs to be made for each path that is sampled. The total number of oracle queries will be the number of paths sampled times the number of queries per path. In Appendix A we show (Theorem 9) the existence of probability distributions which satisfy conditions (a) and (b) with $b = \|\bar{A}\|_q$. So, in terms of query complexity, any matrix A not defined in terms of an oracle is $\mathrm{EPS}_p(\|\bar{A}\|_q,0)$. In the case $p = q = 2$ of relevance to quantum circuits, we have $\|\bar{A}\|_2 = \mathcal{I}_{\max}(A)$, the interference-producing capacity of A. Theorem 9 also shows that any σ not defined in terms of an oracle is $\mathrm{EHT}_2(\|\sigma\|_{\mathrm{Tr}},0)$, where $\|\cdot\|_{\mathrm{Tr}}$ is the trace norm (a generalization is provided for $p \neq 2$).

D. Sufficient conditions for EPS/EHT

We now present theorems that can be used to show that specific operators are EPS or EHT. As stated above, if one is only interested in query complexity, then any matrix A not depending on an oracle is guaranteed to be $\mathrm{EPS}_p(\|\bar{A}\|_q,0)$. However, in terms of time complexity it is possible that the probability distributions that achieve $b = \|\bar{A}\|_q$ cannot be sampled from efficiently (giving large f). For this reason it is worthwhile to introduce probability distributions that are more

[9] Sometimes an alternate definition, $\mathcal{O}'_g = \sum_{x\in X, y\in Y} e^{2\pi i g(x)y/|Y|}|x\rangle\langle x| \otimes |y\rangle\langle y|$, is used. All claims apply to this definition as well, requiring only a modification of (63)–(66).

likely to be efficiently sampled and which in some cases still achieve a small b. In the theorem below each row and column of A is treated as a probability distribution, correcting for phases and normalization. This works well when the absolute row and column sums of A are small.

Theorem 5. Let $1 \leqslant p \leqslant \infty$ and $1/p + 1/q = 1$. Let A be an $M \times N$ matrix. Define the probability distributions

$$P(n|m) = \frac{|A_{mn}|}{\sum_{n'} |A_{mn'}|}, \quad Q(m|n) = \frac{|A_{mn}|}{\sum_{m'} |A_{m'n}|}. \quad (67)$$

Suppose that it is possible in average time $O(f)$ on a classical computer to perform the following operations:

(a) Given m, sample n from the probability distribution $P(n|m)$;

(b) given n, sample m from the probability distribution $Q(m|n)$;

(c) given m,n, compute A_{mn}, $\sum_{n'} |A_{mn'}|$, and $\sum_{m'} |A_{m'n}|$.

Then A is $\mathrm{EPS}_p(b, f)$ with $b = \|A\|_\infty^{1/p} \|A\|_1^{1/q}$. Note that b is the weighted geometric mean of the maximum row and column sums of A.

Proof. This follows directly from plugging the probability distributions (67) into Definition 3, with $K = \{0\}$ (i.e., not making use of the index k). Note that $\|A\|_\infty$ is the maximum absolute row sum and $\|A\|_1$ is the maximum absolute column sum of A. ■

Finally, we present theorems that cover the two most important examples of EHT operators: dyads and density operators.

Theorem 6. Dyads are EHT. Let $|\phi\rangle$ and $\langle\psi|$ be vectors such that the probability distributions $P(n) = |\psi_n|^p / \|\psi\|_p^p$ and $Q(m) = |\phi_m|^q / \|\phi\|_q^q$ can be sampled from and the corresponding ψ_n and ϕ_m can be computed, in average time $O(f)$. Then the dyad $|\phi\rangle\langle\psi|$ is $\mathrm{EHT}_p(\|\psi\|_p \|\phi\|_q, f)$.

Proof. This can be seen immediately by plugging the given probability distributions into Definition 4, with $K = \{0\}$ (i.e., without making use of index k). This is the best possible value of b, which can be seen by applying Theorem 9(a) and using $\|(|\phi\rangle\langle\psi|)\|_q = \|\psi\|_p \|\phi\|_q$. ■

Corollary 2. Estimate matrix entries. Let A be $\mathrm{EPS}_p(b, f)$. Then, given any indices i, j, the value of the matrix entry A_{ij} can be estimated to within additive error $\epsilon > 0$, with probability less than $\delta > 0$ of exceeding the error bound, in average time $O(\log_2(\delta^{-1})\epsilon^{-2} b^2 f)$.

Proof. By Theorem 6 the dyad of computational basis vectors $|j\rangle\langle i|$ is $\mathrm{EHT}_p(1, \log_2(N))$. Note: $f \geqslant \log_2(N)$ in all cases (unless one is dealing with query complexity) since it takes $O(\log_2(N))$ time to even write down the indices i and j, which are $\log_2(N)$ bits long. By Lemma 1, $A_{ij} = \mathrm{Tr}(A|j\rangle\langle i|)$ can be estimated in time $O(\log_2(\delta^{-1})\epsilon^{-2} b^2 [f + \log_2(N)]) = O(\log_2(\delta^{-1})\epsilon^{-2} b^2 f)$. ■

Theorem 7. Density operators are EHT. Let σ be a density operator. Suppose that it is possible to sample from the probability distribution $P(n) = \sigma_{nn}$ in average time $O(f)$ and, given i, j, to compute σ_{ij} in average time $O(f)$. Then σ is $\mathrm{EHT}_2(1, f)$.

Proof. This follows from plugging the probability distributions $P(n) = \sigma_{nn}$ and $Q(m) = \sigma_{mm}$ into Definition 4 and using the inequality $|\sigma_{mn}| \leqslant \sqrt{\sigma_{mm}\sigma_{nn}}$, which is satisfied by positive semidefinite matrices. ■

V. SIMULATION OF QUANTUM CIRCUITS

A. Efficiently simulated states and operators

In this section we take up the case $p = q = 2$, which is relevant to quantum circuits, and list several examples of $\mathrm{EHT}_2(b, f)$ states and $\mathrm{EPS}_2(b, f)$ operators where b is small and $f \leqslant \mathrm{polylog}_2(N)$ where N is the dimension of the system (i.e., $N = 2^n$ where n is the number of qubits). By Theorem 2, circuits made of such states and operators can be efficiently simulated. For example, the circuit depicted in Fig. 1 can be simulated in $\mathrm{polylog}_2(N)$ time. After providing several examples of such states and operators, we discuss a few circuits that cannot be efficiently simulated using our technique.

The initial states we are able to efficiently simulate include the *computationally tractable* (CT) states of [15]. We reproduce the definition here.[10]

Definition 5. A normalized state $|\psi\rangle$ of dimension N is called CT if the following conditions hold:

(a) It is possible to sample in $\mathrm{polylog}_2(N)$ time with classical means from the probability distribution $P(i) = |\psi_i|^2$;

(b) upon input of any $i \in \{0, \ldots, N-1\}$, the coefficient ψ_i can be computed in $\mathrm{polylog}_2(N)$ time on a classical computer.

It follows immediately from Theorem 6 that if $|\psi\rangle$ is a CT state then $\rho = |\psi\rangle\langle\psi|$ is $\mathrm{EHT}_2(1, \mathrm{polylog}_2(N))$. For convenience we present here a brief list of examples of such states from [15] and refer the reader to their paper for details:

(i) product states of qubits (we allow also qudits);

(ii) stabilizer states;

(iii) states of the form $|\psi\rangle = \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} e^{i\theta(x)} |x\rangle$, where $e^{i\theta(x)}$ for a given x can be computed in $\mathrm{polylog}_2(N)$ time;

(iv) matrix product states of polynomial bond dimension;

(v) states obtained by applying a polynomial sized nearest-neighbor matchgate circuit to a computational basis state;

(vi) states obtained by applying the quantum Fourier transform to a product state;

(vii) the output of quantum circuits with logarithmically scaling tree-width acting on product input states.

We present a list of examples of $\mathrm{EPS}_2(b, f)$ operators with b small and $f \leqslant \mathrm{polylog}_2(N)$. All proofs are in Appendix C.

(i) If A is $\mathrm{EPS}_p(b, f)$, then $I \otimes \cdots \otimes I \otimes A \otimes I \otimes \cdots \otimes I$ is $\mathrm{EPS}_p(b, \max\{f, \log_2^2(N)\})$ (Corollary 5). In other words, EPS operations on subsystems are EPS. The $\log_2^2(N)$ is due to the amount of time needed to convert indices of $I \otimes \cdots \otimes I \otimes A \otimes I \otimes \cdots \otimes I$ to indices of A.

(ii) Any operator A on a constant number of qubits or qudits is $\mathrm{EPS}_2(\mathcal{I}_{\max}(A), 1)$, where $\mathcal{I}_{\max}(A) = \|\bar{A}\|_2$ is the interference-producing capacity of A. In other words, the simulation cost due to such an operator is equal to the fourth power of its interference-producing capacity [because of the b_t^4 term in (69)].

[10] Their definition referred to qubits. We generalize slightly to the abstract case where the decomposition into subsystems is not defined; only the total dimension of the space matters.

(iii) If A is an $M \times M$ matrix with maximum singular value bounded by 1 [e.g., a unitary, projector, or positive operator-valued measure (POVM) element], then $\mathcal{I}_{\max}(A) \leqslant \sqrt{M}$. This inequality is saturated when A is a unitary with rows forming a basis mutually unbiased to the computational basis (e.g., a Hadamard or Fourier transform).

(iv) In terms of query complexity rather than time complexity, any operator A not depending on an oracle is $\mathrm{EPS}_2(\mathcal{I}_{\max}(A),0)$ by Theorem 9. The oracles themselves are $\mathrm{EPS}_2(1,1)$.

(v) Efficiently computable sparse matrices as defined in [15] are $\mathrm{EPS}_p(\mathrm{polylog}_2(N),\mathrm{polylog}_2(N))$ (Theorem 14). These include the following.

(a) Permutation matrices are $\mathrm{EPS}_p(1,f)$ as long as the permutation and its inverse can be computed in time $O(f)$.

(b) Diagonal unitary matrices are $\mathrm{EPS}_p(1,f)$ as long as the phases can be computed in time $O(f)$.

(c) Pauli matrices are $\mathrm{EPS}_p(1,1)$.

(vi) Grover reflections $I - 2(|+\rangle\langle+|)^{\otimes n}$ are $\mathrm{EPS}_2(3,n)$ (Theorem 16).

(vii) The Haar wavelet transform on n qubits (Definition 10) is $\mathrm{EPS}_2(\sqrt{n+1},n)$ (Theorem 17).

(viii) One-dimensional projectors onto CT states are $\mathrm{EPS}_2(1,\mathrm{polylog}_2(N))$ since CT dyads are $\mathrm{EHT}_2(1,\mathrm{polylog}_2(N))$ and EHT operators are EPS (Theorem 3).

(ix) Rank r projectors onto spaces defined by CT states are $\mathrm{EPS}_2(r,\mathrm{polylog}_2(N))$ [by applying the sum rule Theorem 4(a) to the previous item].

(x) Block diagonal matrices where each block is $\mathrm{EPS}_p(b,f)$, and in which matrix indices can be converted to and from block indices in time $O(f)$, are $\mathrm{EPS}_p(b,f)$.

(xi) As a special case of block diagonal matrices, projectors of the form $\sum_x |x\rangle\langle x| \otimes |\phi_x\rangle\langle\phi_x|$, where the $|x\rangle$'s are computational basis states and each $|\phi_x\rangle$ is a CT state, are $\mathrm{EPS}_2(1,\mathrm{polylog}_2(N))$. For example, given an even number of qubits, measure half of the qubits in the computational basis to get x, measure the other half in the Fourier basis to get y, return true if $y = g(x)$ for some function g computable in $\mathrm{polylog}_2(N)$ time (Corollary 4). In this example, $|\phi_x\rangle = F|g(x)\rangle$. The measurement depicted in Fig. 1 is of this form.

B. Simulation techniques

As a matter of convenience, we present a theorem that is essentially a direct corollary of Theorem 2, but written in the language of quantum circuits.

Theorem 8. Consider a quantum circuit using states of dimension N [i.e., $\log_2(N)$ qubits or $\log_d(N)$ qudits]. Let $|\psi\rangle$ be a CT state. For $t \in \{1,\ldots,T\}$ let $U^{(t)}$ be an $\mathrm{EPS}_2(b_t,\mathrm{polylog}_2(N))$ unitary and let M be an $\mathrm{EPS}_2(b_M,\mathrm{polylog}_2(N))$ Hermitian observable. It is possible, with probability less than $\delta > 0$ of exceeding the error bound, to estimate

$$\langle\psi|U^{(1)\dagger}\cdots U^{(T)\dagger} M U^{(T)}\cdots U^{(1)}|\psi\rangle \tag{68}$$

to within additive error $\epsilon > 0$ in average time

$$O\left(T\log_2(\delta^{-1})\epsilon^{-2}\mathrm{polylog}_2(N) b_M^2 \prod_{t=1}^{T} b_t^4\right). \tag{69}$$

In particular, if b_M, $\prod_t b_t$, and T are $\mathrm{polylog}_2(N)$, and if δ and ϵ are constant, then the simulation time is $\mathrm{polylog}_2(N)$ on average.

Note that in (69) each unitary $U^{(t)}$ incurs a cost of b_t^4 rather than b_t^2 since it appears twice in (68). If M is a rank 1 projector onto a CT state, $M = |\phi\rangle\langle\phi|$, then it is much more efficient to compute (68) as the absolute square of

$$\mathrm{Tr}\{|\psi\rangle\langle\phi|U^{(T)}\cdots U^{(1)}\}. \tag{70}$$

Since $|\psi\rangle\langle\phi|$ is $\mathrm{EHT}_2(1,\mathrm{polylog}_2(N))$, and since each unitary only occurs once, Theorem 2 gives that this expression can be estimated in average time,

$$O\left(T\log_2(\delta^{-1})\epsilon^{-2}\mathrm{polylog}_2(N)\prod_{t=1}^{T} b_t^2\right), \tag{71}$$

which is much better than (69). If M is a low-rank projector, the same trick can be used by decomposing M as the sum of rank 1 projectors and computing each resulting term individually. The complexity of such a technique will scale proportional to the rank of M.

Theorem 8 is just an application of Theorem 2 with $p = q = 2$. One may wonder whether other values of p,q would lead to a lower simulation cost. Ignore for the moment the efficient sampling conditions (c) and (d) of Definitions 3 and 4. When estimating (68), the optimal probability distributions give (by Theorem 9)

$$b := b_\psi b_{U^{(1)}}\cdots b_{U^{(T)}} b_M b_{U^{(T)}}\cdots b_{U^{(1)}} b_\psi \tag{72}$$

$$= \|\psi\|_p \|\bar{U}^{(1)\dagger}\|_q \cdots \|\bar{U}^{(T)\dagger}\|_q \|\bar{M}\|_q \|\bar{U}^{(T)}\|_q \cdots \|\bar{U}^{(1)}\|_q \|\psi\|_q. \tag{73}$$

This achieves its minimum value at $p = q = 2$, since

$$b = \|\psi\|_p\|\psi\|_q\|\bar{U}^{(1)}\|_p\|\bar{U}^{(1)}\|_q\cdots\|\bar{U}^{(T)}\|_p\|\bar{U}^{(T)}\|_q(\|\bar{M}\|_p\|\bar{M}\|_q)^{1/2} \quad (\text{using} \quad \|\mathrm{A}^\dagger\|_\mathrm{q} = \|\mathrm{A}\|_\mathrm{p}) \tag{74}$$

$$\geqslant \langle\psi|\psi\rangle\|\bar{U}^{(1)}\|_p\|\bar{U}^{(1)}\|_q\cdots\|\bar{U}^{(T)}\|_p\|\bar{U}^{(T)}\|_q(\|\bar{M}\|_p\|\bar{M}\|_q)^{1/2} \quad (\text{Hülder's inequality}) \tag{75}$$

$$\geqslant \langle\psi|\psi\rangle\,\|\bar{U}^{(1)}\|_2^2\cdots\|\bar{U}^{(T)}\|_2^2\|\bar{M}\|_2 \quad (\text{Riesz-Thorin theorem}) \tag{76}$$

$$= \|\psi\|_2\|\bar{U}^{(1)\dagger}\|_2\cdots\|\bar{U}^{(T)\dagger}\|_2\|\bar{M}\|_2\|\bar{U}^{(T)}\|_2\cdots\|\bar{U}^{(1)}\|_2\|\psi\|_2. \tag{77}$$

On the other hand, when estimating an expression of the form (70), each unitary is no longer repeated twice and Riesz-Thorin cannot be applied. In this case the minimum value of b does not necessarily occur at $p = 2$.

