On nonlinear transformations in quantum computation

Zoë Holmes,* Nolan Coble,† Andrew T. Sornborger, and Yiğit Subaşı‡
Information Sciences, Los Alamos National Laboratory, Los Alamos, NM, USA.
(Dated: December 24, 2021)

While quantum computers are naturally well-suited to implementing linear operations, it is less clear how to implement nonlinear operations on quantum computers. However, nonlinear subroutines may prove key to a range of applications of quantum computing from solving nonlinear equations to data processing and quantum machine learning. Here we develop algorithms for implementing nonlinear transformations of input quantum states. Our algorithms are framed around the concept of a *weighted state*, a mathematical entity describing the output of an operational procedure involving both quantum circuits and classical post-processing.

I. INTRODUCTION

Quantum computers are naturally adept at performing linear operations because quantum mechanics is inherently linear. That is, the time evolution of a quantum system is governed by the Schrödinger equation, a linear equation. Or, equivalently, quantum states evolve under unitary operations which are necessarily linear. However, to exploit the full potential of quantum computing, we need to be able to also twist the arm of quantum devices into implementing nonlinear operations.

Nonlinear subroutines are likely to play a key role in a range of quantum algorithms. For example, the ability to efficiently implement nonlinear operations would open up new methods for solving nonlinear equations on quantum hardware, with applications in areas from fluid dynamics to finance. Alternatively, nonlinear subroutines could prove valuable for developing new techniques for error mitigation by providing a means of amplifying a signal in the presence of background noise. Finally, there is much excitement currently about the potential of quantum neural networks. However, classical neural networks inherit much of their power from the use of nonlinear activation functions. Replicating this on quantum hardware necessitates the ability to implement nonlinear quantum operations.

While quantum mechanics is fundamentally linear, quantum systems often appear to evolve nonlinearly. These apparent nonlinearities are typically induced through measurements and coarsegraining. In the context of quantum computing, in addition to these tools, nonlinear effects can also be introduced using classical post-processing and by collectively manipulating multiple copies of a given input state.

There is a growing body of research into developing new methods for introducing nonlinearities into quantum algorithms. In the context of quantum machine learning, convolutional [1] or dissipative [2] quantum neural networks, that disregard qubits as the network grows, have been proposed. While these methods do introduce nonlinearities, the exact form of the nonlinearity is not readily controllable. Algorithms for computing specific nonlinear functions of the elements of a quantum state have been proposed in the context of solving nonlinear equations [3] and studying chaotic systems [4, 5]. In this paper we tackle the more general problem of implementing nonlinear transformations of quantum states. This problem has been previously explored using block encoding methodologies [6, 7].

In this paper we introduce a new approach for preparing nonlinear functions of quantum states, using what we call *weighted states*. These are matrices, describing the output of a quantum instrument, that act like density operators but need not be normalized nor Hermitian. The notion of weighted states is a generalization that includes conditional and marginal states as special cases and shares similarities with virtual state distillation [8] and state broadcasting [9–11] in the sense that these entities allow one to reconstruct expectation values without explicitly preparing the state in a conventional sense. We show that through an appropriate choice of quantum instrument and its inputs it is possible to construct weighted states corresponding to a nonlinear transformation of a density matrix. In particular, our algorithms may be used to implement arbitrary polynomials of the amplitudes of a set of pure input states.

The outline of the paper is as follows. In Subsection II A we formally define the functionals of quantum states that we aim to implement and in Subsection II B we define the notion of weighted states. In Subsection II C we present an algorithm to implement the Hadamard product of states, in Subsection II D we present an algorithm to implement the generalized transpose of an input density operator, and in Subsection II E we describe an algorithm that can be used to prepare a linear combination of products of density matrices. We use these algorithms to show how one can implement more general polynomial functions of input quantum states. In Section III we analyze the sampling complexity associated with the weighted state approach. We conclude in Section IV.

* zholmes@lanl.gov
† ncoble@terpmail.umd.edu
‡ ysubasi@lanl.gov

II. WEIGHTED STATES

A. Statement of problem

Given a set of density operators as input $\rho^{\text{in}} = \rho^{(0)} \otimes \cdots \otimes \rho^{(K-1)}$, we are interested in preparing weighted states (to be explained in detail later) of the form

$$\rho^{\text{in}} \to \tilde{\tau} = \sum_{i,j=0}^{d_{\text{out}}-1} f(\boldsymbol{v}_{\rho^{\text{in}}})_{i,j} \, |i\rangle\langle j| \, , \tag{1}$$

where $\boldsymbol{v}_{\rho^{\text{in}}}$ is the vector containing all entries of the input density operators in the computational basis, i.e. $\boldsymbol{v}_{\rho^{\text{in}}} = (\rho^{(0)}_{0,0}, \rho^{(0)}_{0,1}, ..., \rho^{(K-1)}_{d,d})$ with $\rho^{(k)}_{i,j} = \langle i|\rho^{(k)}|j\rangle$, and $f : \mathbb{C}^{kd^2} \to \mathbb{C}^{d^2_{\text{out}}}$ is a multilinear function of its arguments. Here d is the Hilbert space dimension[1] of the individual inputs $\rho^{(k)}$ and d_{out} is the Hilbert space dimensions of the output system $\tilde{\tau}$.

Crucially, if ρ^{in} includes multiple copies of the same state ρ, and we consider f to be a function of the *unique* input states, then the transformation functions will in general be nonlinear in the elements of ρ. For example, in the case where the input states are all identical, i.e. $\rho_{\text{in}} = \rho^{\otimes K}$, we can consider the transformation

$$\rho \to \tilde{\tau} = \sum_{i,j=0}^{d_{\text{out}}-1} g(\boldsymbol{u}_{\rho})_{i,j} \, |i\rangle\langle j| \, , \tag{2}$$

where $\boldsymbol{u}_{\rho} = (\rho_{0,0}, \rho_{0,1}, ..., \rho_{d,d})$ is now the vector containing all entries of ρ in the computational basis, and the transformation functions $g : \mathbb{C}^{d^2} \to \mathbb{C}^{d^2_{\text{out}}}$ are in general nonlinear.

Note that this operation may depend on the chosen computational basis. In general, $\tilde{\tau}$ is not normalized, positive definite, and might not even be Hermitian. We use tilde to indicate such quantum objects that are not proper quantum states that exist on a register of a quantum computer.

A special case is pure states. In this case we assume as input a set of pure states $|\psi^{\text{in}}\rangle = |\psi^{(0)}\rangle \otimes \cdots \otimes |\psi^{(K-1)}\rangle$, and we are interested in transformations of the form

$$|\psi^{\text{in}}\rangle \to |\tilde{\phi}\rangle = \sum_{i} h(\boldsymbol{v}_{\psi^{\text{in}}})_i \, |i\rangle \, , \tag{3}$$

where $\boldsymbol{v}_{\psi^{\text{in}}}$ is the vector containing all amplitudes of the input states in the computational basis, i.e. $\boldsymbol{v}_{\psi^{\text{in}}} = (\psi^{(0)}_0, \psi^{(0)}_1, ..., \psi^{(K-1)}_{d_{\text{in}}})$ where $\psi^{(k)}_i = \langle i|\psi^{(k)}\rangle$, and $h : \mathbb{C}^{kd} \to \mathbb{C}^{d_{\text{out}}}$ is a multilinear function of its arguments. Again, if $|\psi^{\text{in}}\rangle$ includes multiple copies of the same state $|\psi\rangle$, h may be nonlinear with respect to the amplitudes of $|\psi\rangle$.

[1] The general framework allows for input systems with different Hilbert space dimensions, however we won't be considering such cases in this work.

B. General framework

In this section, we introduce the concept of 'weighted states' which provides the backbone of all the algorithms presented later on. To motivate this concept we start by emphasizing that one only ever has access to quantum states through measurement outcomes. Thus, suppose we are interested in the outcome of measurements on the state τ, it is not strictly necessary to prepare τ, rather it suffices to propose an operational strategy for calculating $\langle O\rangle_\tau = \text{Tr}[\tau O]$ for any measurement operator O.

When the state we are interested in is a nonlinear functional of the input state as in Eq. (2), we cannot strictly prepare it unless it happens to be a proper quantum state, i.e. unless $\tilde{\tau} = \tau$ is positive semi-definite, Hermitian matrix with unit trace. Instead, we first prepare a quantum state in an extended Hilbert space composed of Ancilla register (A) and an Input register (I). As shown in Fig. 1a), the initial composite state of the registers $\sigma^f_A \otimes \rho^{\text{in}}_I$ evolves unitarily under U^f such that the final state is $\rho^{\text{out}} = U^f(\sigma^f_A \otimes \rho^{\text{in}}_I)U^{f\dagger}$. Here σ^f is the ancilla state to be chosen as part of the algorithm. We then regroup the registers of ρ^{out} into a system (S), an environment (E), and the garbage (G). Finally, we pick a measurement operator M^f on the environment such that the following is satisfied:

$$\text{Tr}[\tilde{\tau}O] = \text{Tr}_{SEG}[\rho^{\text{out}}(O \otimes M^f_E \otimes \mathbb{I}_G)] \quad \forall \;\; O \, . \tag{4}$$

In other words, we map the expectation value of any system operator O in the (possibly unphysical) state $\tilde{\tau}$ to the expectation value of the operator $O \otimes M^f \otimes \mathbb{I}$ in the composite state ρ^{out}. Since Eq. (4) is true for any system operator O, we can also express the weighted state as

$$\tilde{\tau} = \text{Tr}_{EG}(\rho^{\text{out}}(\mathbb{I}_S \otimes M^f_E \otimes \mathbb{I}_G)) \, . \tag{5}$$

We note that the measurement of O need not be made straightaway, rather the weighted state $\tilde{\tau}$ can be used as the input to another algorithm and so may be further processed by additional subroutines before measurement.

Alternatively, one can think of weighted states in terms of quantum instruments which are quantum operations with both classical and quantum outputs [12]. Any quantum instrument can be realized by applying a joint unitary transformation on the input system plus an ancilla, followed by a projective measurement on part of the resulting joint system as in Fig. 1b). We indicate such an instrument[2] as $\mathcal{I}(\sigma^f, U^f, M^f)$. In general, the classical and quantum outputs are correlated. If the j'th

[2] Clearly, there is a degeneracy in the choice of the set $\{\sigma^f, U^f, M^f\}$ and different choices can result in the same instrument. One could eliminate this degeneracy by adopting the convention that M^f is diagonal and $\sigma^f = |0\rangle\langle 0|$, which is always possible as the unitary diagonalizing M^f and the one that prepares the purification of σ^f can always be absorbed into U^f. However, we believe that this convention makes it harder to build intuition about the algorithms we describe later in the paper.