Certain algorithms, such as Shor's algorithm, consist of a quantum circuit terminating in a many-outcome measurement (e.g., measurement in the computational basis of several different qubits) which is then postprocessed by a classical computer to produce a final result. This does not immediately fit into our scheme of estimating expectation values. However, in the case where the final result is a two-outcome yes/no answer (e.g., "does N have a prime factor in the range $[a,b]$"), the final measurement and classical postprocessing can be combined into a single collective projector or POVM element as follows. Suppose the final state is measured using a POVM $\{F_i\}$. A classical postprocessing step then inspects the measurement outcome i and returns "yes" or "no." Denote by R the set of measurement outcomes that will result in "yes." The classical postprocessing can be absorbed into the measurement, resulting in the POVM element $F' = \sum_{i\in R} F_i$. The expectation value of F' gives the probability that a measurement of $\{F_i\}$ would yield "yes" after postprocessing.

In some cases F' may be efficiently simulated, a (somewhat contrived) example being the final stage of the circuit of Fig. 1. Note that this example involves a Fourier transform, which by itself cannot be efficiently simulated by our technique since it has large interference-producing capacity. However, when the Fourier transform is followed by the particular classical postprocessing depicted in Fig. 1, the resulting composite operator *can* be efficiently simulated (Corollary 4). Shor's algorithm also has a Fourier transform followed by classical postprocessing; however, in that case the composite operator (Fourier transform followed by postprocessing) has large interference-producing capacity and so *cannot* be efficiently simulated (by our algorithm).

C. Circuits that our technique cannot efficiently simulate

Many examples of efficiently simulatable circuits can be constructed, but it is probably more enlightening to instead discuss examples of circuits that cannot be efficiently simulated using our technique. Since the efficiency of our technique depends upon choice of basis and on choice of representation (see Sec. VI A), a circuit which our technique cannot simulate efficiently in one basis may be efficiently simulatable in another basis. In this section we choose to focus only on the computational basis. That being said, most of the examples in this section have been proved (relative to an oracle) to have no efficient classical solution.

We cannot efficiently simulate Shor's algorithm. The reason for this is that the Fourier transform has high interference-producing capacity: The Fourier transform F on n qubits has $\mathcal{I}_{\max}(F) = 2^{n/2}$. Replacing the Fourier transform by the Haar wavelet transform (Fig. 2) yields a circuit that can be efficiently simulated, since the Haar transform has low interference-producing capacity, $\mathcal{I}_{\max}(G_n) = \sqrt{n+1}$. Note that this circuit no longer factors numbers (and probably does nothing at all useful). The Fourier and Haar transforms play similar roles in classical signal processing, with the latter providing spatially localized rather than global information for the high-frequency components. The fact that replacing the Fourier transform enables efficient classical simulation points to the Fourier transform as being the source of the quantum speedup in Shor's algorithm (for a contrasting point of view, see [21,22]).

Deutsch-Jozsa provides an oracle relative to which deterministic quantum computation is more powerful than deterministic classical computation. Our algorithm can efficiently simulate the Deutsch-Jozsa algorithm, but not deterministically.[11] The Deutsch-Jozsa algorithm consists of an initial CT state $|+\rangle^{\otimes n} \otimes |-\rangle$, acted upon by an oracle $\sum_{xy} |x\rangle\langle x| \otimes |y + g(x)\rangle\langle y|$, followed by a rank 1 projective measurement onto the state $|+\rangle^{\otimes n} \otimes |-\rangle$. The initial state is $\mathrm{EHT}_2(1,n)$ and the operators are $\mathrm{EPS}_2(1,n)$, so we can efficiently simulate this algorithm. However, the simulation will always have a small chance of error due to the δ in Theorem 8.

Our simulation algorithm performs very poorly when applied to Grover's algorithm. Each iteration of Grover's algorithm consists of an oracle query followed by a Grover reflection. These operations have low interference-producing capacity: 1 for the oracle and just under 3 for the Grover reflection. However, our algorithm is exponentially slow in the circuit length, due to the $\prod_t b_t^4$ factor in (69). Since the Grover reflection is used $\Theta(\sqrt{N})$ times, the simulation would run in time $\exp[\Theta(\sqrt{N})]$. Even though each iteration of Grover's algorithm produces small interference, the total interference of the whole circuit, by Definition 1, is $\exp[\Theta(\sqrt{N})]$.

In [23] a quantum random walk is presented that provides an exponential speedup over any possible classical algorithm for the graph traversal problem. The walk is carried out by evolving the initial state with a Hamiltonian that is defined in terms of an oracle. We cannot efficiently simulate this algorithm for the same reason that we cannot efficiently simulate Grover: The runtime of the quantum algorithm increases with the problem size, and our simulation must pay an exponentially large penalty for this due to the $\prod_t b_t^4$ factor in (69). On the other hand, short-time and low-energy Hamiltonian evolutions can be efficiently simulated by our technique. In particular, Theorem 4(c) gives that if H is $\mathrm{EPS}_p(b,f)$, then e^{iHt} is $\mathrm{EPS}_p(e^{bt}, btf)$. In terms of query complexity the Hamiltonian in the algorithm of [23] is $\mathrm{EPS}_2(O(1),1)$, so we could feasibly simulate e^{iHt} for small t. However, their algorithm has $t = \Theta(n^4)$, so our simulation would have query complexity $e^{\Theta(n^4)}$, making it unfeasibly slow.

VI. APPLICATIONS AND DISCUSSION

A. Wigner representation

An $N \times N$ matrix can also be viewed as an N^2-dimensional vector, so we can write, for instance, $\langle \boldsymbol{M}|\boldsymbol{\rho}\rangle$ in place of $\mathrm{Tr}\{M\rho\}$. Superoperators become $N^2 \times N^2$ matrices in this representation, and we can write $\langle \boldsymbol{M}|\boldsymbol{V}\boldsymbol{U}|\boldsymbol{\rho}\rangle = \mathrm{Tr}\{MVU\rho U^\dagger V^\dagger\}$. Simulating a quantum circuit using this representation offers an alternative to the customary representation that was the focus of Sec. V.

Any basis can be used (even ones that are not orthonormal), although some choices of basis may yield more efficient

[11] This was discussed in [15], which our paper extends. However, we mention it here for completeness.

simulation. One notable choice is given by the discrete Wigner representation, which is only defined for qudits of odd dimension. We do not describe the details here but refer the reader to [8,10], in which it is shown that in the discrete Wigner representation stabilizer states become probability distributions and Clifford operations become permutation matrices.

It was shown independently in [9,10] that when operations in the Wigner representation are given by non-negative matrices, such matrices are stochastic and therefore can be efficiently simulated. Our algorithm, taking $p = \infty$ and $q = 1$, extends this result by also allowing states and operations in which the Wigner representation contains a small quantity of negative values, although ours is weaker in that it only computes expectation values rather than allowing sampling of a many-outcome measurement. With $q = 1$ rather than $q = 2$, the difficulty of simulating an operation is given not by $\mathcal{I}_{\max}(A) = \|\bar{A}\|_2$ but rather by $\|\bar{A}\|_1 = \|A\|_1$, the maximum absolute column sum. In cases where the matrix in the Wigner representation is non-negative, the matrix will be left-stochastic and $\|A\|_1 = 1$, such matrices will not increase the number of samples needed. If there are some negative values, then $\|A\|_1$ will be larger.

After the present work was completed, the quantity $\log_2 \|\boldsymbol{\rho}\|_1$ was investigated in [24]. This quantity was termed "mana" and was shown to be monotone under Clifford operations and to be monotone on average under stabilizer measurements, thus providing bounds on magic state distillation by Clifford circuits. Given the results of the present paper, it should perhaps make sense to extend the concept of mana also to quantum operations, defining their mana to be $\log_2 \|A\|_1$. Then Clifford operations have zero mana and in general the following monotonicity relation is satisfied:

$$\log_2 \|A\boldsymbol{\rho}\|_1 \leqslant \log_2(\|A\|_1\|\boldsymbol{\rho}\|_1) = \log_2 \|A\|_1 + \log_2 \|\boldsymbol{\rho}\|_1. \tag{78}$$

So $\log_2 \|A\|_1$, which is the Wigner representation analog of the $\log_2$ of interference-producing capacity, bounds the amount by which the operator A may increase the mana of a state. For each A there will be some $\boldsymbol{\rho}$ that saturates this inequality (by the definition of operator norm), but it is not clear whether this would correspond to a physical state.

Stated in this language, Theorem 2, applied in the Wigner representation, gives that quantum circuits may be efficiently simulated classically in time polynomial in $\|\boldsymbol{M}\|_\infty$ (where $\boldsymbol{M}$ is the final measurement) and exponential in the sum of the mana of the initial state and the mana of each operation. Specifically, write $\langle \boldsymbol{M}|\boldsymbol{V}\boldsymbol{U}|\boldsymbol{\rho}\rangle = \mathrm{Tr}\{|\boldsymbol{\rho}\rangle\langle \boldsymbol{M}|\boldsymbol{V}\boldsymbol{U}\}$. Then, ignoring for the moment conditions (c) and (d) of Definition 3 and (c) and (d) of Definition 4, we have (by Theorem 9) that $|\boldsymbol{\rho}\rangle\langle \boldsymbol{M}|$ is $\mathrm{EHT}_\infty(\|\boldsymbol{\rho}\|_1\|\boldsymbol{M}\|_\infty, f)$ and $\boldsymbol{U}$ is $\mathrm{EPS}_\infty(\|\boldsymbol{U}\|_1, f)$ (similarly for $\boldsymbol{V}$). So by Theorem 2 this can be simulated in time

$$O(\log_2(\delta^{-1})\epsilon^{-2}\|\boldsymbol{M}\|_\infty\|\boldsymbol{U}\|_1\|\boldsymbol{V}\|_1\|\boldsymbol{\rho}\|_1 f). \tag{79}$$

This complements the result of [24], which showed mana to be a necessary resource for magic state distillation but did not show that circuits of low total mana have no quantum speedup (although the zero mana case was treated in [9,10]).

B. Communication complexity

Consider a scenario in which two parties, Alice and Bob, are to cooperatively evaluate a Boolean function. Specifically, suppose that Alice receives input x, Bob receives input y, and they are to evaluate $g(x,y)$, where the function $g : X \times Y \to \{0,1\}$ is known to the two parties ahead of time. They must provide the correct answer with a probability of at least $2/3$. For nontrivial functions this will require communication, which can be either quantum or classical. The communication complexity of g is the number of bits of communication required by the optimal protocol, with no regard for the amount of time Alice and Bob spend on local computations. For some problems quantum communication is exponentially more efficient than classical communication [25].

Consider a quantum communication protocol as depicted by Fig. 3. The initial state, denoted $|\psi\rangle$, is a pure (but possibly entangled) state on three subsystems $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$. Subsystems $\mathcal{H}_A$ and $\mathcal{H}_B$ are owned by Alice and Bob, respectively, and subsystem $\mathcal{H}_C$ is passed between Alice and Bob through a noiseless quantum channel for each round of communication. Alice begins by performing a unitary operation $A^{(1,x)}$, which can depend on her input x on subsystems $\mathcal{H}_A \otimes \mathcal{H}_C$. She then sends the $\mathcal{H}_C$ subsystem to Bob, who performs a unitary operation $B^{(2,y)}$, which can depend on his input y, on subsystems $\mathcal{H}_B \otimes \mathcal{H}_C$. Bob sends $\mathcal{H}_C$ back to Alice, who then performs $A^{(3,x)}$ and so on. Finally, the last party (say, Bob) performs a two-outcome projective (or POVM) measurement $\{M^{(y)}, I - M^{(y)}\}$, which can depend on y, on subsystems $\mathcal{H}_B \otimes \mathcal{H}_C$ and reports the outcome. The expectation value of the final measurement is given by

$$\begin{aligned}&\langle\psi|A^{(1,x)\dagger}B^{(2,y)\dagger}A^{(3,x)\dagger}\cdots A^{(T,x)\dagger}M^{(y)}A^{(T,x)}\\ &\quad\cdots A^{(3,x)}B^{(2,y)}A^{(1,x)}|\psi\rangle\end{aligned} \tag{80}$$

and must be $\leqslant 1/3$ if $g(x,y) = 0$ and $\geqslant 2/3$ if $g(x,y) = 1$. The communication complexity of the protocol is the number of qubits transmitted, $T \log_2[\dim(\mathcal{H}_C)]$, where T is the number of rounds of communication. The dimensionality of the subsystems $\mathcal{H}_A$ and $\mathcal{H}_B$ is not taken into consideration.

The algorithm of this paper can be adapted to provide classical communication simulations of quantum communication protocols, in the case where the quantum protocols are built using operators having low interference-producing capacity, and making a certain assumption regarding the initial state $|\psi\rangle$. Since the expectation value of the final measurement in the quantum protocol will be either $\leqslant 1/3$ or $\geqslant 2/3$, a classical simulation of the quantum protocol can with probability $\geqslant 2/3$ determine $g(x,y)$ if it can, with chance of error $\delta \leqslant 1/3$, estimate the expectation value of the quantum protocol to within additive error $\epsilon < 1/6$. This is exactly the type of estimation provided by the algorithm of this paper; we need only adapt it to the communication scenario.

The algorithm presented in Sec. III D involves computing $O(b_{\max}^2)$ path samples,[12] each of which require evaluation of a left-to-right or a right-to-left Markov chain. Crucially, each

[12]Specifically, $O(\log_2(\delta^{-1})\epsilon^{-2}b_{\max}^2)$ samples are needed. However, in order to achieve the goal of guessing $g(x,y)$ with probability $\geqslant 2/3$, it suffices to set constant $\delta < 1/3$ and $\epsilon < 1/6$.

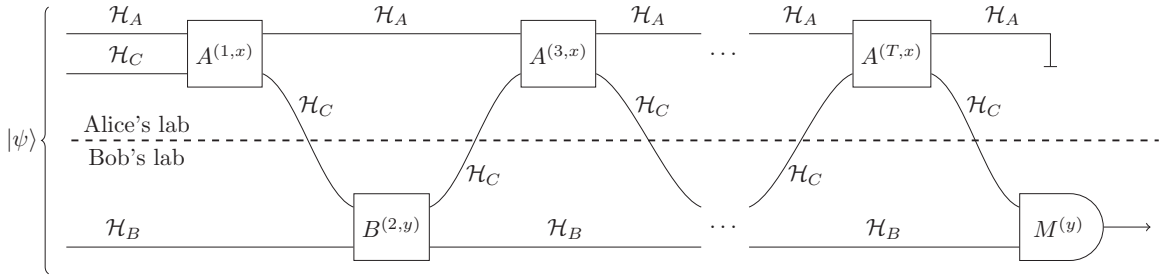

FIG. 3. A quantum communication protocol. The expectation value of the final measurement is given by (80).

transition operator in these chains is defined solely in terms of a single operator of (80). Therefore, each transition can be computed by Alice alone (for the $A^{(t,x)}$ operators) or by Bob alone [for the $B^{(t,y)}$ and $M^{(y)}$ operators]. The state space of the Markov chains consists of indices corresponding to computational basis states of $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$, so the indices can be thought of as triples (i_A,i_B,i_C) of indices over $\mathcal{H}_A$, $\mathcal{H}_B$, and $\mathcal{H}_C$. Since Alice's operators $A^{(t,x)}$ act only on subsystems $\mathcal{H}_A \otimes \mathcal{H}_C$, the corresponding transition operators in the Markov chain involve only indices i_A and i_C. Similarly, Bob's transition operators involve only i_B and i_C. Therefore, Alice and Bob need to communicate only the index i_C for each transition of the Markov chain.