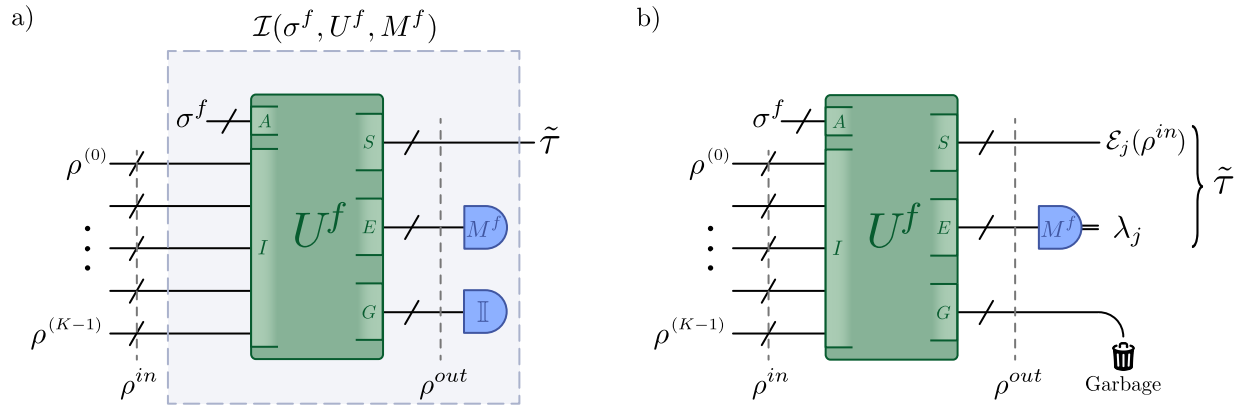

FIG. 1. **Weighted state framework.** Here we illustrate two equivalent ways of interpreting weighted states. a) A weighted state $\tilde{\tau} = \mathrm{Tr}_{EG}(\rho^{\mathrm{out}}(\mathbb{I}_S \otimes M_E^f \otimes \mathbb{I}_G))$ is equivalent to the output of a composite state ρ^{out} after the measurement $M^f \otimes \mathbb{I}$ on the environment and garbage auxiliary systems. b) A weighted state $\tilde{\tau} = \sum_j \lambda_j \mathcal{E}_j(\rho^{\mathrm{in}})$ is the output of a quantum instrument where conditional on measuring the j'th output on the environment register E, the quantum output in register S is weighted by λ_j.

possible measurement outcome λ_j is observed in a realization of the measurement of M^f, the conditional quantum state on the S register of the joint system is given by $\mathcal{E}_j(\rho^{\mathrm{in}})/\mathrm{Tr}(\mathcal{E}_j(\rho^{\mathrm{in}}))$, where $\mathcal{E}_j(\square_I) = \mathrm{Tr}_{EG}(U^f(\sigma_A^f \otimes \square_I)U^{f\dagger}(\mathbb{I}_S \otimes |\lambda_j\rangle\langle\lambda_j|_E \otimes \mathbb{I}_G))$ is a completely positive trace-non-increasing map and $\mathrm{Tr}_S(\mathcal{E}_j(\rho^{\mathrm{in}}))$ is the probability p_j of observing said outcome [13]. Since probabilities add up to one, $\mathcal{E} = \sum_j \mathcal{E}_j$ is completely positive and trace preserving and $\mathcal{E}(\rho^{\mathrm{in}}) = \mathrm{Tr}_{EG}(\rho^{\mathrm{out}})$ describes the marginal state on the S register. What we refer to as the weighted state can be expressed as

$$\tilde{\tau} = \sum_j \lambda_j p_j \frac{\mathcal{E}_j(\rho^{\mathrm{in}})}{\mathrm{Tr}(\mathcal{E}_j(\rho^{\mathrm{in}}))} = \sum_j \lambda_j \mathcal{E}_j(\rho^{\mathrm{in}})\,. \tag{6}$$

In this manner, the weighted state is obtained by weighting the quantum outcomes of the quantum instrument with its classical outcomes. In practice this weighting is implemented in post-processing. It is evident that conditional and marginal states are special cases of weighted states whereby the measurement operator M_f is a projector and identity, respectively. In many ways, one can think of the use of conditional states as being the analog of rejection sampling whereas weighted states are analogous to importance sampling.

The goal is then to find a quantum instrument $\mathcal{I}(\sigma^f, U^f, M^f)$ that implements the transformation of Eq. (1) for a given function f for all inputs ρ^{in}. First, let us assume that M^f is a normal operator and hence can be diagonalized by a unitary. A subsequent measurement in the computational basis, with outcome j, is interpreted as the measurement of λ_j, the j'th eigenvalue of M^f. The physical states $\mathcal{E}_j(\rho^{\mathrm{in}})/\mathrm{Tr}(\mathcal{E}_j(\rho^{\mathrm{in}}))$ can then be used, together with the classical outputs λ_j of the instrument, to emulate any quantum computation that involves the (possibly unphysical) state $\tilde{\tau}$.

The method described in the previous paragraph assumes M^f is a normal operator. More generally, one can emulate an arbitrary operator M^f by randomizing over quantum instruments. Let $M^{f,(k)}$ be a normal operator corresponding to the k'th instrument and q_k be the probability with which this instrument is sampled. Then the net effect of using a randomized instrument is captured by $M^f = \sum_k q_k M^{f,(k)}$ in Eqs. (4,5). Note that, in principle, any operator can be represented this way as $M^f = (1/2)(M^f + M^{f\dagger}) + (1/2)(M^f - M^{f\dagger})$, where both the Hermitian and skew-Hermitian parts of M are normal (but other decompositions may prove to be more efficient). In Appendix. A 4 we show that this randomization is not necessary and in fact a single quantum instrument can always be constructed that implements the transformation that a fictitious instrument with an arbitrary non-normal M^f would. Hence, unless stated otherwise, we will henceforth assume M^f is normal.

Finally, it is important to note that ρ^{in} can itself be a weighted state and Eqs. (5,6) still hold in this case. Thus algorithms implementing basic nonlinear transformations can be concatenated to prepare a large class of weighted states of the form Eqs. (1, 3) corresponding to complex nonlinear transformations of input states. Below we will describe three such basic algorithms. The first algorithm multiplies two input state density operators entry-wise in the computational basis (Hadamard product), the second implements a generalization of the transpose operation, and the third algorithm outputs polynomials of input density matrices.

In the rest of the paper we will drop the tilde on $\tilde{\tau}$ as there is no need to make a distinction between physical and weighted states in the framework. We will also drop the superscript f and let $\{\sigma^f, U^f, M^f\} \to \{\sigma, U, M\}$

for brevity. It is always to be understood that this set depends on f.

C. Quantum Hadamard Product (QHP)

We define the Quantum Hadamard Product (QHP) of two states $\rho^{(0)}$ and $\rho^{(1)}$ as

$$\rho^{(0)} \odot \rho^{(1)} = \sum_{ij} \rho^{(0)}_{ij} \rho^{(1)}_{ij} \, |i\rangle\langle j| \, . \tag{7}$$

That is, the matrix elements of the QHP in the computational basis are the product of the matrix elements of the input density operators.

In Fig. 2 we present a circuit for implementing the Hadamard product. The output state after applying the ladder of CNOT gates, $U = \mathrm{CNOT}^{\otimes n}$, is

$$\rho^{\mathrm{out}} = \sum_{i_0, j_0, i_1, j_1} \rho^{(0)}_{i_0 j_0} \rho^{(1)}_{i_1 j_1} \, |i_0, i_0 \oplus i_1\rangle\langle j_0, j_0 \oplus j_1| \tag{8}$$

The measurement operator is $M = |0\rangle\langle 0|$ and there is no garbage register. It is straightforward to verify that the weighted state is the desired Quantum Hadamard Product, that is we have

$$\tau = \mathrm{Tr}_{EG}(\rho^{\mathrm{out}}(\mathbb{I}_S \otimes M)) = \rho^{(0)} \odot \rho^{(1)} \, . \tag{9}$$

Note that M is a projector. As such it amounts to performing a post-selection. In that viewpoint, one would obtain the normalized state $\rho^{(0)} \odot \rho^{(1)} / \mathrm{Tr}[\rho^{(0)} \odot \rho^{(1)}]$ whenever the second register is measured in the all 0's state. Then one measures the observable O in this state. This process is repeated until enough statistics is collected. Finally, one would have to multiply the outcome with the normalization factor $\mathrm{Tr}[\rho^{(0)} \odot \rho^{(1)}]$ which can be estimated by the relative frequency of the success of post-selection. Thus the weighted state formalism includes methods based on post-selection as a special case, whereby M is proportional to a projector.

In the special case where the inputs to QHP are pure states, i.e. $\rho^{(0)} = |\psi^{(0)}\rangle\langle\psi^{(0)}|$ and $\rho^{(1)} = |\psi^{(1)}\rangle\langle\psi^{(1)}|$, the QHP transforms them into another pure state given by

$$|\phi\rangle = |\psi^{(0)} \odot \psi^{(1)}\rangle \equiv \sum_i \psi^{(0)}_i \psi^{(1)}_i \, |i\rangle \, . \tag{10}$$

Weighted states $|\psi^{(0)} \odot \psi^{(1)} \odot \ldots \psi^{(K-1)}\rangle$ that are the Hadamard product of K states can be prepared by using the QHP algorithm iteratively as shown in Fig. 3. If the input states are defined on n qubits, this algorithm can be implemented using $2n$ qubits, independent of K, by resetting and reusing qubits [14]. Of particular interest is the potential to use iterative applications of QHP to generate powers of a state, that is to prepare

$$\underbrace{|\psi \odot \psi \odot \cdots \odot \psi\rangle}_{p \text{ times}} \equiv |\psi^p\rangle \equiv \sum_i \psi^p_i \, |i\rangle \, . \tag{11}$$

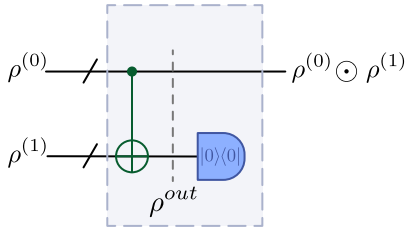

FIG. 2. **Quantum Hadamard Product (QHP).** Here we show the depth 1 circuit to implement QHP. Note that the CNOT gates act on all pairs of qubits.

More generally, using the circuit in Fig. 3 with different input states we can prepare weighted states whose amplitudes are products of the amplitudes of different states such as

$$\sum_i \prod_j \left(\psi^{(j)}_i\right)^{p_j} |i\rangle \, . \tag{12}$$

Such products may be used as building blocks for preparing arbitrary polynomial functions.