Also needed is selection of the initial index according to the probability distribution $P(i_A,i_B,i_C) = |\langle i_A,i_B,i_C|\psi\rangle|^2$ [with Alice getting (i_A,i_C) and Bob getting i_B], as well as evaluation of $\langle i_A,i_B,i_C|\psi\rangle$ for a given (i_A,i_B,i_C) triple [where Alice knows (i_A,i_C) and Bob knows i_B]. If the initial state is a product state, $|\psi\rangle = |\psi_{AC}\rangle \otimes |\psi_B\rangle$, these tasks are easily accomplished using no communication. In fact, even if $|\psi\rangle$ is entangled between Alice and Bob, these two tasks can in some cases be accomplished using only a small amount of communication. Alice and Bob both know $|\psi\rangle$ (since it does not depend on x or y), so they can individually sample from $P(i_A,i_B,i_C)$. If Alice and Bob are granted access to shared randomness (aka public coins), they can sample from $P(i_A,i_B,i_C)$ in a synchronous way (i.e., they both get the same outcome). Computation of $\langle i_A,i_B,i_C|\psi\rangle$ for a given (i_A,i_B,i_C) triple, with (i_A,i_C) known to Alice and i_B known to Bob, is trickier and how much communication is needed depends on $|\psi\rangle$. For example, let $\mathcal{H}_A = \mathcal{H}_{A'} \otimes \mathcal{H}_{A''}$ and $\mathcal{H}_B = \mathcal{H}_{B'} \otimes \mathcal{H}_{B''}$ and consider an initial state of the form

$$|\psi\rangle = |\psi_{A'}\rangle \otimes |\psi_{B'}\rangle \otimes |\psi_C\rangle \otimes \sum_i \alpha_i |i\rangle_{A''} \otimes |i\rangle_{B''}, \quad (81)$$

with $|i\rangle_{A''}$ and $|i\rangle_{B''}$ denoting computational basis vectors. This is the most common type of initial state for quantum protocols that make use of shared entanglement. Then

$$\langle i_A,i_B,i_C|\psi\rangle = \langle i_{A'}|\psi_{A'}\rangle\langle i_{B'}|\psi_{B'}\rangle\langle i_C|\psi_C\rangle\alpha_{i_{A''}}\delta(i_{A''},i_{B''}), \quad (82)$$

where δ is the Kronecker δ. This can be computed using shared randomness and $O(1)$ communication by making use of a bounded error protocol for testing equality of $i_{A''}$ and $i_{B''}$ (Example 3.13 of [26]).

Since each unitary appears twice in (80), evaluation of the entire Markov chain is accomplished with twice as much communication as the classical protocol, or $2T \log_2[\dim(\mathcal{H}_C)]$ bits. The algorithm also requires computing the amplitude associated with the path, as well as the probability of the path. However, this requires only transmission of $O(T)$ scalar quantities from Alice to Bob, using $O(T)$ bits of communication.[13] The total classical communication complexity of this simulation protocol is therefore $O(b_{\max}^2 T \log_2[\dim(\mathcal{H}_C)])$, a factor $O(b_{\max}^2)$ greater than that of the quantum protocol. Using the optimal probability distributions defined in Appendix A, $b_{\max}$ is upper bounded by the product of the interference-producing capacities of the operators in (80). The communication complexity of the classical simulation is then

$$O\big(T \log_2[\dim(\mathcal{H}_C)] \max_{x,y} \big\{ \|\bar{A}^{(1,x)}\|_2^4 \|\bar{B}^{(2,y)}\|_2^4 \|\bar{A}^{(3,x)}\|_2^4 \times \cdots \|\bar{A}^{(T,x)}\|_2^4 \|\bar{M}^{(y)}\|_2^2 \big\}\big). \quad (83)$$

The consequence of this construction is that any quantum communication protocol exhibiting superpolynomial advantage in communication complexity over any classical protocol must have a superpolynomial value of $b_{\max}$ (i.e., the product of the interference-producing capacities of the quantum operators must be high) or must make use of an initial state not of the form (81). There is, however, an interesting caveat to this claim. Due to the fact that each unitary, as well as the initial state, appears twice in (80), our classical simulation will require twice as many communication rounds as the quantum protocol.[14] Our technique, therefore, does not apply if one limits the number of rounds. For example, the quantum protocol for the Perm-Invariance problem described in [27] has $b_{\max} = 1$ yet is exponentially more efficient than any one-round classical protocol.

There is a way to avoid the doubling of the number of rounds of communication, but at a price. Consider a one-round quantum protocol in which Alice sends a state $|\psi\rangle$ and Bob measures a projector (or POVM element) M. The expectation value is $\langle\psi|M|\psi\rangle = \mathrm{Tr}\{|\psi\rangle\langle\psi|M\}$. As described in the previous section, the state $|\psi\rangle\langle\psi|$ and operator M can be vectorized to give $\langle\boldsymbol{\rho}|\boldsymbol{M}\rangle = \mathrm{Tr}\{|\psi\rangle\langle\psi|M\}$. By taking $p = 1$ and $q = \infty$ instead of $p = q = 2$ our algorithm can

[13] Actually, a careful look shows that only $O(1)$ communication is needed. Alice can locally multiply her transition probabilities and the amplitudes for her operators for the given path and report these $O(1)$ values to Bob, who is then able to complete the computation.

[14] Note that independent evaluations of the Markov chain can be run in parallel; otherwise, the number of rounds would scale as $O(b_{\max}^2)$.

estimate $\langle \boldsymbol{\rho}|\boldsymbol{M}\rangle$ using only a left-to-right Markov chain, thus requiring only a single round of communication, from Alice to Bob. However, since $p = 1$ and $q = \infty$, the number of bits communicated is $O(\|\boldsymbol{\rho}\|_1^2\|\boldsymbol{M}\|_\infty^2 n)$, with n being the number of qubits in $|\psi\rangle$. The reason we cannot efficiently simulate the quantum protocol of [27] using this technique is that $\|\boldsymbol{\rho}\|_1$ is exponentially large. Interestingly, [28] provides a one-round protocol that can estimate $\langle \boldsymbol{\rho}|\boldsymbol{M}\rangle$ using $O(\|\boldsymbol{\rho}\|_2^2\|\boldsymbol{M}\|_2^2)$ bits of classical communication. However, this again fails to provide an efficient simulation since $\|\boldsymbol{M}\|_2$ is exponentially large.

C. Continuity of $\mathcal{I}$ and $\mathcal{I}_{\max}$

Our measures $\mathcal{I}_{\max}$ of Definition 2 (which we have related to quantum speedup) and $\mathcal{I}$ of Definition 1 (which we have conjectured to be related to quantum speedup) are continuous as a function of the states and operators of a circuit. To our knowledge, this is the first continuous quantity that has been identified as being a necessary resource for quantum speedup, other resources such as Schmidt rank [1] or tree width [6,7] being discrete valued.

An argument was put forth in [17] as to why most continuous quantities could not be considered as a necessary resource for quantum speedup. Although their argument focuses on functions of the state vector, such as entanglement entropy, rather than of the operators, it is still worthwhile to examine whether it is applicable to the present work. We paraphrase their argument here, modifying it slightly to fit the circuit paradigm that we have been using in this paper. Consider a quantum circuit with initial state $|0\rangle^{\otimes n}$, followed by several unitaries, terminated by a final measurement having expectation value v. Add a control to all of the operators in the circuit, $I \otimes |0\rangle\langle 0| + U \otimes |1\rangle\langle 1|$ in place of U for each unitary and similarly for the final measurement. All operators are controlled by an ancillary qubit initially in the state $\sqrt{1-\epsilon}|0\rangle + \sqrt{\epsilon}|1\rangle$. By repeating execution of the circuit $O(\epsilon^{-2})$ times, the value of v can be recovered to high accuracy. However, by setting ϵ to a sufficiently low value, the state at all times during the computation will be arbitrarily close to $|0\rangle^{\otimes n+1}$ and thus will have arbitrarily low entanglement. The most commonly used entanglement measures take values that depend polynomially on ϵ, so entanglement can be made quite low without $O(\epsilon^{-2})$ growing to an unfeasible magnitude. As a consequence, it is not possible to claim without qualification that entanglement is necessary for quantum speedup.

This construction has no effect on the interference-producing capacity of the operators of the circuit since $\mathcal{I}_{\max}(I \otimes |0\rangle\langle 0| + U \otimes |1\rangle\langle 1|) = \mathcal{I}_{\max}(U)$. For this reason, our main result regarding $\mathcal{I}_{\max}$ as a necessary resource for quantum speedup is immune to the above argument. On the other hand, the interference measure $\mathcal{I}$ of Definition 1, which is the subject of the conjectures of Sec. VII, is immune to this argument for a different reason. The value of $\mathcal{I}$ can be exponentially high in the number of qubits or number of unitaries of a circuit. In order to make $\mathcal{I}$ small, ϵ would have to be exponentially small, in turn requiring an exponentially large number of repetitions of the circuit. So the construction of [17] is not able to significantly lower the interference of a circuit without also losing the quantum speedup.

D. Connection to decoherence functional

There is a close connection between the interference $\mathcal{I}$ of Definition 1 and the decoherence functional introduced by Gell-Mann and Hartle.[15] The latter represents an extension of the Born rule so as to be able to define probabilities for a sequence of events in a closed quantum system. Consider a *family of histories* corresponding to projection onto the computational basis at each step (i.e., after the initial state and after each unitary) of a quantum circuit $\mathrm{Tr}\{U^{(1)\dagger}\cdots U^{(T)\dagger} M U^{(T)}\cdots U^{(1)}\rho\}$. In this case the *decoherence functional* is defined as

$$\mathcal{D}(\boldsymbol{j};\boldsymbol{k}) = \mathrm{Tr}[M W(\boldsymbol{j}) \rho W^\dagger(\boldsymbol{k})], \tag{84}$$

where ρ is the initial state, M is a projector, and

$$W(\boldsymbol{j}) = |j_T\rangle\langle j_T|U^T\cdots|j_2\rangle\langle j_2|U^{(2)}|j_1\rangle\langle j_1|U^{(1)}|j_0\rangle\langle j_0|. \tag{85}$$

It is convenient to think of $\mathcal{D}(\boldsymbol{j};\boldsymbol{k})$ as a matrix with rows labeled by $\boldsymbol{j}$ and columns by $\boldsymbol{k}$, and then it is not difficult to show that

$$\sum_{\boldsymbol{j}}\sum_{\boldsymbol{k}} \mathcal{D}(\boldsymbol{j};\boldsymbol{k}) = \mathrm{Tr}\{U^{(1)\dagger}\cdots U^{(T)\dagger} M U^{(T)}\cdots U^{(1)}\rho\}. \tag{86}$$

If the *consistency condition*

$$\mathcal{D}(\boldsymbol{j};\boldsymbol{k}) = 0 \quad \text{whenever} \quad \boldsymbol{j} \neq \boldsymbol{k} \tag{87}$$

is satisfied, then each diagonal element $\mathcal{D}(\boldsymbol{j};\boldsymbol{j})$ can be interpreted (up to normalization) as the probability of the history corresponding to $\boldsymbol{j}$ occurring. The sum of these diagonal elements is then equal to the expectation value of the final observable, the right side of (86), since the off-diagonal terms vanish.

It is straightforward to show that $\mathcal{I}$ of Definition 1 is equal to

$$\mathcal{I}(U^{(1)\dagger},\ldots,U^{(T)\dagger},M,U^{(T)},\ldots,U^{(1)},\rho) = \sum_{\boldsymbol{j}}\sum_{\boldsymbol{k}} |\mathcal{D}(\boldsymbol{j};\boldsymbol{k})|. \tag{88}$$

When the consistency condition (87) is satisfied, this will be equal to $\sum_{\boldsymbol{j}} \mathcal{D}(\boldsymbol{j};\boldsymbol{j})$ (since the diagonal entries are always positive), which, in turn, is equal to the right-hand side of (86). In general, (88) gives a measure of how badly the consistency condition is violated.

VII. CONJECTURES

We have shown that quantum speedup requires circuit elements with a large interference-producing capacity. In this section we formally state our conjecture that low interference (rather than low interference-producing capacity) is sufficient

[15] See [29]. Here we use the notation of Chaps 7, 8, and 10 of [30], which is more convenient for our purposes because it employs the Schrödinger rather than the Heisenberg representation.

to ensure efficient simulation of a quantum circuit. In general, we are interested in circuits of arbitrary length, but for concreteness consider the task of estimating sums of the form

$$\langle\psi|U^\dagger M U|\psi\rangle = \sum_{ijkl} V(i,j,k,l), \tag{89}$$

$$V(i,j,k,l) = \psi_i^* U_{ij}^\dagger M_{jk} U_{kl} \psi_l. \tag{90}$$

As discussed in Sec. II, this sum can be estimated by considering a number of randomly chosen paths $\pi = (i,j,k,l)$. If these paths are chosen according to the optimal probability distribution $R_{\text{opt}}(\pi)$ of (10), then the number of samples required to estimate (89) to within error ϵ (with probability δ of exceeding this error bound) is $O(\log_2(\delta^{-1})\epsilon^{-2}\mathcal{I}^2)$, where $\mathcal{I} = \langle\bar{\psi}|\bar{U}^\dagger \bar{M}\bar{U}|\bar{\psi}\rangle$ is the interference of the circuit as given by Definition 1. The difficulty with this strategy is that we do not know how to efficiently sample paths according to the distribution $R_{\text{opt}}(\pi)$, or anything sufficiently close to it. In other words, we do not have a strategy for finding the most relevant paths. However, we conjecture that there is a way.

Loosely speaking, we conjecture that a quantum circuit can be simulated in time $\text{poly}(\log_2(\delta^{-1})\epsilon^{-1}\mathcal{I})$ as long as the initial state and operators meet some computational tractability conditions, analogous to conditions (c) and (d) of Definitions 3 and 4 . Exactly which tractability conditions should be required is difficult to know ahead of time for the following reason. In Secs. II and III a simulation algorithm was developed, which required certain tasks to be performed involving the initial state and the operators of the circuit being simulated. The need to efficiently perform these tasks led directly to the definition of conditions (c) and (d). Now we conjecture a better algorithm, whose specific structure is not known ahead of time. Not knowing the specifics of this conjectured algorithm, it is not clear what should be required in place of conditions (c) and (d). The intuition is that we assume any necessary task involving any individual operator in the circuit can be efficiently performed, but we make no assumption regarding the interactions between several operators.

This can be made more precise. Section IV C (on query complexity) and Sec. VI B (on communication complexity) each provided a framework in which the computational tractability conditions (c) and (d) were not relevant. We could use either of these to form a conjecture that avoids the need to state similar conditions. Of these two, communication complexity is representative of a certain algorithmic structure. Consider algorithms that involve dealing with the elements of a circuit one at a time. For instance, when estimating (89) one could imagine carrying out some calculations involving $|\psi\rangle$, making notes of the result, carrying out further calculations involving U, and so on. The time complexity of such an algorithm is lower bounded by the amount of notes taken and the number of times attention is shifted from one circuit element to another. This can be quantified by imagining that each of $|\psi\rangle$, U, and M are stored in separate rooms, and considering how many notes need to be carried back and forth between the rooms by somebody who seeks to estimate (89). Equivalently, stated in terms of communication complexity, imagine that Alice has $|\psi\rangle$, Bob has U, and Charlie has M. How much communication is needed in order to estimate (89)? We conjecture that the amount of communication needed is polynomial in the interference of the circuit.

Conjecture 1. Suppose that Alice has a classical description of a vector $|\psi\rangle$ of dimension N, Bob has a description of an $N \times N$ POVM element M, and T other parties have descriptions of $N \times N$ unitary matrices $U^{(1)}, \ldots, U^{(T)}$. Then, with probability less than δ of exceeding the error bound, the value of

$$\langle\psi|U^{(1)\dagger}\cdots U^{(T)\dagger} M U^{(T)}\cdots U^{(1)}|\psi\rangle \tag{91}$$

can be estimated to within additive error ϵ using $\text{poly}(\log_2(\delta^{-1})\epsilon^{-1}\max\{1,\mathcal{I}\}\log_2(N))$ bits of classical communication where $\mathcal{I}$ is the interference of (91) as given by Definition 1.

The reader may worry that this communication scenario has little bearing on the problem of simulating quantum circuits; however, it is expected that any proof in the positive of this conjecture will be adaptable into an algorithm that can be used in the computation context. Indeed, the Markov chain technique of Sec. III was first developed as a solution to a problem resembling Conjecture 1.

We have been unable to prove this conjecture even for the simple case where there are no unitary operations and the goal is to estimate the expectation value $\langle\psi|M|\psi\rangle$. We present this simplified case formally, as it deserves some discussion.

Conjecture 2. Conjecture 1 holds in the case $T = 0$. In other words, suppose that Alice has a classical description of a vector $|\psi\rangle$ of dimension N and Bob has a classical description of an $N \times N$ POVM element M. Then, with probability less than δ of exceeding the error bound, the value $\langle\psi|M|\psi\rangle$ can be estimated to within additive error ϵ using $\text{poly}(\log_2(\delta^{-1})\epsilon^{-1}\max\{1,\mathcal{I}\}\log_2(N))$ bits of classical communication, where $\mathcal{I} = \langle\bar{\psi}|\bar{M}|\bar{\psi}\rangle$ is the interference of $\langle\psi|M|\psi\rangle$ as given by Definition 1.

Conjecture 2, being weaker than Conjecture 1, should be easier to prove true. However, it would probably be very difficult to prove false since a proof that estimating $\langle\psi|M|\psi\rangle$ requires a large amount of classical communication in the general case (not assuming low interference) remained open for 11 years [25].

Conjecture 2 would be false if only one round of communication was allowed, from Alice to Bob. In [27] the Perm-Invariance problem was defined and shown to be solved efficiently by a one-round quantum protocol; however, no efficient one-round classical protocol exists. The quantum protocol has Bob measuring a POVM element M on a state $|\psi\rangle$ sent by Alice and this protocol is low interference, $\mathcal{I} = \langle\bar{\psi}|\bar{M}|\bar{\psi}\rangle \leqslant 1$. However, there can be no efficient one-round classical protocol for estimating $\langle\psi|M|\psi\rangle$, since such a protocol would efficiently solve Perm Invariance. This does not provide a counterexample to Conjecture 2 since we allow multiple rounds of communication, and there is, indeed, an efficient classical two-round protocol, which can be constructed using the technique of Sec. VI B.

A potential problem with Conjecture 1 is that the unitary portion of the circuit could create very large interference which could be masked by the final measurement. For example, consider the initial state $|\psi\rangle = |0\rangle^{\otimes n}$, acted upon by an arbitrary circuit involving all but the first qubit, followed by measurement of the observable $M = |1\rangle\langle 1| \otimes I^{\otimes n-1}$. For this

circuit $\mathcal{I} = 0$ so Conjecture 1 says the expectation value can be computed in poly$(\log_2(\delta^{-1})\epsilon^{-1}n)$ time, as indeed it can in this case. However, it seems there may be similar situations in which $\mathcal{I}$ is small because of the final measurement, but the circuit is nevertheless difficult to simulate. For this reason we provide an alternate definition that quantifies the interference just before the final measurement, computed by substituting $M = I$ in Definition 1. This will be used to form a weaker conjecture.

Definition 6. The *interference* of a quantum circuit without a measurement, $U^{(T)} \cdots U^{(1)} \rho U^{(1)\dagger}, \ldots, U^{(T)\dagger}$, is

$$\mathcal{J}(U^{(T)}, \ldots, U^{(1)}, \rho) = \mathrm{Tr}\{\bar{U}^{(T)} \cdots \bar{U}^{(1)} \bar{\rho} \bar{U}^{(1)\dagger} \cdots \bar{U}^{(T)\dagger}\}. \tag{92}$$

In other words, $\mathcal{J}$ is the amount by which normalization is spoiled when destructive interference is turned into constructive interference by means of the absolute value applied to each path. This is nondecreasing in time,

$$\mathcal{J}(U^{(T)}, \ldots, U^{(1)}, \rho) \geqslant \mathcal{J}(U^{(T-1)}, \ldots, U^{(1)}, \rho), \tag{93}$$

and $\mathcal{J} = 1$ if all of the unitaries are permutation matrices as in a classical computation. We conjecture that a circuit can be efficiently simulated when $\mathcal{J}$ is small. Since $\mathcal{J}$ does not see the final measurement M, we need an extra constraint. We require M to be a projector diagonal in the computational basis.

Conjecture 3. Suppose that Alice has a classical description of a vector $|\psi\rangle$ of dimension N, Bob has a description of an $N \times N$ projector M that is diagonal in the computational basis, and T other parties have descriptions of $N \times N$ unitary matrices $U^{(1)}, \ldots, U^{(T)}$. Then, with probability less than δ of exceeding the error bound, the value of

$$\langle\psi|U^{(1)\dagger} \cdots U^{(T)\dagger} M U^{(T)} \cdots U^{(1)}|\psi\rangle \tag{94}$$

can be estimated to within additive error ϵ using poly$(\log_2(\delta^{-1})\epsilon^{-1}\mathcal{J}\log_2(N))$ bits of classical communication where $\mathcal{J} = \mathcal{J}(U^{(T)}, \ldots, U^{(1)}, |\psi\rangle\langle\psi|)$ is the interference of (94) just before the final measurement, as given by Definition 6.

VIII. SUMMARY AND OPEN PROBLEMS

We have provided an algorithm for efficiently simulating quantum circuits in which each operator has low interference-producing capacity. Therefore, interference-producing capacity is identified as a resource necessary for quantum speedup. The runtime of the simulation is quadratic in the interference-producing capacities of each operator, so it is typically exponentially slow in the length of the circuit. However, for constant-length circuits making use of operators with low interference-producing capacity (many such operators are listed in Sec. V), the simulation runs in time polynomial in the number of qubits.

In general, our technique is able to estimate expressions of the form $\langle\psi|A \cdots Z|\phi\rangle$, of which quantum circuits $\langle\psi|U^{(1)\dagger} \cdots U^{(T)\dagger} M U^{(T)} \cdots U^{(1)}|\psi\rangle$ are a special case, in time proportional to $\|\psi\|_p^2 \|\bar{A}\|_q^2 \cdots \|\bar{Z}\|_q^2 \|\phi\|_q^2$ for any $1/p + 1/q = 1$, where a bar over a vector or operator denotes entrywise absolute value in the computational basis and where $\|\cdot\|_p$ denotes the ℓ^p norm for vectors and the induced norm for operators. The choice $p = q = 2$ is most relevant for quantum mechanics, and $\|\bar{A}\|_2$ gives the interference-producing capacity of A. The technique was also generalized to expressions of the form $\mathrm{Tr}\{A \cdots Z\sigma\}$.

We formalized the conditions necessary for efficient simulation by introducing two definitions: EHT for the initial state σ and EPS for the operators $A, \ldots, Z$. These definitions consist of requirements having to do with the number of samples needed as well as requirements having to do with efficient computability. The latter requirements can for the most part be ignored if one is concerned with query complexity or communication complexity rather than time complexity. A wide range of initial states and operators are EHT or EPS; many examples are listed in Sec. V. In addition to discussing circuits which can be efficiently simulated, we gave several examples of circuits which we cannot efficiently simulate and explained why.

The choice $p = q = 2$ makes the most sense for simulating expressions of the form $\langle\psi|U^\dagger V^\dagger M V U|\psi\rangle$. However, using the Wigner representation this expression can also be written as $\langle \boldsymbol{M}|\boldsymbol{VU}|\boldsymbol{\rho}\rangle$, and here the choice $p = \infty$ and $q = 1$ works well, allowing efficient simulation of circuits that consist mainly of Clifford operations. We showed how our simulation technique can be applied to communication problems, with the conclusion that there can be no superpolynomial advantage of quantum communication over classical communication unless the quantum protocol uses operations with high interference-producing capacity. Curiously, this result does not apply to one-round communication, since our simulation requires doubling the number of rounds. Indeed, there is an example of a one-round quantum protocol with low interference-producing capacity which is exponentially more efficient than any one-round classical protocol.

Finally, we would like to suggest three open questions.

(1) Can it be shown that interference, rather than interference-producing capacity, is necessary for quantum speedup? In Sec. VII we formalized a series of conjectures on this topic, using the framework of communication complexity.

(2) While we have shown interference-producing capacity to be a necessary resource for quantum speedup, it is also fruitful to investigate sufficient resources for quantum speedup. For example, Ref. [31], building on the work of [32], showed that any operator U having the property that $\max_{ij} |U_{ij}|$ is sufficiently small can be used to exhibit exponential quantum speedup. Can the gap between necessary (e.g., our result) and the sufficient (e.g., [31]) conditions for quantum speedup be narrowed?

(3) Can our technique be combined with existing Monte Carlo or other techniques to provide an improved simulation algorithm for systems of physical interest? Our algorithm in its present form is not likely to be more efficient than existing techniques for such problems.

ACKNOWLEDGMENTS

The author thanks Robert Griffiths and Scott Cohen for many helpful comments and suggestions. This research received financial support from the National Science Foundation through Grant No. PHY-1068331.

APPENDIX A: GENERALIZED SINGULAR VECTORS

The goal of this Appendix is to determine the minimum value of b such that a given operator A is $\mathrm{EPS}_p(b,f)$ and bounds on b such that an operator σ is $\mathrm{EHT}_p(b,f)$. We show that conditions (a) and (b) of Definition 3 require $b \geqslant \|\bar{A}\|_q$ and construct probability distributions that satisfy this with equality. Whether these also satisfy conditions (c) and (d) of Definition 3 needs to be determined on a case-by-case basis. Note that when $p = q = 2$ we have $\|\bar{A}\|_2 = \mathcal{I}_{\max}(A)$, the interference-producing capacity of A. The end result of this appendix is the following theorem.[16]

Theorem 9. Let A and σ be matrices, $p,q \in [1,\infty]$, and $1/p + 1/q = 1$. Then the following conditions apply.

(a) It is not possible to satisfy conditions (a) and (b) of Definition 3 unless $b \geqslant \|\bar{A}\|_q$. The same goes for (a) and (b) of Definition 4 since they are stricter (i.e., $b \geqslant \|\bar{\sigma}\|_q$).

(b) It is possible to satisfy conditions (a) and (b) of Definition 3 with $b = \|\bar{A}\|_q$. The k index is not needed (i.e., $k \in K = \{0\}$ and $\alpha_{mnk} = A_{mn}$).

(c) If one is concerned with query complexity rather than time complexity, and if A is not defined in terms of an oracle, then conditions (c) and (d) of Definition 3 can be ignored, as explained in Sec. IV C. Therefore, A is $\mathrm{EPS}_p(\|\bar{A}\|_q,0)$.

(d) Let w be the smallest value such that σ/w is a convex combination of normalized dyads. That is to say, let

$$w = \min\left\{ \sum_i |s_i| \,|\, s_i \in \mathbb{C}, \sigma = \sum_i s_i \boldsymbol{v}^{(i)} \boldsymbol{u}^{(i)\top}, \right.$$
$$\left. \|\boldsymbol{u}^{(i)}\|_p = \|\boldsymbol{v}^{(i)}\|_q = 1 \right\}. \tag{A1}$$

It is possible to satisfy conditions (a) and (b) of Definition 4 with $b = w$ (although this is not necessarily the smallest possible value of b). The k index is not needed (i.e., $k \in K = \{0\}$ and $\alpha_{mnk} = \sigma_{mn}$). Note that when $p = q = 2$, w is the trace norm of σ.

(e) If one is concerned with query complexity rather than time complexity, and if σ is not defined in terms of an oracle, then conditions (c) and (d) of Definition 4 can be ignored. Therefore, σ is $\mathrm{EHT}_p(w,0)$ (although this is not necessarily the smallest possible value of b).

We present immediately the proofs of parts (a), (d), and (e). Parts (b) and (c) will require more preliminary discussion.

Proof of Theorem 9(a). Let A be an $M \times N$ matrix. Suppose conditions (a) and (b) of Definition 3 are satisfied by some b, K, α_{mnk}, $P(n,k|m)$, and $Q(m,k|n)$. Then, for all $m \in \{1,\ldots,M\}$, $n \in \{1,\ldots,N\}$, and $k \in K$, we have $A_{mn} = \sum_{k' \in K} \alpha_{mnk'}$ and

$$\frac{|\alpha_{mnk}|}{P(n,k|m)^{1/p} Q(m,k|n)^{1/q}} \leqslant b. \tag{A2}$$

Rearranging this expression yields

$$|\alpha_{mnk}| \leqslant b \cdot P(n,k|m)^{1/p} Q(m,k|n)^{1/q}. \tag{A3}$$

Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be non-negative vectors satisfying $\|\boldsymbol{u}\|_p = \|\boldsymbol{v}\|_q = 1$ and $\boldsymbol{u}^\top \bar{A} \boldsymbol{v} = \|\bar{A}\|_q$ (that such vectors exist is well known, but is also a consequence of Theorem 10). Multiply both sides of (A3) by $u_m v_n$ and sum over m,n,k to get

$$\sum_{mnk} u_m |\alpha_{mnk}| v_n$$
$$\leqslant b \sum_{mnk} u_m P(n,k|m)^{1/p} Q(m,k|n)^{1/q} v_n \tag{A4}$$
$$= b \sum_{mnk} \left[P(n,k|m) u_m^p\right]^{1/p} \left[Q(m,k|n) v_n^q\right]^{1/q} \tag{A5}$$
$$\leqslant b \sum_{mnk} \left[\frac{1}{p} P(n,k|m) u_m^p + \frac{1}{q} Q(m,k|n) v_n^q\right] \tag{A6}$$
$$= b \sum_m \frac{1}{p} u_m^p + \sum_n \frac{1}{q} v_n^q \tag{A7}$$
$$= b(1/p + 1/q) \tag{A8}$$
$$= b, \tag{A9}$$

where (A6) follows from the inequality of arithmetic and geometric means. We now place a lower bound on the left-hand side. By the triangle inequality, $\sum_k |\alpha_{mnk}| \geqslant |\sum_k \alpha_{mnk}| = |A_{mn}|$ for all m,n. Since $\boldsymbol{u}$ and $\boldsymbol{v}$ are non-negative,

$$b \geqslant \sum_{mnk} u_m |\alpha_{mnk}| v_n \tag{A10}$$
$$\geqslant \sum_{mn} u_m |A_{mn}| v_n \tag{A11}$$
$$= \|\bar{A}\|_q. \tag{A12}$$

■

Proof of Theorems 9(d) and 9(e) . Let σ be an $M \times N$ matrix. Let s_i, $\boldsymbol{u}^{(i)}$, and $\boldsymbol{v}^{(i)}$ take values achieving the minimum in (A1). By absorbing a phase into $\boldsymbol{u}^{(i)}$ we can assume that the s_i are positive. We then have $w = \sum_i s_i$, $\|\boldsymbol{u}^{(i)}\|_p = \|\boldsymbol{v}^{(i)}\|_q = 1$, and $\sigma = \sum_i s_i \boldsymbol{v}^{(i)} \boldsymbol{u}^{(i)\top}$. Define

$$P(n) = \sum_i \frac{s_i}{w} \left|u_n^{(i)}\right|^p, \tag{A13}$$

$$Q(m) = \sum_i \frac{s_i}{w} \left|v_m^{(i)}\right|^q. \tag{A14}$$

Since $\boldsymbol{u}^{(i)}$ and $\boldsymbol{v}^{(i)}$ are normalized for all i, and since $\sum_i s_i/w = 1$, these $P(n)$ and $Q(m)$ are convex combinations of probability distributions and hence are probability distributions themselves.

For any $m \in \{1,\ldots,M\}, n \in \{1,\ldots,N\}$, Hölder's inequality gives

$$\sum_i \frac{s_i^{1/p}}{w^{1/p}} \left|u_n^{(i)}\right| \frac{s_i^{1/q}}{w^{1/q}} \left|v_m^{(i)}\right|$$
$$\leqslant \left[\sum_i \left(\frac{s_i^{1/p}}{w^{1/p}} \left|u_n^{(i)}\right|\right)^p\right]^{1/p} \left[\sum_i \left(\frac{s_i^{1/q}}{w^{1/q}} \left|v_m^{(i)}\right|\right)^q\right]^{1/q} \tag{A15}$$

$$\Rightarrow \sum_i \frac{s_i}{w} \left|u_n^{(i)} v_m^{(i)}\right| \leqslant \left[\sum_i \frac{s_i}{w} \left|u_n^{(i)}\right|^p\right]^{1/p} \left[\sum_i \frac{s_i}{w} \left|v_m^{(i)}\right|^q\right]^{1/q} \tag{A16}$$

[16] In the case $p = q = 2$, claims (a) and (b) of Theorem 9 are similar to results of [33], although the techniques are different.

$$\Rightarrow \left|\sum_i \frac{s_i}{w} u_n^{(i)} v_m^{(i)}\right| \leqslant P(n)^{1/p} Q(m)^{1/q} \tag{A17}$$

$$\Rightarrow \frac{|\sigma_{mn}|}{w} \leqslant P(n)^{1/p} Q(m)^{1/q} \tag{A18}$$

$$\Rightarrow \frac{|\sigma_{mn}|}{P(n)^{1/p} Q(m)^{1/q}} \leqslant w. \tag{A19}$$

Therefore, conditions (a) and (b) of Definition 4 are satisfied with $\alpha_{mn0} = \sigma_{mn}$ and $b = w$.

If one is concerned with query complexity rather than time complexity, and if σ is not defined in terms of an oracle, then conditions (c) and (d) of Definition 4 are satisfied trivially with $f = 0$ since no oracle queries are needed in order to carry out the required operations. So σ is $\mathrm{EHT}_p(w,0)$. ■

We now begin construction of the probability distributions satisfying conditions (a) and (b) of Definition 3 with $b = \|\bar{A}\|_q$. The bulk of the discussion concerns the $p \in (1,\infty)$ case; the reader interested only in $p = 1$ or $p = \infty$ may skip directly to the second half of the proof of Theorems 9(b) and 9(c) at the end of this section.