If we treat $\rho^{(0)} \to \sigma$ as part of the ancilla register (A) and only $\rho^{(1)} \to \rho$ as the input (I) of the quantum instrument, then the same quantum circuit as QHP shown in Fig. 2 implements the following linear transformation

$$\rho \to \sigma \odot \rho = \sum_{ij} \sigma_{ij} \rho_{ij} \, |i\rangle\langle j| \, . \tag{13}$$

The only difference between the two quantum instruments is the interpretation of what constitutes the input of the instrument.

D. Generalized Quantum Transpose (GQT)

In this section we present an algorithm for implementing a transformation related to the QHP that we call the Generalized Quantum Transpose (GQT). Given an input state ρ and ancillary state σ, the GQT is defined as

$$\rho^{(T)}_\sigma = \sigma \odot \rho^T \, . \tag{14}$$

That is, the output of GQT is the transpose of the input density operator ρ in the computational basis, with the elements of ρ^T weighted by the elements of σ on an element-by-element basis. When σ is chosen to be the plus state $\sigma = |+\rangle\langle+|$ with $|+\rangle = \frac{1}{\sqrt{d}} \sum_i |i\rangle$, the output is the transpose of ρ up to the dimension of the input, that is, $\rho^{(T)}_{|+\rangle\langle+|} = \frac{1}{d}\rho^T$. In the special case of pure input states, i.e. $\rho = |\psi\rangle\langle\psi|$, the GQT implements a weighted complex conjugation operation such that the output is $\sigma \odot |\psi^*\rangle\langle\psi^*|$.

In Fig. 4 we present a circuit for implementing the GQT. Specifically, the weighted state prepared by this

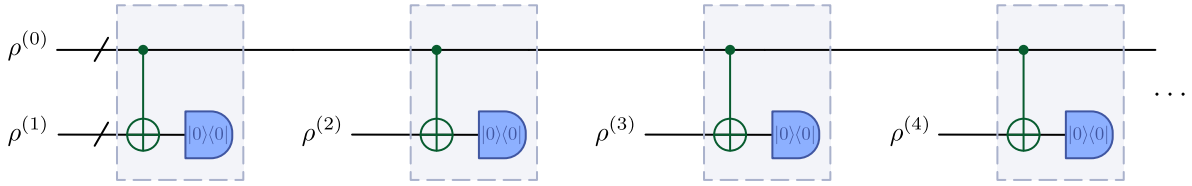

FIG. 3. **Iterated Quantum Hadamard Product.** The circuit consists of $n-1$ repetitions of the QHP subroutine shown in Fig. 2. This circuit can be used with $\rho^{(1)} = \rho^{(2)} = ... = \rho^{(k)} = |\psi\rangle\langle\psi|$ for preparing a quantum state $\sum_i \psi_i^k |i\rangle$ in which all the amplitudes in the computational basis are raised to some power k. We note that one could also reset alternating qubits. These two instruments would produce the same final weighted state on an ideal device but may perform differently in the presence of noise [14].

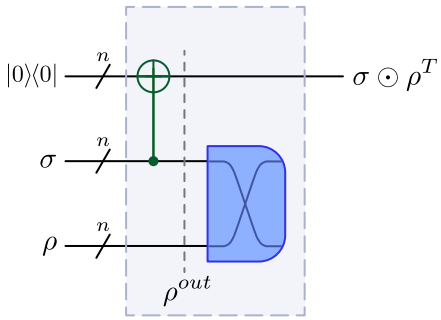

FIG. 4. **Generalized Quantum Transpose (GQT).** Here we show the depth 3 circuit to implement GQT. The purple box indicates a SWAP measurement (for breakdown see Fig. 5).

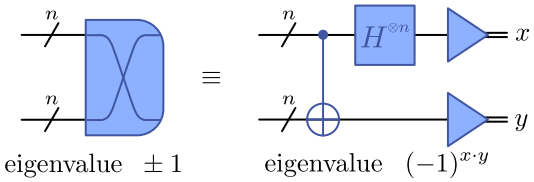

FIG. 5. **Swap measurement.** The implementation of the SWAP measurement with a circuit of depth 2 [15]. The readouts are $x = x_1, \ldots, x_n$ and $y = y_1, \ldots, y_n$, and $x \cdot y = x_1 y_1 + \cdots + x_n y_n$.

circuit can be shown to be:

$$\rho^{\text{out}} = \sum_{ii'jj'} \sigma_{ij}\rho_{i'j'} |iij\rangle\langle i'i'j'| \tag{15}$$

$$\tilde{\tau} = \text{Tr}_{2,3}(\rho^{\text{out}}(\mathbb{I} \otimes \text{SWAP})) \tag{16}$$

$$= \sum_{iji'j'} \sigma_{ij}\rho_{i'j'} \langle i'j'| \text{SWAP} |ij\rangle |i\rangle\langle i'| \tag{17}$$

$$= \sum_{ij} \sigma_{ij}\rho_{ji} |i\rangle\langle j| = \sigma \odot \rho^T , \tag{18}$$

where the trace in second equation is taken over the last two registers. Note that the Bell basis states are eigenstates of the SWAP operator with eigenvalues ± 1. Hence the SWAP measurement on the bottom two registers can be realized with a depth two quantum circuit as seen in Fig. 5 [15].

GQT can be used in tandem with the other algorithms described in this work to enlarge the set of functions f that can be implemented. Although GQT, in its current formulation, implements a linear transformation, if the ancilla state is chosen to be ρ, or GQT is applied iteratively taking multiple copies of the state ρ as inputs, the resulting transformation will be nonlinear in ρ. It is worth noting that GQT can be applied to a subsystem of a larger system to implement a partial transpose. Thus GQT could be leveraged to witness entanglement in mixed quantum states [16, 17].

E. Quantum State Polynomial (QSP)

In this section we present an algorithm that takes as input two states $\rho^{(0)}$ and $\rho^{(1)}$ and prepares weighted states

$$\begin{aligned}\tilde{\tau} =& \alpha_{00}\rho^{(0)} + \alpha_{11}\rho^{(1)} \\ &+ \alpha_{01}\rho^{(0)}\rho^{(1)} + \alpha_{10}\rho^{(1)}\rho^{(0)} .\end{aligned} \tag{19}$$

Note that the weighted state τ is a multilinear polynomial of its two inputs $\rho^{(0,1)}$. Consider the algorithm in Fig. 6. Specifically, the pre-measurement output of the circuit is

$$\begin{aligned}\rho^{\text{out}} = \sum_{ii'jj'} &\sigma_{00}\rho^{(0)}_{ij}\rho^{(1)}_{i'j'} |i0j\rangle\langle i'0j'| + \sigma_{01}\rho^{(0)}_{ij}\rho^{(1)}_{i'j'} |i0j\rangle\langle j'1i'| \\ &+\sigma_{10}\rho^{(0)}_{ij}\rho^{(1)}_{i'j'} |j1i\rangle\langle i'0j'| + \sigma_{11}\rho^{(0)}_{ij}\rho^{(1)}_{i'j'} |j1i\rangle\langle j'1i'| .\end{aligned} \tag{20}$$

Performing the measurement M we generate the weighted state

$$\begin{aligned}\tilde{\tau} =& \sigma_{00}M_{00}\,\text{Tr}(\rho^{(1)})\rho^{(0)} + \sigma_{11}M_{11}\,\text{Tr}(\rho^{(0)})\rho^{(1)} \\ &+\sigma_{01}M_{10}\rho^{(0)}\rho^{(1)} + \sigma_{10}M_{01}\rho^{(1)}\rho^{(0)} .\end{aligned} \tag{21}$$

Hence by appropriately choosing the initial state σ and measurement operator M it is possible to produce states

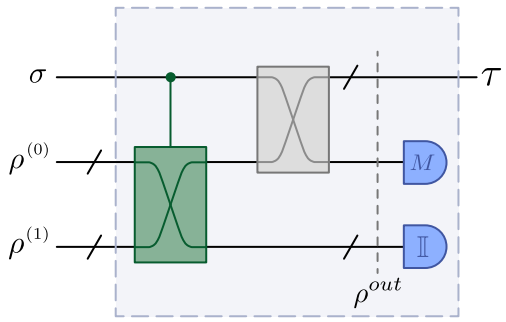

FIG. 6. **Quantum State Polynomial (QSP).** Here we show the circuit to implement QSP. The controlled SWAP gate can be implemented with three Toffoli gates. The final swap does not need to be implemented but rather is included to match the register labelling convention chosen in Fig. 1.

of the form Eq. (19) as claimed. More concretely, interpreting the coefficients α_{ij} as entries of a matrix α we require that

$$\alpha = \sigma \odot M^T \odot \gamma^{\text{in}} . \tag{22}$$

Here γ^{in} is a matrix where the off-diagonal elements are input independent with $\gamma^{\text{in}}_{01} = \gamma^{\text{in}}_{10} = 1$ and the diagonal entries depend on the input states with $\gamma^{\text{in}}_{00} = \text{Tr}(\rho^{(1)})$ and $\gamma^{\text{in}}_{11} = \text{Tr}(\rho^{(0)})$.

Multilinear polynomials of more than two input states, generalizing Eq. (19), can be obtained by concatenating this algorithm. However, only a limited subset of all multilinear polynomials can be generated this way. In order to obtain any arbitrary multilinear polynomial, one can either randomize over instruments, see Appendix A 4, or forego concatenation and implement the transformation directly with an instrument that takes all inputs at once, see Appendix B 1.

A number of interesting families of weighted states may be generated with QSP when σ is a physical density operator and M is an Hermitian observable. Here we list several pertinent examples.

Mixtures: One does not need weighted states to prepare the mixture of two given states but here we show how that scheme fits in the larger framework of weighted states. Here we want $\tilde{\tau} = p\rho^{(0)} + (1-p)\rho^{(1)}$ for some probability $0 < p < 1$. This corresponds to $\alpha = \left(\begin{smallmatrix} p & 0 \\ 0 & 1-p \end{smallmatrix}\right)$. We can achieve this by choosing $\sigma_A = \left(\begin{smallmatrix} p & 0 \\ 0 & 1-p \end{smallmatrix}\right)$ and $M = \mathbb{I} = \left(\begin{smallmatrix} 1/\text{Tr}(\rho^{(1)}) & 0 \\ 0 & 1/\text{Tr}(\rho^{(0)}) \end{smallmatrix}\right)$. When the inputs are physical states with unit trace, the measurement M is just the identity operator and so the environment register may simply be traced out.