It suffices to let k take only a single value, say $k = 0$, and to set $\alpha_{mn0} = A_{mn}$. Making this simplification, and plugging in the desired bound $b = \|\bar{A}\|_q$, conditions (a) and (b) of Definition 3 become

$$\max_{mn}\left\{\frac{|A_{mn}|}{P(n|m)^{1/p} Q(m|n)^{1/q}}\right\} \leqslant \|\bar{A}\|_q. \tag{A20}$$

It will be convenient to derive the probability distributions from a pair of vectors. With A being an $M \times N$ matrix, let $\boldsymbol{u}$ be a positive vector of dimension M and let $\boldsymbol{v}$ be a positive vector of dimension N. Taking the probability distributions

$$P(n|m) = |A_{mn}| v_n / [\bar{A}\boldsymbol{v}]_m, \tag{A21}$$

$$Q(m|n) = |A_{mn}| u_m / [\bar{A}^\top \boldsymbol{u}]_n, \tag{A22}$$

brings (A20) to the form

$$\max_{mn}\left\{\left(\frac{[\bar{A}\boldsymbol{v}]_m}{v_n}\right)^{1/p}\left(\frac{[\bar{A}^\top \boldsymbol{u}]_n}{u_m}\right)^{1/q}\right\} \leqslant \|\bar{A}\|_q. \tag{A23}$$

Consider for a moment the case $p = q = 2$. If $\bar{A}$ is not block diagonal (even under permutations of rows and columns) then the left and right singular vectors of $\bar{A}$ will be positive. Taking these for $\boldsymbol{u}$ and $\boldsymbol{v}$ it is easy to see that (A23) holds. If $p \neq 2$ we can use a sort of generalization of singular vectors: We show the existence of positive vectors satisfying

$$(\bar{A}^\top \boldsymbol{u})_n \leqslant v_n^{q/p} \|\bar{A}\|_q, \tag{A24}$$

$$(\bar{A}\boldsymbol{v})_m \leqslant u_m^{p/q} \|\bar{A}\|_q. \tag{A25}$$

These vectors are easily seen to satisfy (A23). If $\bar{A}$ is not block diagonal then $\boldsymbol{u}$ and $\boldsymbol{v}$ can be computed using the power method [34,35] since $\bar{A}$ is non-negative. In this case the inequalities (A24) and (A25) become equalities. On the other hand, if $\bar{A}$ is block diagonal then $\boldsymbol{u}$ and $\boldsymbol{v}$ can be built from the generalized left and right singular vectors of each block. The rest of this section is devoted to proving the existence of such vectors.

First we need some basic facts about ℓ^p norms. If $\boldsymbol{v}$ is a real vector normalized under the ℓ^2 norm, then $\boldsymbol{u} = \boldsymbol{v}$ is the unique ℓ^2-normalized vector with the property that $\boldsymbol{u}^\top \boldsymbol{v} = 1$. This generalizes to arbitrary ℓ^p norms, with some adaptation.

Definition 7. Let $p,q \in [1,\infty]$ and $1/p + 1/q = 1$. Let $\boldsymbol{v} \in \ell^q$. Any $\boldsymbol{u} \in \ell^p$ satisfying the conditions $\boldsymbol{u}^\top \boldsymbol{v} = \|\boldsymbol{v}\|_q$ and $\|\boldsymbol{u}\|_p = 1$ is called a *support functional* of $\boldsymbol{v}$.

Lemma 2. Let $p,q \in (1,\infty)$ and $1/p + 1/q = 1$. For any nonzero $\boldsymbol{v} \in \ell^q$, the vector $\boldsymbol{u} \in \ell^p$ defined by

$$u_i = \|\boldsymbol{v}\|_q^{-q/p} |v_i|^{q/p} \mathrm{sgn}(v_i) \tag{A26}$$

is the unique support functional of $\boldsymbol{v}$. Similarly, for any nonzero $\boldsymbol{u} \in \ell^p$, the vector $\boldsymbol{v} \in \ell^q$ defined by

$$v_i = \|\boldsymbol{u}\|_p^{-p/q} |u_i|^{p/q} \mathrm{sgn}(u_i) \tag{A27}$$

is the unique support functional of $\boldsymbol{u}$.

Proof. Uniqueness of the support functional when $1 < p < \infty$ follows from strict convexity of the norm (Chap. 11 of [36]). That the specific vectors (A26) and (A27) are support functionals is easily verified through direct computation [37]. ■

We now describe generalized singular vectors. Ordinary ($p = 2$) left and right singular vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ satisfy $\|A\boldsymbol{v}\|_2 = \|A^\top \boldsymbol{u}\|_2 = \|A\|_2$, furthermore $\boldsymbol{u}$ is the support functional of $A\boldsymbol{v}$ (since $p = 2$ this just means that $\boldsymbol{u} \propto A\boldsymbol{v}$), and $\boldsymbol{v}$ is the support functional of $A^\top \boldsymbol{u}$. These properties generalize to arbitrary ℓ^p norms, as we now show.

Theorem 10. Let $p,q \in [1,\infty]$ and $1/p + 1/q = 1$. Let A be a matrix. Then there are vectors $\boldsymbol{u} \in \ell^p$ and $\boldsymbol{v} \in \ell^q$ such that

(a) $\|\boldsymbol{u}\|_p = \|\boldsymbol{v}\|_q = 1$;
(b) $\boldsymbol{u}^\top A\boldsymbol{v} = \|A^\top \boldsymbol{u}\|_p = \|A\boldsymbol{v}\|_q = \|A\|_q = \|A^\top\|_p$;
(c) $\boldsymbol{u}$ is a support functional of $A\boldsymbol{v}$;
(d) $\boldsymbol{v}$ is a support functional of $A^\top \boldsymbol{u}$;
(e) if A is non-negative, then $\boldsymbol{u}$ and $\boldsymbol{v}$ are non-negative.

Proof. Let $\boldsymbol{v}$ be a vector satisfying $\|\boldsymbol{v}\|_q = 1$ and $\|A\boldsymbol{v}\|_q = \|A\|_q$. Such a vector is guaranteed to exist (see Definition 5.6.1 of [38]). Let $\boldsymbol{u}$ be a support functional of $A\boldsymbol{v}$. By the definition of a support functional, $\|\boldsymbol{u}\|_p = 1$, so claims (a) and (c) have been proved. With these two vectors defined, we have

$$\|A\|_q = \|A\boldsymbol{v}\|_q \tag{A28}$$

$$= \boldsymbol{u}^\top A\boldsymbol{v} \quad (\boldsymbol{u} \text{ is the support functional of } A\boldsymbol{v}) \tag{A29}$$

$$= \boldsymbol{v}^\top (A^\top \boldsymbol{u}) \tag{A30}$$

$$\leqslant \|\boldsymbol{v}\|_q \|A^\top \boldsymbol{u}\|_p \quad (\text{Hölder's inequality}) \tag{A31}$$

$$= \|A^\top \boldsymbol{u}\|_p \tag{A32}$$

$$\leqslant \|A^\top\|_p \|\boldsymbol{u}\|_p \tag{A33}$$

$$= \|A^\top\|_p. \tag{A34}$$

By symmetry we also have $\|A^\top\|_p \leqslant \|A\|_q$; therefore, the inequalities become equalities. Claim (b) is proved. Since $\|\boldsymbol{v}\|_q = 1$ and $\boldsymbol{v}^\top (A^\top \boldsymbol{u}) = \|A^\top \boldsymbol{u}\|_p$, claim (d) is proved as well.

To prove claim (e), assume that A is non-negative. Then $\|\bar{\boldsymbol{u}}\|_p = \|\bar{\boldsymbol{v}}\|_q = 1$ and $\|A\bar{\boldsymbol{v}}\|_q \geqslant \bar{\boldsymbol{u}}^\top A\bar{\boldsymbol{v}} \geqslant \boldsymbol{u}^\top A\boldsymbol{v} = \|A\|_q$. It follows that $\|A\bar{\boldsymbol{v}}\|_q = \|A\|_q$; thus, $\bar{\boldsymbol{u}}$ is a support functional of $A\bar{\boldsymbol{v}}$. Therefore, $\bar{\boldsymbol{u}}$ and $\bar{\boldsymbol{v}}$ could have been taken instead of $\boldsymbol{u}$ and $\boldsymbol{v}$ in the first steps of this proof, justifying the claim that $\boldsymbol{u}$ and $\boldsymbol{v}$ can be chosen to be non-negative. ■

The Perron-Frobenius theorem states that an irreducible non-negative matrix has a first eigenvector that has positive components. A similar statement holds for the first singular vector: If $\bar{A}$ is a non-negative matrix that is not block diagonal, then the left and right singular vectors associated with the largest singular value of $\bar{A}$ have positive entries. This is true also for our generalized singular vectors, as we now show.

Definition 8. A matrix A is *block diagonal* if there are permutation matrices σ and τ such that A can be decomposed as $\bar{A} = \sigma^\top(A^{(1)} \oplus \cdots \oplus A^{(L)} \oplus \mathbf{0}^{M\times N})\tau$, where the $A^{(l)}$ are nonzero and have nonvanishing dimension, and at least one of the inequalities $L > 1$, $M > 0$, or $N > 0$ holds.[17] A matrix is *not block diagonal* if no such decomposition is possible. In particular, a matrix that is not block diagonal has no totally zero rows or columns.

Lemma 3. Let $q \in (1,\infty)$. Let $\bar{A}$ be a non-negative matrix that is not block diagonal. Let $\boldsymbol{v}$ be a nonzero, non-negative vector that maximizes $\|\bar{A}\boldsymbol{v}\|_q/\|\boldsymbol{v}\|_q$. Then $\boldsymbol{v}$ is, in fact, a positive vector (has no zero entries).

Proof. Let $Z = \{i : v_i = 0\}$. This will be a proof by contradiction; suppose that $\boldsymbol{v}$ has at least one zero entry, so that Z is nonempty. Since $\boldsymbol{v} \neq 0$, the complement Z^C is nonempty; therefore, Z and Z^C partition the entries of $\boldsymbol{v}$ into two nonempty sets. Also, Z and Z^C can be considered as a partition of the columns of $\bar{A}$. Since $\bar{A}$ is not block diagonal, there must be indices $i \in Z$, $j \notin Z$, and k such that $\bar{A}_{ki} > 0$ and $\bar{A}_{kj} > 0$. We show that $\boldsymbol{v}$ cannot maximize $\|\bar{A}\boldsymbol{v}\|_q/\|\boldsymbol{v}\|_q$ by showing that $\boldsymbol{v}$ is not a critical point of $\|\bar{A}\boldsymbol{v}\|_q/\|\boldsymbol{v}\|_q$ or, equivalently, of $\|\bar{A}\boldsymbol{v}\|_q^q/\|\boldsymbol{v}\|_q^q$. Without loss of generality, take $\|\boldsymbol{v}\|_q = 1$. Let $\hat{\imath}$ be the unit vector corresponding to i. We have

$$\frac{\partial}{\partial\alpha}\frac{\|\bar{A}(\boldsymbol{v}+\alpha\hat{\imath})\|_q^q}{\|\boldsymbol{v}+\alpha\hat{\imath}\|_q^q}\bigg|_{\alpha=0}$$

$$= \frac{\left(\frac{\partial}{\partial\alpha}\|\bar{A}(\boldsymbol{v}+\alpha\hat{\imath})\|_q^q\right)\|\boldsymbol{v}\|_q^q - \|\bar{A}\boldsymbol{v}\|_q^q\left(\frac{\partial}{\partial\alpha}\|\boldsymbol{v}+\alpha\hat{\imath}\|_q^q\right)}{\|\boldsymbol{v}\|_q^{2q}}\bigg|_{\alpha=0} \tag{A35}$$

$$= \frac{\partial}{\partial\alpha}\|\bar{A}(\boldsymbol{v}+\alpha\hat{\imath})\|_q^q\big|_{\alpha=0} \tag{A36}$$

$$= \frac{\partial}{\partial\alpha}\sum_l([\bar{A}\boldsymbol{v}]_l + \alpha\bar{A}_{li})^q|_{\alpha=0} \tag{A37}$$

$$= \sum_l q\bar{A}_{li}[\bar{A}\boldsymbol{v}]_l^{q-1} \tag{A38}$$

$$\leqslant q\bar{A}_{ki}[\bar{A}\boldsymbol{v}]_k^{q-1} \tag{A39}$$

$$\leqslant q\bar{A}_{ki}(\bar{A}_{kj}v_j)^{q-1} \tag{A40}$$

$$> 0. \tag{A41}$$

Equality (A36) follows from $\|\boldsymbol{v}\|_q = 1$ as well as $(v_i = 0 \Rightarrow \partial\|\boldsymbol{v}+\alpha\hat{\imath}\|_q^q/\partial\alpha = 0)$. Inequality (A39) follows from each term of the previous summation being non-negative. Inequality (A40) follows from each term of the sum $[\bar{A}\boldsymbol{v}]_k = \sum_n \bar{A}_{kn}v_n$ being non-negative. ■

[17] If $M > 0, N = 0$, then $\oplus\mathbf{0}^{M\times N}$ adds M rows of zeros. Similarly, if $M = 0, N > 0$ then $\oplus\mathbf{0}^{M\times N}$ adds N columns of zeros.

Theorem 11. Let $p,q \in (1,\infty)$ and $1/p + 1/q = 1$. Let $\bar{A}$ be a non-negative matrix that is not block diagonal. Then there are positive vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ satisfying

$$(\bar{A}^\top\boldsymbol{u})_n = v_n^{q/p}\|\bar{A}\|_q, \tag{A42}$$

$$(\bar{A}\boldsymbol{v})_m = u_m^{p/q}\|\bar{A}\|_q. \tag{A43}$$

Note that if $p = q = 2$, then $\boldsymbol{u}$ and $\boldsymbol{v}$ will be the left and right singular vectors associated with the largest singular value of $\bar{A}$.

Proof. Theorem 10 guarantees the existence of non-negative vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ that satisfy $\|\boldsymbol{u}\|_p = \|\boldsymbol{v}\|_q = 1$ and $\boldsymbol{u}^\top\bar{A}\boldsymbol{v} = \|\bar{A}\|_q = \|\bar{A}^\top\|_p$, with $\boldsymbol{u}$ being the support functional of $A\boldsymbol{v}$ and $\boldsymbol{v}$ being the support functional of $A^\top\boldsymbol{u}$. Lemma 2 gives the exact form of these support functionals:

$$u_m = \|\bar{A}\boldsymbol{v}\|_q^{-q/p}(\bar{A}\boldsymbol{v})_m^{q/p}\mathrm{sgn}(\bar{A}\boldsymbol{v}), \tag{A44}$$

$$v_n = \|\bar{A}^\top\boldsymbol{u}\|_p^{-p/q}(\bar{A}^\top\boldsymbol{u})_n^{p/q}\mathrm{sgn}(\bar{A}^\top\boldsymbol{u}). \tag{A45}$$

Since $\bar{A}, \boldsymbol{u}$, and $\boldsymbol{v}$ are non-negative, the sgn functions disappear. Theorem 10 gives $\|\bar{A}\boldsymbol{v}\|_q = \|\bar{A}^\top\boldsymbol{u}\|_p = \|\bar{A}\|_q$. With these simplifications, we get (A42) and (A43). That $\boldsymbol{u}$ and $\boldsymbol{v}$ have nonzero entries follows from Lemma 3 . ■

We now generalize Theorem 11 to matrices that are not block diagonal. This is done by applying Theorem 11 to each individual block of the matrix. Each block of $\bar{A}$ may have a different operator norm, but each of these is upper bounded by $\|\bar{A}\|_q$. For this reason, we end up with an inequality rather than an equality when generalizing (A42) and (A43).