Anti-commutator: The weighted state $\tilde{\tau} = \{\rho^{(0)}, \rho^{(1)}\} = \rho^{(0)}\rho^{(1)} + \rho^{(1)}\rho^{(0)}$ corresponds to $\alpha = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. This may be prepared using $\sigma = |+\rangle\langle+| = (1/2)\left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ and $M = 2X = 2\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. We note that one-half of the anti-commutator with identical inputs ρ is the weighted state $\tilde{\tau} = \rho^2$. In this special case, the algorithm reduces to established algorithms for computing Renyi entropies [14] and virtual state distillation [8, 18].

Commutator: The weighted state $\tilde{\tau} = [\rho^{(0)}, \rho^{(1)}] = (\rho^{(0)}\rho^{(1)} - \rho^{(1)}\rho^{(0)})$ corresponds to $\alpha = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$. This may be prepared with $\sigma = |+\rangle\langle+| = (1/2)\left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ and $M = 2iY = 2\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$.

Linear Combinations of Pure States: A particularly promising application of the QSP algorithm is to prepare linear combinations of pure quantum states. In this case the weighted state is $\tau = |\psi\rangle\langle\psi|$ where $|\psi\rangle = \alpha_0|\psi^{(0)}\rangle + \alpha_1|\psi^{(1)}\rangle$, corresponding to

$$\alpha = \begin{pmatrix} |\alpha_0|^2 & \frac{\alpha_0\alpha_1^*}{\langle\psi_0|\psi_1\rangle} \\ \frac{\alpha_1\alpha_0^*}{\langle\psi_1|\psi_0\rangle} & |\alpha_1|^2 \end{pmatrix} . \tag{23}$$

Note that unlike previous algorithms we presented, in this case, as α depends on the overlap between the input states, the functional form of g in Eq. (3) depends on the input state. In particular, one needs to know in advance, or estimate, the overlaps between the input states in order to specify the quantum instrument that prepares the desired weighted state.

We have freedom in how to pick the initial state σ and measurement M. Suppose we take σ to be the arbitrary pure state $\sigma = |\beta\rangle\langle\beta|$ with $|\beta\rangle = \beta_0|0\rangle + \beta_1|1\rangle$ then we need

$$M = \begin{pmatrix} \frac{|\alpha_0|^2}{|\beta_0|^2\langle\psi_1|\psi_1\rangle} & \frac{\alpha_1\alpha_0^*}{\beta_1\beta_0^*\langle\psi_1|\psi_0\rangle} \\ \frac{\alpha_0\alpha_1^*}{\beta_0\beta_1^*\langle\psi_0|\psi_1\rangle} & \frac{|\alpha_1|^2}{|\beta_1|^2\langle\psi_0|\psi_0\rangle} \end{pmatrix} . \tag{24}$$

Note that M is a Hermitian operator. In Section III we discuss how to choose $|\beta\rangle$ in order to minimize the sampling complexity.

We note that the denominator for the off diagonal elements in Eq. (24) will vanish if $|\psi_0\rangle$ and $|\psi_1\rangle$ are orthogonal. Even if the denominator does not vanish, for close to orthogonal states it can become very small. As we will demonstrate in the error analysis in Section III this leads to precision issues. First, the overlap $\langle\psi_0|\psi_1\rangle$ will have to be estimated with high precision. Second, even if this overlap is known exactly the eigenvalues of M, i.e. the weights, will be large. This in turn increases the sampling complexity. Hence, this method is not recommended for preparing linear combinations of orthogonal, or close to orthogonal, states. In Appendix B we discuss alternative methods for preparing such combinations.

One can prepare linear combinations of many states by iterating this method for preparing the linear combination of a pair of states. This ability to take the linear combinations of states is expected to prove valuable for a number of applications since it can be used as a primitive to prepare arbitrary polynomials of quantum states. In particular, by taking the linear combination of powers

of quantum states (generated via the QHP algorithm) one may approximately implement any function of the quantum state that may be expanded as a power series as

$$|\psi\rangle \to |g(\psi)\rangle = \sum_i g(\psi_i)\,|i\rangle \tag{25}$$

where $g(\psi_i) = \sum_k \alpha_k(\psi_i)^k$. For example, this could be used to approximately implement the reciprocal operation $g(x) = \frac{1}{x}$ to amplify basis states with small amplitudes. Or, one could potentially use this method to implement the $g(x) = \tanh(x)$ activation function used to introduce nonlinearities into neural networks. More generally the QSP, GQT, and QHP may be concatenated to implement complex nonlinear transformations of the form

$$|\psi\rangle \to |h(\psi)\rangle = \sum_{ikl} \alpha_{kl}(\psi_i)^k(\psi_i^*)^l\,|i\rangle\ . \tag{26}$$

Physically measurements are associated with normal operators. If we require M to be a normal operator in the QSP circuit, then it follows from Eq. (22) that not all α matrices can be obtained. In Appendix D we provide a detailed classification of the classes of weighted states that can be realised when M is a normal operator. To go beyond, as previously noted, we can express arbitrary operators as sums of normal operators N_l via $M = \sum_l c_l N_l$. The right hand side of Eq. (4) now has a summation, which means that we need to operate multiple quantum instruments to obtain the weighted state. As an example of weighted states that can only be prepared via QSP through such means consider products of density matrices, i.e. $\tau = \rho^{(0)}\rho^{(1)}$.

III. SAMPLING COMPLEXITY ANALYSIS

We are interested in the expectation value of an Hermitian operator O in a weighted state τ, i.e. $\mathrm{Tr}[\tau O]$. Due to Eq. (4) this quantity is nothing but the expectation value of $O\otimes M\otimes I$ in the state ρ^{out} and can be estimated by repeatedly running the quantum instrument of Fig. 1 and then measuring the operator O on the quantum output, i.e. the system register. This estimator takes the form

$$\widehat{\mathcal{O}} = \frac{1}{s}\sum_i^s \lambda_{m(i)}\mu_{o(i)} \tag{27}$$

where $\lambda_{m(i)}$ and $\mu_{o(i)}$ are the random outcomes of M and O measurements at the i'th run of the circuit, respectively, and s is the total number of runs. Note that $\widehat{\mathcal{O}}$ is an unbiased estimator since it estimates a quantum expectation value in the standard way:

$$E[\widehat{\mathcal{O}}] = \mathrm{Tr}[\rho^{\mathrm{out}}(O\otimes M\otimes \mathbb{I})] = \mathrm{Tr}[\tau O] \tag{28}$$

where $E[\cdot]$ denotes the expectation value.

The variance of $\widehat{\mathcal{O}}$ is

$$\mathrm{Var}(\widehat{\mathcal{O}}) = \frac{1}{s}\left(\mathrm{Tr}[\rho^{\mathrm{out}}(O\otimes M\otimes\mathbb{I})^2] - \mathrm{Tr}[\tau O]^2\right)\ . \tag{29}$$

Note that, in general, the variance of an observable can not be expressed in terms of the weighted state τ alone. Consequently, two instruments that implement the same transformation and hence prepare the same weighted state, can result in estimators with different variances. Since variance is inversely proportional to sampling complexity we would like to find instruments that minimize it. As is clear from Eq. (29), the variance manifestly depends on the measurement O performed on the weighted state and so to compare the sampling complexity of different instruments it is desirable to derive an operator independent bound on the variance. To do so we note that the variance of the product of random variables may be bounded [19] as

$$\mathrm{Var}(\widehat{XY}) \le 2\mathrm{Var}(\widehat{X})||Y||_\infty^2 + 2\langle X\rangle^2\mathrm{Var}(\widehat{Y})\,. \tag{30}$$

Assuming that $||O||_\infty \le 1$, it follows that

$$\mathrm{Var}(\widehat{\mathcal{O}}) \le \frac{2\,\mathrm{Tr}[\rho^{\mathrm{out}}(\mathbb{I}\otimes M\otimes\mathbb{I})^2]}{s}\,. \tag{31}$$

We use Eq. (29) (Eq. (31)) to compute (bound) the number of samples needed for QHP, GQT, and QSP algorithms in order to achieve a desired precision. Here we restrict our scope to weighted states that are prepared using quantum instruments whose inputs are physical states[3] and for which M is a Hermitian operator. We also suppose that the system observable O is Hermitian. Here we summarise our findings. The more detailed and general analysis can be found in Appendix. C.

A. Quantum Hadamard Product

For the Quantum Hadamard Product algorithm the variance takes the form

$$\mathrm{Var}[\hat{\mathcal{O}}] = \frac{1}{s}(\mathrm{Tr}[\tau O^2] - \mathrm{Tr}[\tau O]^2) \tag{32}$$

Thus the variance has the form of a standard quantum observable, except that in this case τ may be a sub-normalized state with $\mathrm{Tr}[\tau] \le 1$. This expression, Eq. (32), also holds when the QHP circuit is applied iteratively to generate higher order powers/polynomials of the input states.

[3] Note that if the input states are weighted states themselves, one needs to take into account the quantum instruments associated with them in order to analyze the variance. This can be achieved by treating the concatenated quantum instruments as a single quantum instrument with more registers.

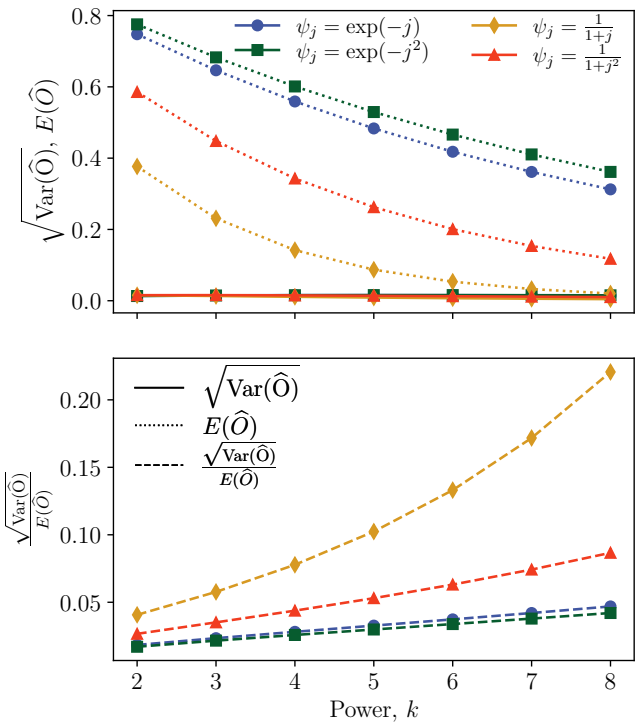

FIG. 7. **Error analysis for powers of states.** Here we plot the absolute sampling error $\sqrt{\mathrm{Var}(\widehat{\mathcal{O}})}$ (solid), the analytically computed expectation value $\mathrm{E}(\widehat{\mathcal{O}})$ (dotted) and the relative error $\sqrt{\mathrm{Var}(\widehat{\mathcal{O}})}/\mathrm{E}(\widehat{\mathcal{O}})$ (dashed) after running the powers of state circuit for the normalized states $|\psi\rangle \propto \sum_j \psi_j |j\rangle$, with the functions ψ_j indicated in the legend, as a function of power k. In all cases we consider an $n = 6$ qubit state $|\psi\rangle$, we measure the all zero projector $O = |0\rangle\langle 0|$ and suppose $s = 1000$ shots are used.

In Fig. 7 we plot the absolute and relative sampling errors when applying the QHP algorithm for generating increasing powers of different input states. While the additive error (solid lines) remains approximately constant as k is increased, the relative error

$$\frac{\mathrm{Var}(\widehat{\mathcal{O}})}{\mathrm{Tr}[\tau O]^2} = \frac{1}{s}\left(\frac{\mathrm{Tr}[\tau O^2]}{\mathrm{Tr}[\tau O]^2} - 1\right) \tag{33}$$

typically increases dramatically with power (dashed lines). This follows from the fact that the amplitudes of $|\psi\rangle$ are typically less than 1 (for proper quantum states) and so $\mathrm{Tr}[\tau] = \sum_i |\psi_i|^{2k}$ typically decreases exponentially with k (dotted lines).

B. Generalized Quantum Transpose

For the Generalized Quantum Transpose circuit shown in Fig. 2, the variance is given by

$$\mathrm{Var}(\widehat{\mathcal{O}}) = \frac{1}{s}\left(\mathrm{Tr}\left[D(\sigma)O^2\right] - \mathrm{Tr}\left[\left(\sigma \odot \rho^T\right) O\right]^2\right) \tag{34}$$

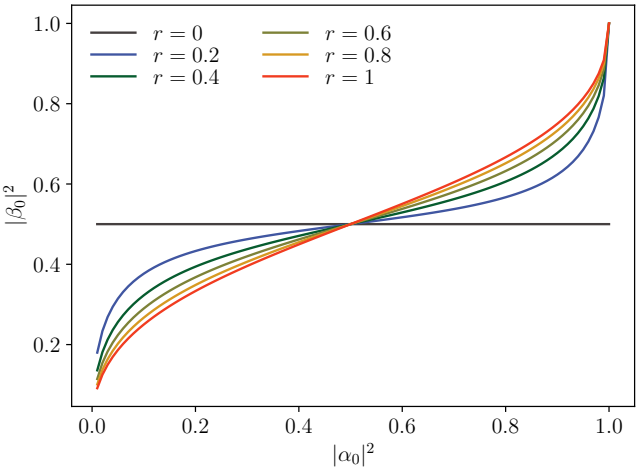

FIG. 8. **Optimum choice in β parameter.** We plot β_0^{opt}, the optimum β_0 parameter that minimises $\mathcal{B}_{\mathrm{Var}}$, the upper bound on the error for the linear combinations of states circuit, as a function of $|\alpha_0|^2$ for different overlaps r.

where $D(\sigma)$ is the dephased version of the ancilla state σ, i.e. $D(\sigma) = \sum_i \sigma_{ii}|i\rangle\langle i|$, and we used the fact that $\mathrm{SWAP}^2 = \mathbb{I}$.

We recall that the Generalized Quantum Transpose may be used to compute expectation values of the transpose of the state ρ if σ is chosen to be the plus state $|+\rangle\langle +|$ and the observable dO is measured. The factor of d compensates for the normalization factor in $\rho^{(T)}_{|+\rangle\langle +|} = \frac{1}{d}\rho^T$. While this is one possible use of GQT, in this case the variance is given by

$$\mathrm{Var}(\widehat{\mathcal{O}}) = \frac{1}{s}\left(d\,\mathrm{Tr}\left[O^2\right] - \mathrm{Tr}\left[\rho^T O\right]^2\right), \tag{35}$$

such that the sampling error scales with d and so exponentially in the number of qubits of the input system. Thus for large scale problems GQT is not efficient. However, if GQT is used to implement the partial transpose to a d_s dimensional subsystem then the variance will scale as d_s. Hence even for large scale problems it should be possible to implement the partial transpose of a constant sized subsystem.

C. Linear combination of states

Here we focus on analysing the sampling complexity of the linear combination of states algorithm - see Appendix C for a presentation of the errors in the general case of the quantum state polynomial circuit. Using Eq. (31), one can show that for the linear combination of states algorithm the variance may be bounded as

$$\begin{aligned}\mathrm{Var}(\widehat{\mathcal{O}}) \leq \frac{2}{s}\Bigg(|\alpha_0|^2\left(\frac{|\alpha_0|^2}{|\beta_0|^2} + \frac{|\alpha_1|^2}{q_1 r}\right) \\ + |\alpha_1|^2\left(\frac{|\alpha_0|^2}{|\beta_0|^2 r} + \frac{|\alpha_1|^2}{|\beta_1|^2}\right)\Bigg) := \mathcal{B}_{\mathrm{Var}}\,,\end{aligned} \tag{36}$$

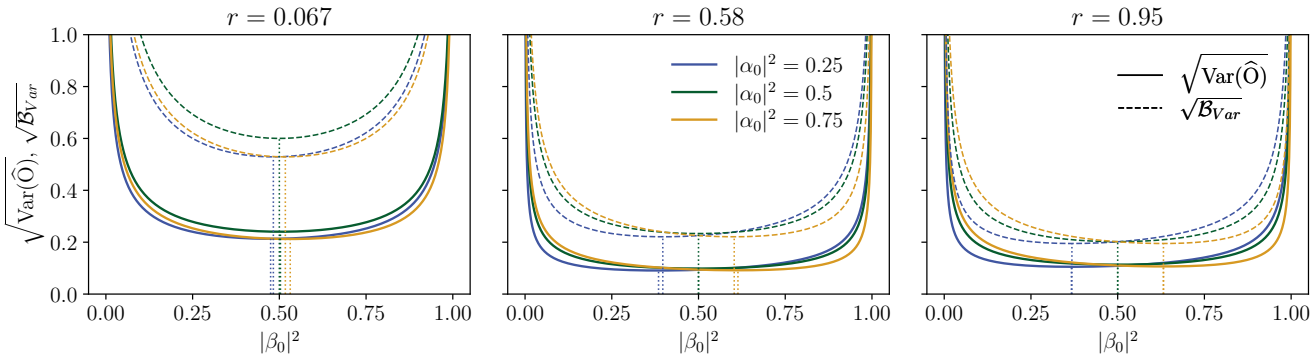

FIG. 9. **Error Analysis for Linear Combination of Unitaries.** Here we plot the the actual standard deviation of the estimator, $\sqrt{\mathrm{Var}(\widehat{\mathcal{O}})}$, (solid) and its upper bound $\sqrt{\mathcal{B}_{\mathrm{Var}}}$ (dashed) for the linear combination of states algorithm for an $n = 6$ qubit circuit evaluated using $N = 100$ shots when O is a random separable measurement operator. We plot the results for pairs of randomly generated states with a small overlap ($r = 0.067$, left), median overlap ($r = 0.58$, middle) and maximum overlap ($r = 0.95$, right) for three different choices in amplitude: $\alpha_0 = 0.25$ (blue), $\alpha_0 = 0.5$ (green) and $\alpha_0 = 0.95$ (yellow). The dotted lines indicate the value of $|\beta_0|^2$ corresponding to the minimum error and minimum of the bound respectively, with the close agreement between these two minima indicating that the optimum $|\beta_0|^2$ determined from the bound, Eq. (C20), is close to the true optimum. In agreement with Fig. 8 the optimal $|\beta_0|^2$ value is close to 0.5 for small r but has a stronger dependence on α_0 for larger r values.

where $r = |\langle\psi_0|\psi_1\rangle|^2$ is the overlap between the input states. As remarked earlier, for this algorithm there is a freedom in how the ancilla state and measurement operator may be chosen. We can use the above bound to approximately determine the optimum β_0 value which minimises the variance of the estimator. Specifically, we find that the β_0 value that minimizes $\mathcal{B}_{\mathrm{Var}}$ is

$$\beta_0^{\mathrm{opt}}(p,r) = \sqrt{\frac{p - p^2 + p^2 r + r^{-1}h(p,r)}{2p - 2p^2 + r - 2pr + 2p^2 r + 2r^{-1}h(p,r)}} \tag{37}$$

where we use the shorthand $p := |\alpha_0|^2$ and define

$$h(p,r) := \sqrt{(p^2 - p)(-p + p^2 - r + 2pr - 2p^2 r - pr^2 + p^2 r^2)}\,. \tag{38}$$

We plot β_0^{opt} against $|\alpha_0|^2$ in Fig. 8. For states with small overlaps, $r \to 0$, the approximately optimum strategy is to use $\beta_0 = 1/\sqrt{2}$ for any choice in α_0, whereas, for states with large overlaps, $r \to 1$, the optimum β_0 value increases monotonically with increasing α_0.

In Fig. 9 we plot the exact variance of the estimator (for the analytic expression see Appendix C) and the bound on the variance $\mathcal{B}_{\mathrm{Var}}$ from Eq.(36) as a function of β_0, for different overlaps r and weights α_0. While the bound is not tight it is useful in coming up with strategies to minimize the variance of the estimator. In particular, the optimum β_0 given by Eq. (37), found by minimizing the bound Eq.(36), closely agrees with the true optimum (found by numerically minimising the true variance for the given examples). As expected given that the algorithm breaks down for orthogonal input states, both the bound and exact variance diverge for input states with vanishing overlap ($r \to 0$).

IV. DISCUSSION

In this work we introduced a framework for implementing nonlinear transformations in quantum computers by associating so-called weighted states with the output of quantum instruments. More specifically, weighted states are quantum objects describing the output of an operational procedure involving quantum circuits, measurements and classical post-processing. While playing a similar role to standard density matrices, weighted states are liberated from the constraints of positivity, Hermiticity and normalization and hence can be generic functions of input states.