Theorem 12. Let $p,q \in (1,\infty)$ and $1/p + 1/q = 1$. Let $\bar{A}$ be a non-negative matrix that can possibly be block diagonal and that may have some totally zero rows or columns. Then there are positive vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ satisfying

$$(\bar{A}^\top\boldsymbol{u})_n \leqslant v_n^{q/p}\|\bar{A}\|_q, \tag{A46}$$

$$(\bar{A}\boldsymbol{v})_m \leqslant u_m^{p/q}\|\bar{A}\|_q. \tag{A47}$$

Proof. Let σ and τ be permutation matrices that bring out the block structure of $\bar{A}$, and let $A^{(1)}, \ldots, A^{(L)}$ be the blocks. Specifically, suppose $\sigma^\top(A^{(1)} \oplus \cdots \oplus A^{(L)} \oplus \mathbf{0}^{M\times N})\tau = \bar{A}$, where the $A^{(1)} \cdots A^{(L)}$ matrices are not block diagonal and $\mathbf{0}^{M\times N}$ is an M-by-N matrix of zeros (if there is no zero block, then just take $M = N = 0$). It is easy to see that $\|A^{(l)}\|_q \leqslant \|\bar{A}\|_q$ for all $l \in \{1, \ldots, L\}$.

By Theorem 11, there are positive vectors $\boldsymbol{u}^{(1)}, \ldots, \boldsymbol{u}^{(L)}$ and $\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(L)}$ such that

$$(A^{(l)\top}\boldsymbol{u}^{(l)})_n = v_n^{(l)q/p}\|A^{(l)}\|_q \tag{A48}$$

$$\leqslant v_n^{(l)q/p}\|\bar{A}\|_q, \tag{A49}$$

$$(A^{(l)}\boldsymbol{v}^{(l)})_m = u_m^{(l)p/q}\|A^{(l)}\|_q \tag{A50}$$

$$\leqslant u_m^{(l)p/q}\|\bar{A}\|_q, \tag{A51}$$

for all $l \in \{1, \ldots, L\}$. Define $\boldsymbol{u} = \sigma^\top(\boldsymbol{u}^{(1)} \oplus \cdots \oplus \boldsymbol{u}^{(L)} \oplus \mathbf{1}^M)$ and $\boldsymbol{v} = \tau^\top(\boldsymbol{v}^{(1)} \oplus \cdots \oplus \boldsymbol{v}^{(L)} \oplus \mathbf{1}^N)$, where $\mathbf{1}^M$ and $\mathbf{1}^N$ are the all-ones vectors of lengths M and N, respectively. Then (A48)–(A51) imply (A46) and (A47). Since the $\boldsymbol{u}^{(1)}, \ldots, \boldsymbol{u}^{(L)}$ and $\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(L)}$ are positive, $\boldsymbol{u}$ and $\boldsymbol{v}$ are positive. ■

We are now ready to complete the proof of Theorem 9.

Proof of Theorems 9(a) and 9(b). Let A be a matrix. Set $K = \{0\}$ and $\alpha_{mn0} = A_{mn}$. Clearly, condition (a) of Definition 3 is satisfied.

Consider the case $p \in (1,\infty)$. Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be positive vectors satisfying (A46) and (A47). The existence of such vectors is guaranteed by Theorem 12. Define the probability distributions

$$P(n|m) = |A_{mn}|v_n/[\bar{A}\boldsymbol{v}]_m, \tag{A52}$$

$$Q(m|n) = |A_{mn}|u_m/[\bar{A}^\top \boldsymbol{u}]_n. \tag{A53}$$

These satisfy condition (b) of Definition 3 with $b = \|\bar{A}\|_q$ since

$$\max_{mnk}\left\{\frac{|\alpha_{mnk}|}{P(n|m)^{1/p}Q(m|n)^{1/q}}\right\}$$

$$= \max_{mn}\left\{\frac{|A_{mn}|}{P(n|m)^{1/p}Q(m|n)^{1/q}}\right\} \tag{A54}$$

$$= \max_{mn}\left\{\left(\frac{[\bar{A}\boldsymbol{v}]_m}{v_n}\right)^{1/p}\left(\frac{[\bar{A}^\top\boldsymbol{u}]_n}{u_m}\right)^{1/q}\right\} \tag{A55}$$

$$\leqslant \max_{mn}\left\{\left(\frac{u_m^{p/q}\|\bar{A}\|_q}{v_n}\right)^{1/p}\left(\frac{v_n^{q/p}\|\bar{A}\|_q}{u_m}\right)^{1/q}\right\} \tag{A56}$$

$$= \|\bar{A}\|_q. \tag{A57}$$

Now consider the case $p = 1$, $q = \infty$ (the case $p = \infty$, $q = 1$ follows by a symmetrical argument). Define $P(n|m) = |A_{mn}|/\sum_{n'}|A_{mn'}|$ and define $Q(m|n)$ arbitrarily. Condition (b) of Definition 3 is satisfied with $b = \|\bar{A}\|_\infty$ since

$$\max_{mnk}\left\{\frac{|\alpha_{mnk}|}{P(n|m)^{1/p}Q(m|n)^{1/q}}\right\}$$

$$= \max_{mn}\left\{\frac{|A_{mn}|}{P(n|m)^1 Q(m|n)^0}\right\} \tag{A58}$$

$$= \max_{mn}\left\{\frac{|A_{mn}|}{|A_{mn}|/\sum_{n'}|A_{mn'}|}\right\} \tag{A59}$$

$$\leqslant \|\bar{A}\|_\infty. \tag{A60}$$

If one is concerned with query complexity rather than time complexity, and if A is not defined in terms of an oracle, then conditions (c) and (d) of Definition 3 are satisfied trivially with $f = 0$ since no oracle queries are needed in order to carry out the required operations. So A is $\mathrm{EPS}_p(\|\bar{A}\|_q,0)$. ■

APPENDIX B: PROOFS FOR SEC. IV

In this section we prove Theorem 4 and Lemma 1. The proofs are conceptually rather simple; however, they are notationally tedious. Since we will at times be manipulating infinite series, we begin by showing that these series converge absolutely. This will be useful, since absolutely convergent series allow permutation of terms and reordering of double summations.

Lemma 4. Let b and α_{mnk} satisfy condition (b) of Definition 3. Then series $\sum_{k\in K}\alpha_{mnk}$ is absolutely convergent for all m,n, and $\sum_{k\in K}|\alpha_{mnk}| \leqslant b$.

Proof. Rearranging (55) of condition (b) gives, for all m,n,k,

$$|\alpha_{mnk}| \leqslant bP(n,k|m)^{1/p}Q(m,k|n)^{1/q} \tag{B1}$$

$$\leqslant b[P(n,k|m)/p + Q(m,k|n)/q]. \tag{B2}$$

Therefore,

$$\sum_{k\in K}|\alpha_{mnk}| \leqslant b\sum_{k\in K}[P(n,k|m)/p + Q(m,k|n)/q] \tag{B3}$$

$$= b[P(n|m)/p + Q(m|n)/q] \tag{B4}$$

$$\leqslant b \tag{B5}$$

$$< \infty. \tag{B6}$$

■

We now prove that linear combinations of EPS operators are EPS. Theorem 4(a), regarding sums of EPS operators, follows as a corollary. This will also be used to prove Theorem 4 (c), regarding exponentials of EPS operators.

Theorem 13. Linear combination of EPS. Let L be a finite or countable set. For $l \in L$, let s_l be a complex number and let $A^{(l)}$ be an $M \times N$ matrix that is $\mathrm{EPS}_p(b_l, f_l)$ for some f_l and b_l. Let $W(l)$ be a probability distribution[18] on l such that $W(l)$ can be sampled from, and $s_l/W(l)$ computed, in average time $O(f_0)$. Let $b := \max_l\{|s_l|b_l/W(l)\} < \infty$ and $f := f_0 + \sum_l W(l)f_l$. Then $\sum_l s_l A^{(l)}$ is $\mathrm{EPS}_p(b,f)$.

Proof. For each $l \in L$, $A^{(l)}$ is $\mathrm{EPS}_p(b_l,f_l)$, so there are K_l, $\alpha^{(l)}_{mnk}$, $P_l(n,k|m)$, and $Q_l(m,k|n)$ satisfying Definition 3. Let $K = L \times \cup_{l\in L}K_l$. For $(l,k) \in K$ define

$$\alpha_{mn(l,k)} = \begin{cases} s_l\alpha^{(l)}_{mnk} & \text{if} \quad k \in K_l, \\ 0 & \text{otherwise.} \end{cases} \tag{B7}$$

$\sum_{(l,k)\in K}\alpha_{mn(l,k)}$ is absolutely convergent, so that it can be expressed as a double sum. By Lemma 4, $\sum_{k\in K_l}|\alpha^{(l)}_{mnk}| \leqslant b_l$ for all $l \in L$; therefore,

$$\sum_{(l,k)\in K}|\alpha_{mn(l,k)}| = \sum_{l\in L}|s_l|\sum_{k\in K_l}\left|\alpha^{(l)}_{mnk}\right| \tag{B8}$$

$$\leqslant \sum_{l\in L}|s_l b_l| \tag{B9}$$

$$\leqslant b. \tag{B10}$$

Since $b < \infty$ by assumption, the series $\sum_{(l,k)\in K}\alpha_{mn(l,k)}$ is absolutely convergent. We can then decompose it as a double series,

$$\sum_{(l,k)\in K}\alpha_{mn(l,k)} = \sum_{l\in L}s_l\sum_{k\in K_l}\alpha^{(l)}_{mnk} \tag{B11}$$

$$= \sum_{l\in L}s_l A^{(l)}, \tag{B12}$$

showing that condition (a) of Definition 3 is satisfied.

[18] The lowest b is obtained when $W(l)$ is proportional to $|s_l|b_l$.

Define the probability distributions

$$P(n,(l,k)|m) = \begin{cases} W(l)P_l(n,k|m) & \text{if } k \in K_l, \\ 0 & \text{otherwise,} \end{cases} \quad \text{(B13)}$$

$$Q(m,(l,k)|n) = \begin{cases} W(l)Q_l(m,k|n) & \text{if } k \in K_l, \\ 0 & \text{otherwise.} \end{cases} \quad \text{(B14)}$$

We now show that condition (b) holds. Let $m \in \{1,\ldots,M\}$, $n \in \{1,\ldots,N\}$, and $(l,k) \in K$. We need only consider $k \in K_l$ since otherwise $\alpha_{mn(l,k)}$ vanishes:

$$\frac{|\alpha_{mn(l,k)}|}{[P(n,(l,k)|m)]^{1/p}[Q(m,(l,k)|n)]^{1/q}}$$

$$= \frac{\left|s_l \alpha^{(l)}_{mnk}\right|}{[W(l)P_l(n,k|m))]^{1/p}[W(l)Q_l(m,k|n)]^{1/q}} \quad \text{(B15)}$$

$$= \frac{|s_l|}{W(l)} \frac{\left|\alpha^{(l)}_{mnk}\right|}{[P_l(n,k|m)]^{1/p}[Q_l(m,k|n)]^{1/q}} \quad \text{(B16)}$$

$$\leqslant |s_l| b_l / W(l) \quad \text{(B17)}$$

$$\leqslant b. \quad \text{(B18)}$$

Condition (c) requires that the distribution $P(n,(l,k)|m)$ can be sampled from, and $\alpha_{mn(l,k)}/P(n,(l,k)|m)$ and $\alpha_{mn(l,k)}/Q(m,(l,k)|n)$ can be computed, in average time $O(f) = O(f_0 + \sum_l W(l) f_l)$. This can be accomplished as follows.

(i) Draw l according to the distribution $W(l)$ and compute $s_l/W(l)$. This can be done in average time $O(f_0)$.

(ii) Draw n,k according to the distribution $P_l(n,k|m)$ and compute $\alpha^{(l)}_{mnk}/P_l(n,k|m)$ and $\alpha^{(l)}_{mnk}/Q_l(m,k|n)$. This can be done in average time $O(f_l)$.

(iii) The quantities $\alpha_{mn(l,k)}/P(n,(l,k)|m)$ and $\alpha_{mn(l,k)}/Q(m,(l,k)|n)$ can be directly computed from (B7), (B13), and (B14) in time $O(1)$ given the quantities that have been computed in the previous two steps.

The average time needed for a given l is $O(f_0 + f_l)$; therefore, the average time needed given that l is drawn according to $W(l)$ is $O(f) = O(f_0 + \sum_l W(l) f_l)$. Condition (c) is satisfied. Condition (d) follows from a symmetric argument. ■

Proof of Theorem 4(a). This follows directly from Theorem 13. Specifically, apply Theorem 13 with $L = \{A,B\}$, $s_A = s_B = 1$, $W(A) = b_A/(b_A + b_B)$, and $W(B) = b_B/(b_A + b_B)$. Then $b = \max_l\{|s_l|b_l/W(l)\} = b_A + b_B$ and $f = O(1) + \sum_l W(l) f_l = O(\max\{b_A,b_B\})$. ■

Proof of Theorem 4(b). Since A is $\text{EPS}_p(b_A,f_A)$, there are K_A, $\alpha^{(A)}_{lmk}$, $P_A(m,k|l)$, and $Q_A(l,k|m)$ satisfying Definition 3 with $l \in \{1,\ldots,L\}$, $m \in \{1,\ldots,M\}$, and $k \in K_A$. Likewise, since B is $\text{EPS}_p(b_B,f_B)$, there are K_B, $\alpha^{(B)}_{mnk}$, $P_B(n,k|m)$, and $Q_B(m,k|n)$ satisfying Definition 3 with $m \in \{1,\ldots,M\}$, $n \in \{1,\ldots,N\}$, and $k \in K_B$.

Let $K = K_A \times K_B \times \{1,\ldots,M\}$ and

$$\alpha_{ln(k',k'',m)} = \alpha^{(A)}_{lmk'} \alpha^{(B)}_{mnk''}. \quad \text{(B19)}$$

We first show that $\sum_{(k',k'',m)\in K} \alpha_{ln(k',k'',m)}$ is absolutely convergent, so that it can be expressed as a double series. By Lemma 4, $\sum_{k'\in K_A} |\alpha^{(A)}_{lmk'}| \leqslant b_A$ and $\sum_{k''\in K_B} |\alpha^{(B)}_{mnk''}| \leqslant b_B$; therefore,

$$\sum_{(k',k'',m)\in K} |\alpha_{ln(k',k'',m)|}$$

$$= \sum_{m\in\{1,\ldots,M\}} \sum_{k'\in K_A} \left|\alpha^{(A)}_{lmk'}\right| \sum_{k''\in K_B} \left|\alpha^{(B)}_{mnk''}\right| \quad \text{(B20)}$$

$$\leqslant M b_A b_B \quad \text{(B21)}$$

$$\leqslant \infty. \quad \text{(B22)}$$

Being absolutely convergent, $\sum_{(k',k'',m)\in K} \alpha_{ln(k',k'',m)}$ can be expressed as a double series, giving

$$\sum_{(k',k'',m)\in K} \alpha_{ln(k',k'',m)}$$

$$= \sum_{m\in\{1,\ldots,M\}} \sum_{k'\in K_A} \alpha^{(A)}_{lmk'} \sum_{k''\in K_B} \alpha^{(B)}_{mnk''} \quad \text{(B23)}$$

$$= \sum_m A_{lm} B_{mn} \quad \text{(B24)}$$

$$= (AB)_{ln}, \quad \text{(B25)}$$

so condition (a) of Definition 3 is satisfied.

Define the probability distributions

$$P(n,(k',k'',m)|l) = P_A(m,k'|l) P_B(n,k''|m), \quad \text{(B26)}$$

$$Q(l,(k',k'',m)|n) = Q_A(l,k'|m) Q_B(m,k''|n). \quad \text{(B27)}$$

These satisfy condition (b) of Definition 3 since for all l,m,n,k',k'',

$$b_A b_B \geqslant \frac{\left|\alpha^{(A)}_{lmk'}\right|}{P_A(m,k'|l)^{1/p} Q_A(l,k'|m)^{1/q}} \times \frac{\left|\alpha^{(B)}_{mnk''}\right|}{P_B(n,k''|m)^{1/p} Q_B(m,k''|n)^{1/q}} \quad \text{(B28)}$$

$$= \frac{|\alpha_{ln(k',k'',m)|}}{P(n,(k',k'',m)|l)^{1/p} Q(l,(k',k'',m)|n)^{1/q}}. \quad \text{(B29)}$$

Condition (c) requires that it be possible in average time $O(f_A + f_B)$ to sample from the probability distribution $P(n,(k',k'',m)|l)$ and to compute $\frac{\alpha_{ln(k',k'',m)}}{P(n,(k',k'',m)|l)}$ and $\frac{\alpha_{ln(k',k'',m)}}{Q(l,(k',k'',m)|n)}$. This can be accomplished as follows.

(i) Draw m,k' from $P_A(m,k'|l)$ and compute $\frac{\alpha^{(A)}_{lmk'}}{P_A(m,k'|l)}$ and $\frac{\alpha^{(A)}_{lmk'}}{Q_A(l,k'|m)}$. This can be done in average time $O(f_A)$.

(ii) Draw n,k'' from $P_B(n,k''|m)$ and compute $\frac{\alpha^{(B)}_{mnk''}}{P_B(n,k''|m)}$ and $\frac{\alpha^{(B)}_{mnk''}}{Q_B(m,k''|n)}$. This can be done in average time $O(f_B)$.

(iii) Compute

$$\frac{\alpha_{ln(k',k'',m)}}{P(n,(k',k'',m)|l)} = \frac{\alpha^{(A)}_{lmk'}}{P_A(m,k'|l)} \frac{\alpha^{(B)}_{mnk''}}{P_B(n,k''|m)}, \quad \text{(B30)}$$

$$\frac{\alpha_{ln(k',k'',m)}}{Q(l,(k',k'',m)|n)} = \frac{\alpha^{(A)}_{lmk'}}{Q_A(l,k'|m)} \cdot \frac{\alpha^{(B)}_{mnk''}}{Q_B(m,k''|n)}. \quad \text{(B31)}$$

This can be done in time $O(1)$ since the factors on the right-hand sides of these expressions have already been computed in the previous two steps.

So condition (c) is satisfied. Condition (d) follows from a symmetric argument. ■

Proof of Theorem 4(c). Let A be a square matrix that is $\mathrm{EPS}_p(b,f)$. We show that e^A is $\mathrm{EPS}_p(e^b,bf)$.

This follows from applying Theorems 13 and 4(b) to $e^A = \sum_{j=0}^{\infty} A^j/j!$. Specifically, let $L = \{0,1,\dots\}$, $A^{(l)} = A^l$, $s_l = 1/l!$, and $W(l) = b^l/(l!e^b)$. By repeated application of Theorem 4(b), $A^{(l)}$ is $\mathrm{EPS}_p(b^l,lf)$. Assume for now that $W(l)$ can be sampled in average time $O(b)$. Then by Theorem 13, $e^A = \sum_{j=0}^{\infty} A^j/j!$ is $\mathrm{EPS}_p(b',f')$ with $b' = \max_l\{|s_l|b_l/W(l)\} = e^b$ and

$$f' = b + \sum_{l=0}^{\infty} W(l) f_l \tag{B32}$$

$$= b + \sum_{l=0}^{\infty} \frac{lfb^l}{l!e^b} \tag{B33}$$

$$= b + \frac{bf}{e^b} \sum_{l=1}^{\infty} \frac{b^{l-1}}{(l-1)!} \tag{B34}$$

$$= b + bf \tag{B35}$$

$$= O(bf). \tag{B36}$$

It remains only to show that $W(l)$ can be sampled in time $O(b)$. The procedure is as follows. Flip a weighted coin that lands heads up with probability $W(0)$, and if it lands heads up take $l = 0$. This can be done in time $O(1)$. If the coin landed tails up, then flip another coin that lands heads up with probability $W(1)/[1 - W(0)]$, and if it lands heads up, take $l = 1$. Continue, in each iteration flipping a coin that lands heads up with probability $W(l)/[1 - \sum_{j=0}^{l-1} W(j)]$. Each iteration requires computing $W(l)/[1 - \sum_{j=0}^{l-1} W(j)]$, which in turn requires computing $W(l)$ and updating the partial sum with the previous $W(l-1)$. This can be done in $O(1)$ time. The expected number of iterations is $\sum_l l W(l) = b$. Therefore, this sampling algorithm takes average time b. ■

Proof of Lemma 1. Since σ is $\mathrm{EHT}_p(b_\sigma, f_\sigma)$, there are $\alpha^{(\sigma)}_{nmk}$, $P_\sigma(m,k)$, and $Q_\sigma(n,k)$, with $k \in K_\sigma$ satisfying Definition 4 (note that m and n have been swapped since σ is an $N \times M$ operator). Similarly, since A is $\mathrm{EPS}_p(b_A, f_A)$ there are $\alpha^{(A)}_{mnk'}$, $P_A(n,k'|m)$, and $Q_A(m,k'|n)$, with $k' \in K_A$ satisfying Definition 3.

We have

$$\mathrm{Tr}(A\sigma) = \sum_{mn} A_{mn}\sigma_{nm} \tag{B37}$$

$$= \sum_{mnkk'} \alpha^{(A)}_{mnk'}\alpha^{(\sigma)}_{nmk}. \tag{B38}$$

Define the probability distribution

$$R(m,n,k,k') = \frac{1}{p} P_\sigma(m,k) P_A(n,k'|m) + \frac{1}{q} Q_\sigma(n,k) Q_A(m,k'|n). \tag{B39}$$

By the inequality of arithmetic and geometric means,

$$R(m,n,k,k') \geqslant [P_\sigma(m,k)P_A(n,k'|m)]^{1/p}[Q_\sigma(n,k)Q_A(m,k'|n)]^{1/q}. \tag{B40}$$

Setting $V(m,n,k,k') = \alpha^{(A)}_{mnk'}\alpha^{(\sigma)}_{nmk}$, we get the bound

$$b_{\max} := \max_{mnkk'} \left\{ \frac{|V(m,n,k,k')|}{R(m,n,k,k')} \right\} \tag{B41}$$

$$\leqslant \max_{mnkk'} \left\{ \frac{\left|\alpha^{(A)}_{mnk'}\alpha^{(\sigma)}_{nmk}\right|}{[P_\sigma(m,k)P_A(n,k'|m)]^{1/p}[Q_\sigma(n,k)Q_A(m,k'|n)]^{1/q}} \right\} \tag{B42}$$

$$\leqslant \max_{mnk'} \left\{ \frac{\left|\alpha^{(A)}_{mnk'}\right|}{P_A(n,k'|m)^{1/p} Q_A(m,k'|n)]^{1/q}} \right\} \max_{mnk} \left\{ \frac{\left|\alpha^{(\sigma)}_{nmk}\right|}{P_\sigma(m,k)^{1/p} Q_\sigma(n,k)^{1/q}} \right\} \tag{B43}$$

$$\leqslant b_A b_\sigma. \tag{B44}$$

By Corollary 1, the sum (B38) can be estimated at the cost of drawing $O(\log_2(\delta^{-1})\epsilon^{-2}b_\sigma^2 b_A^2)$ samples from $R(m,n,k,k')$ and evaluating the corresponding $V(m,n,k,k')/R(m,n,k,k')$. Each of these samples can be computed in average time $O(f_\sigma + f_A)$ as follows.

(i) Flip a weighted coin that lands heads up with probability $1/p$.

(ii) If it lands heads up, sample m,k according to $P_\sigma(m,k)$ and then sample n,k' according to $P_A(n,k'|m)$.

(iii) If it lands tails up, sample n,k according to $Q_\sigma(n,k)$ and then sample m,k' according to $Q_A(m,k'|n)$.

(iv) The previous steps produce a sample according to $R(m,n,k,k')$ and can be accomplished in time $O(f_\sigma + f_A)$ by conditions (c) and (d) of Definition 3 and (c) and (d) of Definition 4, with the side effect of producing values $\alpha^{(\sigma)}_{nmk}/P_\sigma(m,k)$, $\alpha^{(A)}_{mnk'}/P_A(n,k'|m)$, $\alpha^{(\sigma)}_{nmk}/Q_\sigma(n,k)$, and $\alpha^{(A)}_{mnk'}/Q_A(m,k'|n)$.

(v) These values can be used to compute $V(m,n,k,k')/R(m,n,k,k')$ since

$$\frac{V(m,n,k,k')}{R(m,n,k,k')} = \frac{\alpha^{(A)}_{mnk'}\alpha^{(\sigma)}_{nmk}}{R(m,n,k,k')} \tag{B45}$$

$$= \left[\frac{1}{p} \frac{P_A(n,k'|m)}{\alpha^{(A)}_{mnk'}} \frac{P_\sigma(m,k)}{\alpha^{(\sigma)}_{nmk}} + \frac{1}{q} \frac{Q_A(m,k'|n)}{\alpha^{(A)}_{mnk'}} \frac{Q_\sigma(n,k)}{\alpha^{(\sigma)}_{nmk}} \right]^{-1} \tag{B46}$$

Therefore, the sum (B38) can be estimated in average time $O(\log_2(\delta^{-1})\epsilon^{-2}b_\sigma^2 b_A^2(f_\sigma + f_A))$. ■

APPENDIX C: PROOFS FOR SEC. V

In Sec. V several matrices and classes of matrices were claimed to be $\mathrm{EPS}_2(b,f)$ or $\mathrm{EPS}_p(b,f)$ for small values of b and f. In this Appendix we provide proofs for these claims.

We first prove that the efficiently computable sparse (ECS) matrices from [15] (definition reproduced below) are $\mathrm{EPS}_p(\mathrm{polylog}_2(N),\mathrm{polylog}_2(N))$. This covers a rather large class of matrices including permutation matrices, Pauli matrices, controlled phase matrices, and arbitrary unitaries on a constant number of qudits. The original definition from [15] was in terms of qubits, but we adapt it to systems of arbitrary dimension.

Definition 9. ECS. A matrix A is efficiently computable sparse (ECS) if

(a) each row and column of A has at most $\mathrm{polylog}_2(N)$ nonzero entries;

(b) for any given row index m, it is possible in $\mathrm{polylog}_2(N)$ time to list the indices of the nonzero entries in that row, $\{n : A_{mn} \neq 0\}$, and to compute their values A_{mn};

(c) for any given column index n, it is possible in $\mathrm{polylog}_2(N)$ time to list the indices of the nonzero entries in that column, $\{m : A_{mn} \neq 0\}$, and to compute their values A_{mn}.

Theorem 14. ECS is EPS. Let A be an ECS matrix satisfying $\max_{mn}\{|A_{mn}|\} = \mathrm{polylog}_2(N)$. Unitaries and Hermitian matrices whose eigenvalues are in the $[-1,1]$ range satisfy this bound. Then A is $\mathrm{EPS}_p(\mathrm{polylog}_2(N),\mathrm{polylog}_2(N))$ for any $p \in [1,\infty]$.

Proof. Theorem 5 is applicable here with $f = \mathrm{polylog}_2(N)$. Let $P(n|m)$ and $Q(m|n)$ be the probability distributions defined in (67). Given any m and n, the value A_{mn} can be computed in $\mathrm{polylog}_2(N)$ time. Since each row and column contains $\mathrm{polylog}_2(N)$ nonzero entries, which can be enumerated and computed in $\mathrm{polylog}_2(N)$ time, the sums $\sum_{n'}|A_{mn'}|$ and $\sum_{m'}|A_{m'n}|$ can be computed in $\mathrm{polylog}_2(N)$ time. Thus, condition (c) of Theorem 5 is satisfied.

For any given m, the distribution $P(n|m)$ has support of size $\mathrm{polylog}_2(N)$, the indices of which can be enumerated in $\mathrm{polylog}_2(N)$ time, and each individual probability can be computed in time $\mathrm{polylog}_2(N)$. Therefore, this distribution can be sampled from in time $\mathrm{polylog}_2(N)$. Similarly for $Q(m|n)$, so conditions (a) and (b) of Theorem 5 are satisfied and A is $\mathrm{EPS}_p(\|A\|_\infty^{1/p}\|A\|_1^{1/q},\mathrm{polylog}_2(N))$. Each row and column of A has at most $\mathrm{polylog}_2(N)$ nonzero entries, each bounded by $\max_{mn}\{|A_{mn}|\} = \mathrm{polylog}_2(N)$. It follows that $\|A\|_\infty = \mathrm{polylog}_2(N)$ and $\|A\|_1 = \mathrm{polylog}_2(N)$, giving $\|A\|_\infty^{1/p}\|A\|_1^{1/q} = \mathrm{polylog}_2(N)$. ■

A block diagonal matrix is $\mathrm{EPS}_p(b,f)$ if each of its blocks is $\mathrm{EPS}_p(b,f)$. This is rather powerful in that it can be used to show the EPS property for operations on subsystems, for controlled unitaries, and for some rather exotic projectors. This will be the subject of the following theorem and corollaries.

Theorem 15. Block diagonal. For $r \in \{1,\ldots,R\}$, let $A^{(r)}$ be an $\mathrm{EPS}_p(b_r,f)$ matrix of dimension $M_r \times N_r$. Let A be the block diagonal matrix $A = \oplus_r A^{(r)}$ of dimension $\sum_r M_r \times \sum_r N_r$. Suppose that it is possible in time $O(f)$ to convert between row and column indices of A and the corresponding block indices [i.e., $m' \to (r,m)$ and $n' \to (s,n)$ and their inverse maps, with $A_{m'n'} = \delta_{rs}A^{(r)}_{mn}$]. Then A is $\mathrm{EPS}_p(\max_r\{b_r\},f)$.

Proof. Since $A^{(r)}$ is $\mathrm{EPS}_p(b_r,f)$ for each r, there are K_r, $\alpha^{(r)}_{mnk}$, $P_r(n,k|m)$, and $Q_r(m,k|n)$ satisfying Definition 3, with $m \in \{1,\ldots,M_r\}$, $n \in \{1,\ldots,N_r\}$, and $k \in K_r$. Since we can convert between row and column indices of A and the corresponding block indices in time $O(f)$, go ahead and label the indices of A using block indices: $A_{(r,m),(s,n)} = \delta_{rs}A^{(r)}_{mn}$. Define $K = \cup_r K_r$ and

$$\alpha_{(r,m),(s,n),k} = \begin{cases} \alpha^{(r)}_{mnk} & \text{if} \quad r = s \quad \text{and} \quad k \in K_r, \\ 0 & \text{otherwise.} \end{cases} \tag{C1}$$

This satisfies condition (a) of Definition 3 since

$$\sum_{k\in K}\alpha_{(r,m),(s,n),k} = \delta_{rs}\sum_{k\in K_r}\alpha^{(r)}_{mnk} \tag{C2}$$

$$= \delta_{rs}A^{(r)}_{mn} \tag{C3}$$

$$= A_{(r,m),(s,n)}. \tag{C4}$$

Define the probability distributions

$$P((s,n),k|(r,m)) = \delta_{rs}P_r(n,k|m), \tag{C5}$$

$$Q((r,m),k|(s,n)) = \delta_{rs}Q_s(m,k|n). \tag{C6}$$

That $\alpha_{(r,m),(s,n),k}$, $P((s,n),k|(r,m))$, and $Q((r,m),k|(s,n))$ satisfy conditions (c) and (d) of Definition 3 directly follows from the fact that $\alpha^{(r)}_{mnk}$, $P_r(n,k|m)$, and $Q_s(m,k|n)$ satisfy conditions (c) and (d) for all r. Condition (b) is satisfied as well, since

$$\max_{(r,m),(s,n),k}\left\{\frac{|\alpha_{(r,m),(s,n),k}|}{P((s,n),k|(r,m))^{1/p}Q((r,m),k|(s,n))^{1/q}}\right\}$$

$$= \max_r\max_{mnk}\left\{\frac{|\alpha^{(r)}_{mnk}|}{P_r(n,k|m)^{1/p}Q_r(m,k|n)^{1/q}}\right\} \tag{C7}$$

$$\leqslant \max_r\{b_r\}. \tag{C8}$$

■

Corollary 3. For $r \in \{1,\ldots,R\}$, let $A^{(r)}$ be matrices on a space of dimension N. Suppose that each $A^{(r)}$ is $\mathrm{EPS}_p(b,f)$ with $f = \Omega(\log_2^2(N))$. Then $A = \sum_{r=1}^R |r\rangle\langle r| \otimes A^{(r)}$, where the $|r\rangle$ are computational basis states, is $\mathrm{EPS}_p(b,f)$.

Proof. This is essentially a restatement of Theorem 15 for the case where all the $A^{(r)}$ are the same size. We require $f = \Omega(\log_2^2(N))$ because converting row or column indices of A to indices of the blocks (as required for application of Theorem 15) requires the operation of computing the quotient and remainder of division by N. The $f = \Omega(\log_2^2(N))$ requirement can be dropped if one is dealing with query complexity rather than computational complexity. ■

Corollary 4. Let U denote a unitary matrix on n qubits whose rows are CT states (e.g., the Fourier transform). Let $g : \{0,\ldots,2^{n-1}\} \to \{0,\ldots,2^{n-1}\}$ be a poly(n) time computable

function. Then the projector $\sum_{x=0}^{2^{n-1}} |x\rangle\langle x| \otimes U^\dagger|g(x)\rangle\langle g(x)|U$ is $\mathrm{EPS}_s(1,\mathrm{poly}(n))$. This projector corresponds to measuring half of the system in the computational basis to get measurement result x, measuring the other half of the system in the basis determined by U to get y, and returning true if $y = g(x)$. The measurement depicted in Fig. 1 is of this form.