We have introduced three algorithms for implementing nonlinear functionals of the elements of a set of quantum states. The Quantum Hadamard Product algorithm takes two states $\rho^{(0)}$ and $\rho^{(1)}$ as inputs and generates a state $\rho^{(0)} \odot \rho^{(1)}$ as an output. That is, QHP outputs a weighted state where the elements of $\rho^{(0)}$ and $\rho^{(1)}$ have been multiplied in the computational basis. The Generalised Quantum Transpose implements the transpose of an operator ρ in the computational basis, with the elements of ρ^T reweighted by the elements of an operator σ. The Quantum State Polynomial algorithm takes $\rho^{(0)}$ and $\rho^{(1)}$ as inputs and outputs the polynomial $\alpha_{00}\rho^{(0)} + \alpha_{11}\rho^{(1)} + \alpha_{01}\rho^{(0)}\rho^{(1)} + \alpha_{10}\rho^{(1)}\rho^{(0)}$. When applied to pure states, iterative applications of QHP, GQT, QSP can be used to generate arbitrary polynomials of the amplitudes of a set of pure states.

In Appendix E we show results from a proof of principle implementation of QSP and QHP on IBMQ-Bogota. While the implementations correctly capture the qualitative effect of performing nonlinear transformations of quantum states, the low CNOT and qubit reset fidelities

lead to non-negligible deviations from the expected results. However, it is important to note that, due to their amenability to qubit resets [14, 18], these algorithms do not require large numbers of qubits. In particular, QSP, GQT, and QHP can be concatenated and iterated an arbitrarily many times with at most three times the number of qubits of the input states. Thus a polynomial of order p of an n-qubit pure state can be implemented using at most $3n$ qubits (for any p).

It is important to note that distinct quantum instruments can implement the same transformation with differing complexities. In other words, in this framework, the complexity associated with a transformation is not an inherent property of the transformation. However, one can ask what is the optimal quantum instrument for a given transformation. A promising setting in which this question might be tractable is when the cost of implementing unitaries is neglected, in other words we are only concerned with minimizing the sampling complexity. How exactly to do this optimization is an open question; however, in Section III C we were able to do a partial optimization for one of our algorithms (namely, QSP as applied to implementing linear combinations of states).

In this paper we focused on primitives that implement intuitive transformations and require relatively simple quantum circuits for their implementation. These primitives can be concatenated to yield complex nonlinear transformations of input states. While concatenation is qubit-efficient, in Appendix C 7 we provide examples of cases where a transformation can be implemented more efficiently directly. Thus it would be valuable to have a method for directly designing quantum instruments for a target transformation. More generally, it would be interesting to investigate the breadth of applicability of the weighted state methodology by quantifying the full set of transformations that can be implemented within this framework.

V. ACKNOWLEDGEMENTS

We thank Denis Aslangil, Lukasz Cincio, Patrick Coles, Yen Ting Lin, Robert Lowrie, Rolando Somma, and Burak Şahinoğlu for useful discussions. This work was supported by the Beyond Moore's Law project of the Advanced Simulation and Computing Program at Los Alamos National Laboratory (LANL) managed by Triad National Security, LLC, for the National Nuclear Security Administration of the U.S. Department of Energy under contract 89233218CNA000001. Research presented in this article was also supported by the Laboratory Directed Research and Development program of LANL. ZH acknowledges support from the Mark Kac Fellowship. Part of this work was completed while NC was a participant in the 2021 Quantum Computing Summer School at LANL, sponsored by the LANL Information Science & Technology Institute.

[1] I. Cong, S. Choi, and M. D. Lukin, Nature Physics **15**, 1273 (2019).
[2] K. Beer, D. Bondarenko, T. Farrelly, T. J. Osborne, R. Salzmann, D. Scheiermann, and R. Wolf, Nature Communications **11**, 808 (2020).
[3] M. Lubasch, J. Joo, P. Moinier, M. Kiffner, and D. Jaksch, Phys. Rev. A **101**, 010301 (2020).
[4] B. Georgeot and D. L. Shepelyansky, Phys. Rev. Lett. **86**, 2890 (2001).
[5] B. Georgeot and D. L. Shepelyansky, Phys. Rev. Lett. **86**, 5393 (2001).
[6] I. Dodin and E. Startsev, arXiv preprint arXiv:2105.07317 (2021), https://arxiv.org/abs/2105.07317.
[7] N. Guo, K. Mitarai, and K. Fujii, arXiv preprint arXiv:2107.10764 (2021).
[8] W. J. Huggins, S. McArdle, T. E. O'Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, R. Babbush, and J. R. McClean, arXiv preprint arXiv:2011.07064 (2020).
[9] H. Barnum, C. M. Caves, C. A. Fuchs, R. Jozsa, and B. Schumacher, Phys. Rev. Lett. **76**, 2818 (1996).
[10] G. M. D'Ariano, C. Macchiavello, and P. Perinotti, Phys. Rev. Lett. **95**, 060503 (2005).
[11] M. Piani, P. Horodecki, and R. Horodecki, Phys. Rev. Lett. **100**, 090502 (2008).
[12] E. B. Davies and J. T. Lewis, Communications in Mathematical Physics **17**, 239 (1970).
[13] O. Oreshkov, F. Costa, and Č. Brukner, Nature Communications **3**, 1092 (2012).
[14] J. Yirka and Y. Subaşı, Quantum **5**, 535 (2021).
[15] L. Cincio, Y. Subaşı, A. T. Sornborger, and P. J. Coles, New Journal of Physics **20**, 113022 (2018).
[16] A. Peres, Phys. Rev. Lett. **77**, 1413 (1996).
[17] M. Horodecki, P. Horodecki, and R. Horodecki, Physics Letters A **223**, 1 (1996).
[18] P. Czarnik, A. Arrasmith, L. Cincio, and P. J. Coles, arXiv preprint arXiv:2102.06056 (2021).
[19] D. Giraudo, "Bound the variance of the product of two random varables." Mathematics Stack Exchange, https://math.stackexchange.com/q/1044864 (version: 2014-11-30).
[20] A. M. Childs and N. Wiebe, arXiv preprint arXiv:1202.5822 (2012).
[21] A. Arrasmith, L. Cincio, R. D. Somma, and P. J. Coles, arXiv preprint arXiv:2004.06252 (2020).

Appendices

Appendix A: Quantum instruments

In this section we elaborate on the quantum instruments associated with the quantum Hadamard product, quantum generalized transpose, and density matrix polynomial algorithms and provide alternative derivations of these algorithms from this perspective.

1. Quantum Hadamard Product

The quantum instrument for the QHP is shown in Fig. 2. Let x be the outcome of a measurement in the computational basis. Given this classical output of the quantum instrument, the associated quantum output is proportional to:

$$\mathcal{E}_x(\rho^{(0)} \otimes \rho^{(1)}) = \sum_{i,j} \rho^{(0)}_{ij} \rho^{(1)}_{i\oplus x, j\oplus x} |i\rangle\langle j| \tag{A1}$$

The weighting in this case corresponds to post-selection; only $x = 0$ has unit weight and all other outcomes have zero weight.

$$\tilde{\tau} = \sum_x \delta_{x,0} \mathcal{E}_x(\rho^{(0)} \otimes \rho^{(1)}) \tag{A2}$$

$$= \sum_{i,j} \rho^{(0)}_{ij} \rho^{(1)}_{i,j} |i\rangle\langle j| = \rho^{(0)} \odot \rho^{(1)} \tag{A3}$$

Note that one can generate a large family of weighted states by projecting onto states other than $|0\rangle$ and/or nontrivially weighting all outcomes of the measurement.

2. Generalized Quantum Transpose

The quantum instrument for the GQT is shown in Fig. 4. In practice the measurement of SWAP operator is implemented by measuring pairs of qubits in the Bell basis as seen in Fig 5 [15]. This can be achieved by first acting with a CNOT on qubit pairs and then Hadamard gates on the control qubits, and finally measuring all qubits in the computational basis. Let x and y be the binary strings that are the outcome of the measurements in the two registers where SWAP operator is measured. Given this classical output of the quantum instrument, the associated quantum output is proportional to:

$$\mathcal{E}_{x,y}(\sigma \otimes \rho) = \frac{1}{2^n} \sum_{ij} \sigma_{i,j} \rho_{i\oplus y, j\oplus y} (-1)^{x\cdot(i\oplus j)} |i\rangle\langle j| \tag{A4}$$

where n is the number of qubits in each input register. The weighted state is obtained by weighting the quantum outcomes of the instruments as:

$$\tilde{\tau} = \sum_{x,y} (-1)^{x\cdot y} \mathcal{E}_{x,y}(\sigma \otimes \rho) \tag{A5}$$

$$= \sum_y \sum_{ij} \sigma_{i,j} \rho_{i\oplus y, j\oplus y} \left(\frac{1}{2^n} \sum_x (-1)^{x\cdot(i\oplus j\oplus y)} \right) |i\rangle\langle j| \tag{A6}$$

$$= \sum_{ij} \sigma_{i,j} \rho_{j,i} = \sigma \odot \rho^T \tag{A7}$$

where going from the second to the last line we used the fact $\sum_x (-1)^{x\cdot(i\oplus j\oplus y)} = 2^n \delta_{y,i\oplus j}$ and $i \oplus i \oplus j = j$. All summations are modulo 2.

3. Quantum State Polynomial

The quantum instrument for QSP is shown in Fig. 6. Let $|\psi_l\rangle$ be the state that correspond to the l'th outcome of the M measurement. Given this classical output of the quantum instrument, the associated quantum output is proportional to:

$$\mathcal{E}_l(\rho^{(0)} \otimes \rho^{(1)}) = \sum_{i,j} \sigma_{ij} \langle j|\psi_l\rangle \langle\psi_l|i\rangle \left[\delta_{ij} \operatorname{Tr}(\rho^{(i\oplus 1)})\rho^{(i)} + (1-\delta_{ij})\rho^{(i)}\rho^{(j)}\right] \tag{A8}$$

The weighted state is obtained by weighting the quantum outcomes of the instrument as:

$$\tilde{\tau} = \sum_l \lambda_l \mathcal{E}_l(\rho^{(0)} \otimes \rho^{(1)}) \tag{A9}$$

$$= \sum_{i,j} \sigma_{ij} \langle j| \left(\sum_l \lambda_l |\psi_l\rangle\langle\psi_l|\right) |i\rangle \left[\delta_{ij} \operatorname{Tr}(\rho^{(i\oplus 1)})\rho^{(i)} + (1-\delta_{ij})\rho^{(i)}\rho^{(j)}\right] \tag{A10}$$

$$= \sum_{ij} (\sigma \odot M^T)_{ij} \left[\delta_{ij} \operatorname{Tr}(\rho^{(i\oplus 1)})\rho^{(i)} + (1-\delta_{ij})\rho^{(i)}\rho^{(j)}\right] \tag{A11}$$

$$= \sigma_{00} M_{00} \operatorname{Tr}(\rho^{(1)})\rho^{(0)} + \sigma_{11} M_{11} \operatorname{Tr}(\rho^{(0)})\rho^{(1)} + \sigma_{01} M_{10} \rho^{(0)}\rho^{(1)} + \sigma_{10} M_{01} \rho^{(1)}\rho^{(0)} \tag{A12}$$

where going from the second to the last line we used the fact $M = \sum_l \lambda_l |\psi_l\rangle\langle\psi_l|$.