Proof. Apply Corollary 3 with $A^{(x)} = U^\dagger|g(x)\rangle\langle g(x)|U$. $U^\dagger|g(x)\rangle$ is a CT state, so by Theorem 6 $A^{(x)}$ is $\mathrm{EHT}_2(1,\mathrm{poly}(n))$ and therefore also $\mathrm{EPS}_2(1,\mathrm{poly}(n))$. ■

Corollary 5. Let I_{M_1} and I_{M_2} denote the identity operator on spaces of dimensions M_1 and M_2. Let A be an $\mathrm{EPS}_p(b,f)$ matrix of dimension $N_1 \times N_2$ with $f = \Omega(\log_2^2(M_1M_2N_1N_2))$. Then $I_{M_1} \otimes A \otimes I_{M_2}$ is $\mathrm{EPS}_p(b,f)$. This somewhat trivial result is important in that it allows the matrix to act on subsystems of the full state.

Proof. Apply Theorem 15 with all of the $A^{(r)}$ blocks being equal. We require $f = \Omega(\log_2^2(M_1M_2N_1N_2))$ in order to allow converting row or column indices of $I_{M_1} \otimes A \otimes I_{M_2}$ to indices of A in time $O(f)$. ■

We now turn to the Grover reflection operation. We show this operator to be $\mathrm{EPS}_2(3,\log_2(N))$. Since a unitary operator incurs a time expense of b^4 as per (69), each round of Grover's algorithm multiplies the simulation time by $3^4 = 81$. This time is constant in the number of qubits, but is exponential in the number of rounds. Our technique is therefore perfectly capable of simulating a small number of Grover reflections placed anywhere in a circuit, but would perform very poorly, $\exp[\Theta(\sqrt{N})]$ time, if applied to the $\Theta(\sqrt{N})$ rounds required by Grover's algorithm.

Theorem 16. Let $|+\rangle = N^{-1/2}\sum_{i=0}^{N-1}|i\rangle$. The Grover reflection $I - 2|+\rangle\langle+|$ is $\mathrm{EPS}_2(3,\log_2(N))$.

Proof. Let δ_{mn} be the Kronecker δ. The identity operator can be seen to be $\mathrm{EPS}_p(1,\log_2(N))$, for any p but in particular $p = 2$, by simple inspection of Definition 3 with $K = \{0\}$ and $\alpha_{mnk} = P(n,k|m) = Q(m,k|n) = \delta_{mn}$. Note that we must take $f = \log_2(N)$ rather than $f = 1$ since it takes $\Omega(\log_2(N))$ time to even write the indices m and n, which are $\log_2(N)$ bits long.

By Theorem 6, the projector $|+\rangle\langle+|$ is $\mathrm{EHT}_2(1,\log_2(N))$, and therefore also $\mathrm{EPS}_2(1,\log_2(N))$. By Theorem 3 the operator $(-2)|+\rangle\langle+|$ is $\mathrm{EPS}_2(2,\log_2(N))$ and by Theorem 4(a) the operator $I - 2|+\rangle\langle+|$ is $\mathrm{EPS}_2(3,\log_2(N))$. One cannot do much better than $b = 3$ since $\|I - 2(|+\rangle\langle+|)\|_2 \to 3$ as $N \to \infty$. ■

Next we show that the Haar wavelet transform on n qubits, denoted G_n, is $\mathrm{EPS}_2(\sqrt{n+1},n)$. This is the lowest possible value of b, since $\|G_n\|_2 = \sqrt{n+1}$.

Definition 10. The *Haar wavelet transform* on n qubits is defined to be

$$G_n = (|0\rangle\langle+|)^{\otimes n} + \sum_{m=0}^{n-1}(|0\rangle\langle+|)^{\otimes m} \otimes |1\rangle\langle-| \otimes I^{\otimes n-m-1}. \tag{C9}$$

Note that there are other conventions that differ from this by a permutation in the computational basis. Such permutations do not affect whether the Haar transform is $\mathrm{EPS}_2(\sqrt{n+1},n)$.

As an example, the Haar transform on three qubits is implemented by the circuit depicted in Fig. 4 and in the computational basis takes the form

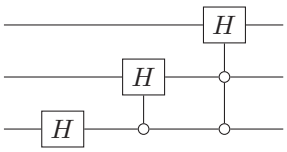

FIG. 4. This circuit implements the Haar transform of Definition 10, on three qubits [39]. The gates in this circuit are controlled-Hadamard gates, and the open circles denote that the Hadamard gates are active when all of the controls are in the $|0\rangle$ state.

$$G_3 = \begin{bmatrix} \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} & \frac{-1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{-1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{-1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{-1}{\sqrt{8}} \\ \frac{1}{\sqrt{4}} & 0 & \frac{-1}{\sqrt{4}} & 0 & \frac{1}{\sqrt{4}} & 0 & \frac{-1}{\sqrt{4}} & 0 \\ 0 & \frac{1}{\sqrt{4}} & 0 & \frac{-1}{\sqrt{4}} & 0 & \frac{1}{\sqrt{4}} & 0 & \frac{-1}{\sqrt{4}} \\ \frac{1}{\sqrt{2}} & 0 & 0 & 0 & \frac{-1}{\sqrt{2}} & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & \frac{-1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & \frac{-1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & \frac{-1}{\sqrt{2}} \end{bmatrix}. \tag{C10}$$

Theorem 17. The Haar transform on n qubits is $\mathrm{EPS}_2(\sqrt{n+1},n)$.

Proof. Since we are dealing with spaces of dimension 2^n, made of qubits, it will be convenient to index the space using bit strings $\boldsymbol{x},\boldsymbol{y} \in \{0,1\}^n$. We denote the corresponding basis vectors using the notation $|\boldsymbol{x}\rangle = |x_0\rangle \otimes \cdots \otimes |x_{n-1}\rangle$. To avoid notational confusion regarding subscripts, define $A = G_n$. Then A_{xy} refers to the matrix element $\langle\boldsymbol{x}|G_n|\boldsymbol{y}\rangle$.

Take $K = \{0\}$ (i.e., do not make use of the index k), and set $\alpha_{xyk} = A_{xy}$. This satisfies condition (a) of Definition 3 trivially. Take the probability distributions $P(\boldsymbol{y}|\boldsymbol{x})$ and $Q(\boldsymbol{x}|\boldsymbol{y})$ to be uniform over the nonzero elements of the given row or column of A_{xy}. Despite the apparent simplicity of this choice, analysis will be tedious due to the somewhat complicated definition of A. These probability distributions can be expressed as follows:

$$P(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{2^n}[\boldsymbol{x} = 0] + \sum_{m=0}^{n-1}\frac{1}{2^{m+1}}\left(\prod_{i=0}^{m-1}[x_i = 0]\right) \times [x_m = 1]\left(\prod_{i=m+1}^{n-1}[y_i = x_i]\right) \tag{C11}$$

$$Q(\boldsymbol{x}|\boldsymbol{y}) = \frac{1}{n+1}\left\{[\boldsymbol{x} = 0] + \sum_{m=0}^{n-1}\left(\prod_{i=0}^{m-1}[x_i = 0]\right) \times [x_m = 1]\left(\prod_{i=m+1}^{n-1}[x_i = y_i]\right)\right\} \tag{C12}$$

These can be sampled from in time $O(n)$. Consider first $P(\boldsymbol{y}|\boldsymbol{x})$. Given an $\boldsymbol{x}$, only a single one of the $n+1$ terms of (C11) does not vanish, and this term can be identified in time $O(n)$, by searching for the smallest (if any) m for which $x_m = 1$. The nonvanishing term defines the value of y_i for some of the i and gives a uniform distribution for each of the remaining y_i. For $Q(\boldsymbol{x}|\boldsymbol{y})$, each of the $n+1$ terms of (C12) is nonvanishing for a single value of $\boldsymbol{x}$, and each occurs with equal probability. Therefore, sampling from $Q(\boldsymbol{x}|\boldsymbol{y})$ is accomplished by drawing from a uniform distribution over $n+1$ possibilities.

To satisfy conditions (c) and (d) of Definition 3 we must also show that $A_{xy}/P(\boldsymbol{y}|\boldsymbol{x})$ and $A_{xy}/Q(\boldsymbol{x}|\boldsymbol{y})$ can be computed in time $O(n)$. We begin by writing an expression for A_{xy}. In the equations below, square brackets denote the Iverson bracket, which takes a value of 1 if the enclosed expression is true and 0 otherwise:

$$A_{xy} = \langle \boldsymbol{x}| \left((|0\rangle\langle +|)^{\otimes n} + \sum_{m=0}^{n-1} (|0\rangle\langle +|)^{\otimes m} \otimes |1\rangle\langle -| \otimes I^{\otimes n-m-1} \right) |\boldsymbol{y}\rangle \tag{C13}$$

$$= \frac{1}{\sqrt{2^n}}[\boldsymbol{x}=0] + \sum_{m=0}^{n-1} \frac{(-1)^{y_m}}{\sqrt{2^{m+1}}} \left(\prod_{i=0}^{m-1} [x_i = 0] \right) \times [x_m = 1] \left(\prod_{i=m+1}^{n-1} [x_i = y_i] \right). \tag{C14}$$

Since only a single term for each of (C11), (C12), and (C14) is nonvanishing for each given $\boldsymbol{x},\boldsymbol{y}$ pair, we can divide these equations term by term to get

$$\frac{A_{xy}}{P(\boldsymbol{y}|\boldsymbol{x})} = \sqrt{2^n}[\boldsymbol{x}=0] + \sum_{m=0}^{n-1} (-1)^{y_m} \sqrt{2^{m+1}} \left(\prod_{i=0}^{m-1} [x_i = 0] \right) \times [x_m = 1] \left(\prod_{i=m+1}^{n-1} [x_i = y_i] \right), \tag{C15}$$

$$\frac{A_{xy}}{Q(\boldsymbol{x}|\boldsymbol{y})} = (n+1) \left\{ \frac{1}{\sqrt{2^n}}[\boldsymbol{x}=0] + \sum_{m=0}^{n-1} \frac{(-1)^{y_m}}{\sqrt{2^{m+1}}} \left(\prod_{i=0}^{m-1} [x_i = 0] \right) \times [x_m = 1] \left(\prod_{i=m+1}^{n-1} [x_i = y_i] \right) \right\}. \tag{C16}$$

At most a single term of these expressions is nonvanishing for each given $\boldsymbol{x},\boldsymbol{y}$ pair, and this term can be identified in time $O(n)$ by searching for the smallest (if any) m for which $x_m = 1$. The value of nonvanishing terms is of the form $\pm\sqrt{2^s}$ or $\pm(n+1)/\sqrt{2^s}$ for some s, and this can be computed in $O(1)$ time.

That condition (b) of Definition 3 is satisfied is checked directly,

$$\max_{xy} \left\{ \frac{|A_{xy}|}{P(\boldsymbol{y}|\boldsymbol{x})^{1/2} Q(\boldsymbol{x}|\boldsymbol{y})^{1/2}} \right\} = \max_{xy} \left\{ \left[\frac{|A_{xy}|}{P(\boldsymbol{y}|\boldsymbol{x})} \frac{|A_{xy}|}{Q(\boldsymbol{x}|\boldsymbol{y})} \right]^{1/2} \right\} \tag{C17}$$

$$= \max_{xy} \left\{ (n+1)^{1/2} \right\} \tag{C18}$$

$$= \sqrt{n+1}, \tag{C19}$$

where (C18) follows from the fact that only a single term from each of (C15) and (C16) is nonvanishing, so they can be multiplied term by term. ■

[1] G. Vidal, Phys. Rev. Lett. **91**, 147902 (2003).
[2] R. Jozsa and N. Linden, Proc. R. Soc. London, Ser. A **459**, 2011 (2003).
[3] B. Eastin, Bull. Am. Phys. Soc. **56**, BAPS.2011.MAR.D29.10 (2011).
[4] D. Gottesman, in *Group22: Proceedings of the XXII International Colloquium on Group Theoretical Methods in Physics*, edited by S. P. Corney, R. Delbourgo, and P. D. Jarvis (International Press, Cambridge, MA, 1999), pp. 32–43, arXiv:quant-ph/9807006.
[5] L. G. Valiant, in *Proceedings of the Thirty-third Annual ACM Symposium on Theory of Computing, STOC '01* (ACM, New York, 2001), pp. 114–123.
[6] I. L. Markov and Y. Shi, SIAM J. Comput. **38**, 963 (2008).
[7] R. Jozsa, arXiv:quant-ph/0603163.
[8] V. Veitch, C. Ferrie, D. Gross, and J. Emerson, New J. Phys. **14**, 113011 (2012).
[9] V. Veitch, N. Wiebe, C. Ferrie, and J. Emerson, New J. Phys. **15**, 013037 (2013).
[10] A. Mari and J. Eisert, Phys. Rev. Lett. **109**, 230503 (2012).
[11] J. F. Fitzsimons, E. G. Rieffel, and V. Scarani, in *Computation for Humanity: Information Technology to Advance Society*, edited by J. Zander and P. J. Mosterman (CRC Press, Boca Raton, 2013), Chap. 11.
[12] C. H. Bennett, Phys. Today **48**, 24 (1995).
[13] L. Fortnow, Theor. Comput. Sci. **292**, 597 (2003).
[14] S. Lloyd, Phys. Rev. A **61**, 010301 (1999).
[15] M. Van den Nest, Quantum Inf. Comput. **11**, 0784 (2011).
[16] D. Braun and B. Georgeot, Phys. Rev. A **73**, 022314 (2006).
[17] M. V. d. Nest, Phys. Rev. Lett. **110**, 060504 (2013).
[18] B. Terhal and D. DiVincenzo, Quantum Inf. Comput. **4**, 134 (2004).

[19] W. Hoeffding, J. Am. Stat. Assoc. **58**, 13 (1963).
[20] D. R. Simon, in *Proceedings of the 35th Annual Symposium on Foundations of Computer Science, 1994* (IEEE, Piscataway, NJ, 1994), pp. 116–123.
[21] D. Aharonov, Z. Landau, and J. Makowsky, arXiv:quant-ph/0611156.
[22] N. Yoran and A. J. Short, Phys. Rev. A **76**, 060302 (2007).
[23] A. M. Childs, R. Cleve, E. Deotto, E. Farhi, S. Gutmann, and D. A. Spielman, in *Proceedings of the 35th Annual ACM Symposium on Theory of Computing, STOC '03* (ACM, New York, 2003), pp. 59–68.
[24] V. Veitch, S. A. H. Mousavian, D. Gottesman, and J. Emerson, New J. Phys. **16**, 013009 (2014).
[25] O. Regev and B. Klartag, in *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing, STOC '11* (ACM, New York, 2011), pp. 31–40.
[26] E. Kushilevitz and N. Nisan, *Communication Complexity* (Cambridge University Press, Cambridge, UK, 2006).
[27] A. Montanaro, Quantum Inf. Comput. **11**, 0574 (2011).
[28] I. Kremer, N. Nisan, and D. Ron, in *Proceedings of the 27th Annual ACM Symposium on Theory of Computing, STOC '95* (ACM, New York, 1995) pp. 596–605.
[29] M. Gell-Mann and J. B. Hartle, Phys. Rev. D **47**, 3345 (1993).
[30] R. B. Griffiths, *Consistent Quantum Theory* (Cambridge University Press, Cambridge, UK, 2003).
[31] F. G. S. L. Brandao and M. Horodecki, Quantum Inf. Comput. **13**, 0901 (2013).
[32] S. Aaronson, in *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC '10* (ACM, New York, 2010), pp. 141–150.
[33] R. Mathias, Linear Algebra Appl. **139**, 269 (1990).
[34] D. W. Boyd, Linear Algebra Appl. **9**, 95 (1974).
[35] A. Bhaskara and A. Vijayaraghavan, in *Proceedings of the 22nd Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '11* (SIAM, Philadelphia, 2011), pp. 497–511.
[36] N. L. Carothers, *A Short Course on Banach Space Theory (London Mathematical Society Student Texts)* (Cambridge University Press, Cambridge, 2004).
[37] J. Armstrong, *The Extremal Case of Hölders Inequality* (2010), accessed: 2012/09/02, https://unapologetic.wordpress.com/2010/09/01/the-extremal-case-of-holders-inequality.
[38] R. A. Horn and C. R. Johnson, *Matrix Analysis*, reprint ed. (Cambridge University Press, Cambridge, UK, 1990).
[39] P. Høyer, arXiv:quant-ph/9702028.