4. Going beyond normal measurement operators

Let us consider a set of quantum instruments $\mathcal{I}_k(\sigma, U, N_k)$, labeled by k, that share the same ancilla state σ and that apply the same unitary U, but differ in terms of the measurement operators M now referred to as N_k to emphasize they are normal. Data from such quantum instruments can be compounded to emulate a fictitious quantum instrument $\mathcal{I}_k(\sigma, U, \bar{M})$ with a non-normal measurement operator $\overline{M} = \sum_k c_k N_k$. This follows from Eq. (4) via linearity. That is, as

$$\sum_k c_k \operatorname{Tr}[\rho^{\text{out}}(O \otimes N_k \otimes \mathbb{I})] = \operatorname{Tr}[\rho^{\text{out}}(O \otimes \sum_k c_k N_k \otimes \mathbb{I})] = \operatorname{Tr}[\rho^{\text{out}}(O \otimes \overline{M} \otimes \mathbb{I})] \tag{A13}$$

it follows that $\sum_k c_k \mathcal{I}_k(\sigma, U, N_k) \cong \mathcal{I}_k(\sigma, U, \bar{M})$. It is easy to see that this allows us to effectively implement any M since any operator can be written as the sum of its Hermitian and anti-Hermitian components via

$$\overline{M} = \frac{1}{2}(\overline{M} + \overline{M}^\dagger) + \frac{1}{2}(\overline{M} - \overline{M}^\dagger)\,. \tag{A14}$$

As both parts are normal, we could pick two instruments $N_0 = (\overline{M} + \overline{M}^\dagger)$ and $N_1 = (\overline{M} - \overline{M}^\dagger)$ with $c_0 = c_1 = 1/2$ to achieve our goal. We emphasize that this example is only intended as a proof of existence and not a recipe. There are an infinite number of ways to satisfy Eq. (A13), and some ways may result in better sampling complexity than others.

However, such non-normal measurement operators are not strictly necessary for full generality. This is because given an instrument $\mathcal{I}(\sigma, U, \overline{M})$, where $\overline{M}$ is a non-normal operator it is possible to define an alternative instrument $\mathcal{I}(\sigma\otimes\sigma', U\otimes\mathbb{I}, M)$, where M is a normal operator, and the alternative instrument implements the same transformation as the original.

To see how let us, without loss of generality, assume $c_k > 0$ and $\sum_k c_k = 1$. The alternative quantum instrument $\mathcal{I}(\sigma_A \otimes \sigma'_{A'}, U_{AI} \otimes \mathbb{I}_{A'}, M_{E'E})$, where we now include subscripts to explicitly denote the relevant subsystem, shown in Fig. 10. There is an additional ancilla register initially prepared in the state $\sigma'_{A'} = \sum_k c_k |k\rangle\langle k|$. The unitary is the same as before and does not act on this new ancilla register. The environment register on the output now includes the added ancilla register. The measurement operator is given by $M_{E'E} = \sum_k |k\rangle\langle k|_{E'} \otimes (N_k)_E$, which is normal by construction. The quantum state before the measurement is $\rho'^{\text{out}}_{SEE'G} = \rho^{\text{out}}_{SEG} \otimes \sigma'_{E'}$ and so we have

$$\operatorname{Tr}_{SEE'G}[(\rho^{\text{out}}_{SEG} \otimes \sigma'_{E'})(O_S \otimes M_{EE'} \otimes \mathbb{I}_G)] = \operatorname{Tr}_{SEG}[\rho^{\text{out}}_{SEG}(O_S \otimes \sum_k c_k (N_k)_E \otimes \mathbb{I}_G)] \tag{A15}$$

$$= \operatorname{Tr}_{SEG}[\rho^{\text{out}}_{SEG}(O_S \otimes \overline{M}_E \otimes \mathbb{I}_G)]\,. \tag{A16}$$

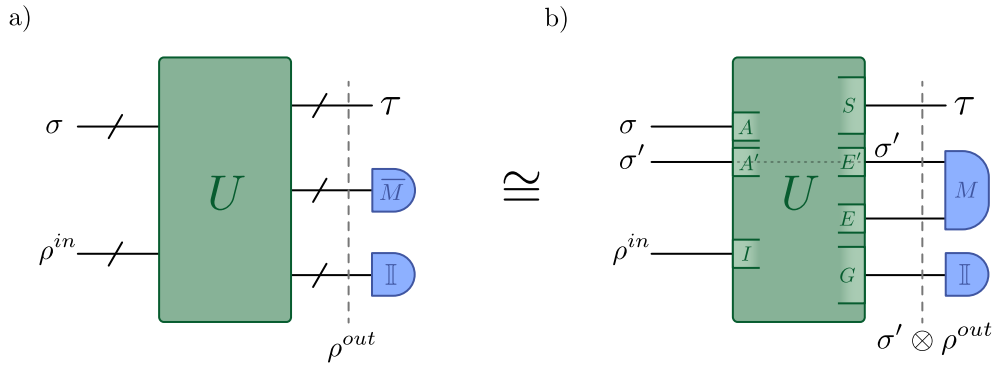

FIG. 10. **Capturing a non-normal measurement within the standard weighted state framework.** Here we show two equivalent quantum instruments $\mathcal{I}(\sigma_A, U_{AI}, \overline{M}_E)$ and $\mathcal{I}(\sigma_A \otimes \sigma'_{A'}, U_{AI} \otimes \mathbb{I}_{A'}, M_{E'E})$.

Thus $\mathcal{I}(\sigma_A, U_{AI}, \overline{M}_E) \cong \mathcal{I}(\sigma_A \otimes \sigma'_{A'}, U_{AI} \otimes \mathbb{I}_{A'}, M_{E'E})$ as claimed. Note that this construction is equivalent to randomly sampling quantum instruments with measurement operator N_k with probability c_k, but the formulation in terms of a single instrument allows us to simplify notation for the rest of the discussion.

Appendix B: Alternative methods for preparing a linear combination of states

Given $L+1$ states $\{|\psi_i\rangle\}_{i=0}^{L}$, our goal is to simulate the effect of having access to a superposition of these states:

$$|\Phi\rangle = \sum_{l=0}^{L} \alpha_l |\phi_l\rangle \ . \tag{B1}$$

The states $\{|\phi_i\rangle\}_{i=0}^{L}$ do not form an orthonormal basis, and $|\Phi\rangle$ does not have to be normalized. In the main text we describe a method for generating such linear combinations of states via repeated applications of the Quantum State Polynomial algorithm. Here we describe three alternative methods for this task.

1. 'All-At-Once' weighted state method

In this section we describe a weighted state methodology where the linear combination of states is generated in a single step via a global measurement. This approach is more general than the iterative application of QSP algorithm, capturing it as a special case.

Consider the quantum circuit shown in Fig. 11. Here $\sigma = |\beta\rangle\langle\beta|$ where

$$|\beta\rangle = \sum_{l=0}^{L} \beta_l |l\rangle \ . \tag{B2}$$

The controlled unitary is given by

$$|l\rangle\langle l| \otimes \hat{\pi}_l \tag{B3}$$

where $\hat{\pi}_l$ is a permutation that maps the l'th state to the 0'th state

$$\hat{\pi}_l |\phi_0\rangle \otimes |\phi_1\rangle \otimes \cdots \otimes |\phi_L\rangle = |\phi_l\rangle \otimes |\phi_{\pi_l(1)}\rangle \cdots \otimes |\phi_{\pi_l(L)}\rangle \tag{B4}$$

i.e. $\pi_l(0) = l$ for the state labels[4]. Beyond this specification, we leave the permutations π_l underspecified. The state

[4] One could equivalently see this operation as mapping the 0'th register to the l'th register.

at the end of this operation is given by

$$\sum_{l=0}^{L} \beta_l \left|l\right\rangle \otimes \left|\phi_l\right\rangle \bigotimes_{k=1}^{L} \left|\phi_{\pi_l(k)}\right\rangle \tag{B5}$$

The next step is to trace out all but the top two registers. To do so, we first write down the density operator for the pure state of all the registers given above

$$\sum_{l,l'=0}^{L} \beta_l \beta_{l'}^* \left|l\right\rangle\!\left\langle l'\right| \otimes \left|\phi_l\right\rangle\!\left\langle\phi_{l'}\right| \bigotimes_{k=1}^{L} \left|\phi_{\pi_l(k)}\right\rangle\!\left\langle\phi_{\pi_{l'}(k)}\right| . \tag{B6}$$

Then, tracing out the registers 1 through L we get

$$\sum_{l,l'=0}^{L} \beta_l \beta_{l'}^* \prod_{k=1}^{L} \left\langle\phi_{\pi_{l'}(k)}|\phi_{\pi_l(k)}\right\rangle \left|l\right\rangle\!\left\langle l'\right| \otimes \left|\phi_l\right\rangle\!\left\langle\phi_{l'}\right| . \tag{B7}$$

Next we swap the top two registers and compute the expectation value of an operator M on the lower register and output the top register. This leaves us with a weighted state given by

$$\sum_{l,l'=0}^{L} \beta_l \beta_{l'}^* \left\langle l'|M|l\right\rangle \prod_{k=1}^{L} \left\langle\phi_{\pi_{l'}(k)}|\phi_{\pi_l(k)}\right\rangle \otimes \left|\phi_l\right\rangle\!\left\langle\phi_{l'}\right| \tag{B8}$$

We want this to match the density operator

$$\left|\Phi\right\rangle\!\left\langle\Phi\right| = \sum_{l,l'=0}^{L} \alpha_l \alpha_{l'}^* \left|\phi_l\right\rangle\!\left\langle\phi_{l'}\right| \tag{B9}$$

Matching term by term we demand

$$\left\langle l'|M|l\right\rangle = \frac{\alpha_l \alpha_{l'}^*}{\beta_l \beta_{l'}^* \prod_{k=1}^{L} \left\langle\phi_{\pi_{l'}(k)}|\phi_{\pi_l(k)}\right\rangle} \tag{B10}$$

We remark that M is Hermitian, i.e. $M_{ll'}^* = \left\langle l|M|l'\right\rangle^* = \left\langle l'|M|l\right\rangle = M_{l'l}$, and hence is a valid observable. If we have access to multiple copies of the states ϕ_l we can estimate the overlaps in the denominator. For small L we can construct the matrix M and compute its eigenvalues efficiently. Moreover, we can classically compute a unitary U_M that diagonalizes M, i.e. $U_M^\dagger M U_M$ is diagonal, and find a quantum circuit that implements it on a quantum computer. Then the measurement of M can be achieved by first applying U_M and then measuring in the computational basis and recording the associated eigenvalue. Thus the circuit shown in Fig. 11 can be used to prepare the weighted state $\left|\Phi\right\rangle$.

Similarly, to the method for generating linear combinations of states via repeated applications of the Quantum State Polynomial algorithm, this algorithm breaks down if any one of the states $\{\left|\phi_l\right\rangle\}_{l=0}^{L}$ is outside the space spanned by the rest of the states. In the next two subsections we propose alternative methods that may be used in these cases.

2. Incoherent post-processing method

If our goal is simply to apply some unitary V to $\left|\Phi\right\rangle$ and make a measurement of O, we can achieve this by applying unitaries to $\{\left|\phi_l\right\rangle\}$ individually and computing matrix elements of M in this new set. That is one can compute

$$\left\langle\Phi|V^\dagger O V|\Phi\right\rangle = \sum_{l,l'=0}^{L} \alpha_l \alpha_{l'}^* \left\langle\phi_{l'}|V^\dagger O V|\phi_l\right\rangle \tag{B11}$$

by computing each $\left\langle\phi_{l'}|V^\dagger O V|\phi_l\right\rangle$ term, reweighting it by $\alpha_l \alpha_{l'}^*$ and then summing together the outputs.

Computing the diagonal terms $\left\langle\phi_l|V^\dagger O V|\phi_l\right\rangle$ is straightforward. This is simply done by evolving $\left|\phi_l\right\rangle$ under V and then measuring O. The off diagonal terms $\left\langle\phi_{l'}|V^\dagger O V|\phi_l\right\rangle$, for $l \neq l'$, can be computed using the Hadamard test if one has access to the controlled versions of the unitaries W_l that prepare $\left|\phi_l\right\rangle$. To do so, we have to first expand

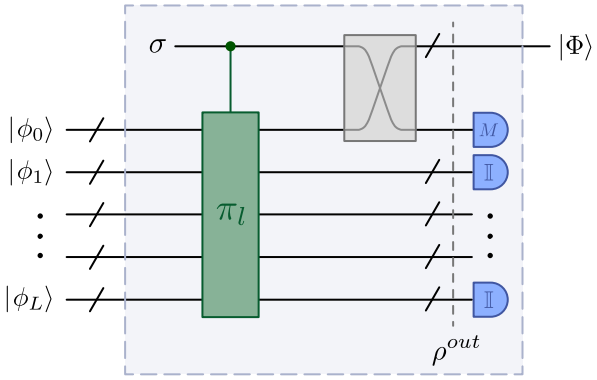

FIG. 11. **Direct Linear combination of states algorithm.** Here we show a circuit to directly compute the linear combination of many states. $\hat{\pi}_l$ is a permutation that maps the l'th state to the 0'th state. The final swap does not need to be implemented but rather is included to match the register labelling convention chosen in Fig. 1.

the measurement operator as as a linear combination of unitaries, e.g. as $O = \sum_i r_i U_i$. Equipped with this linear combination we then have

$$\langle \phi_{l'}|V^\dagger O V|\phi_l\rangle = \sum_i r_i \langle 0|W_{l'}^\dagger V^\dagger U_i V W_l|0\rangle \ . \tag{B12}$$

Thus $\langle \phi_{l'}|V^\dagger O V|\phi_l\rangle$ is a weighted sum of the expectation value of the unitaries $W_{l'}^\dagger V^\dagger U_i V W_l$ in the state $|0\rangle$, and each of these expectation values may be computed with a Hadamard test. In applications where the gate sequence for implementing W_l is known, a controlled version can be obtained with a constant factor overhead.

3. Linear Combination of Unitaries (LCU) method

Here we assume that we have access to the controlled versions of the unitaries W_l that prepare $|\phi_l\rangle$. In order to prepare the linear combination of states in Eq. (B1) we act on the $|0\rangle$ state with the LCU given by

$$\sum_{l=0}^{L} \alpha_l W_l \, . \tag{B13}$$

The LCU method [20] has the same restriction as the incoherent method in that it requires access to the controlled version of the unitaries W_l. The main difference is that the LCU method actually prepares the normalized state

$$|\Phi_N\rangle = \frac{|\Phi\rangle}{\| \, |\Phi\rangle \, \|} = \frac{1}{\sqrt{\sum_{ll'} \alpha_l^* \alpha_{l'} \langle \phi_l|\phi_{l'}\rangle}} \sum_{l=0}^{L} \alpha_l \, |\phi_l\rangle \tag{B14}$$

with probability $(\| \, |\Phi\rangle \, \|/\|\alpha\|_1)^2$, where $\|\alpha\|_1 = \sum_l |\alpha_l|$. This probability can be boosted by amplitude amplification. In addition, note that we need to know $\| \, |\Phi\rangle \, \|$ in order to compute quantities of interest in terms of $|\Phi\rangle$ using the actual state $|\Phi_N\rangle$. This can be done by computing the overlaps $\langle \phi_l|\phi_{l'}\rangle = \langle 0|W_l^\dagger W_{l'}|0\rangle$ using the Hadamard test, since we assume access to the controlled-W_l's.

Appendix C: Sampling complexity analysis

1. General analysis for weighted states

In the approach outlined in Section II the expectation value of an operator O in a weighted state $\tilde{\tau}$ is obtained from the expectation value of the operator $O \otimes M$ in a state ρ^{out} — see Eq. (4). In this appendix we analyze the sampling complexity of estimating expectation values using this approach in general before focusing on specific applications in

the following sections. Unlike in the main text where M was Hermitian, here we allow it to be an arbitrary normal operator. The total operator whose expectation value we need to estimate is given by

$$T = O \otimes M \otimes \mathbb{I} \tag{C1}$$

We estimate $\langle T \rangle$ with $\widehat{\mathcal{T}}$ using

$$\widehat{\mathcal{T}} = \frac{1}{s} \sum_{i=1}^{s} \nu_{t(i)} \ . \tag{C2}$$

Here, s is the number of circuit evaluations and $\nu_{t(i)}$ is the outcome of i'th measurement i.e. the $t(i)$'th eigenvalue of T. Since T is a normal operator $\widehat{\mathcal{T}}$ is a complex random number. This estimator is unbiased since it relies on the standard method for estimating a quantum expectation value. The variance is defined as

$$\mathrm{Var}(\widehat{\mathcal{T}}) = E[|\widehat{\mathcal{T}}|^2] - |E[\widehat{\mathcal{T}}]|^2 \tag{C3}$$

$$= \frac{1}{s} \left(\mathrm{Tr}[\rho^{\mathrm{out}} (OO^\dagger \otimes MM^\dagger \otimes \mathbb{I})] - |\,\mathrm{Tr}[\tau O]|^2 \right) \ . \tag{C4}$$

In Appendix. A 4 we explained how to express a quantum instrument with a non-normal measurement operator $\overline{M} = \sum_k c_k N_k$ in terms of an instrument with a normal measurement operator M. In this case the above analysis can be repeated to yield

$$\mathrm{Var}(\widehat{\mathcal{T}}) = \frac{1}{s} \left(\sum_k c_k \,\mathrm{Tr}[\bar{\rho}^{\mathrm{out}} (OO^\dagger \otimes N_k N_k^\dagger \otimes \mathbb{I})] - |\,\mathrm{Tr}[\tau O]|^2 \right) \tag{C5}$$

Ideally one wants to minimize the number of shots s, i.e. circuit evaluations, needed to achieve a desired variance for the estimator. This can be done, in theory, by optimizing over all quantum instruments $\{\mathcal{I}(\sigma, U, M)\}$ that implement the desired transformation. However, it is an open question how to do this in practice. In addition, there may be other considerations than sampling complexity. For instance, in order to measure a normal operator we first need to apply the diagonalizing unitary, which might be hard to compute classically or hard to implement on a quantum computer. Such considerations should also be taken into account when formulating the optimization task, especially when using noisy quantum devices.

2. Proof of operator independent bound on variance

To compare the sampling complexity associated with different quantum instruments implementing the same transformation and provide an observable independent analysis of the complexity of our algorithms, it is advantageous to derive an operator independent bound on the variance. Here we provide proofs of our operator independent bounds on the variance quoted in the main text. We start by providing a proof [19] for Eq. 30,

$$\sigma_{XY}^2 \leq 2\sigma_X^2 |||Y||_\infty|^2 + 2|\langle X \rangle|^2 \sigma_Y^2 \tag{C6}$$

which we repeat for the general case where X and Y are complex random variables. To derive this we consider two complex random variables $A = (X - \langle X \rangle)Y$ and $B = \langle X \rangle Y$, and note that

$$\mathrm{Var}(XY) = \mathrm{Var}(A + B) \tag{C7}$$

$$\leq 2\mathrm{Var}(A) + 2\mathrm{Var}(B) \tag{C8}$$

$$= 2\mathrm{Var}((X - \langle X \rangle)Y) + 2|\langle X \rangle|^2 \mathrm{Var}(Y) \, . \tag{C9}$$

We then note that,

$$\mathrm{Var}((X - \langle X \rangle)Y) \leq \langle |X - \langle X \rangle)Y|^2 \rangle \tag{C10}$$

$$\leq \mathrm{Var}(X) |||Y||_\infty|^2 \, , \tag{C11}$$

and thus we obtain Eq. (C6). Applying this to evaluating $\mathrm{Var}(\widehat{\mathcal{O}})$, and assuming O is Hermitian, we have that

$$\mathrm{Var}(\widehat{\mathcal{O}}) \leq \frac{2}{s} \left(\sigma_M^2 ||O||_\infty^2 + |\langle M \rangle|^2 \sigma_O^2 \right) \ . \tag{C12}$$

Assuming that the largest eigenvalue of O is bounded by 1 we have that $||O||_\infty^2 \leq 1$ and $\sigma_O^2 \leq 1$ and so

$$\mathrm{Var}(\widehat{\mathcal{O}}) \leq \frac{2}{s}\left(\sigma_M^2 + |\langle M\rangle|^2\right) = \frac{2\langle |M|^2\rangle}{s} := \mathcal{B}_{\mathrm{Var}}\,. \tag{C13}$$

In the following sections we evaluate the variance $\mathrm{Var}(O)$ and its bound Eq. (C13) for the primitive nonlinear subroutines that we have introduced in this manuscript.

3. Quantum state polynomial

For the case of the quantum state polynomial algorithm the variance of the estimator evaluates to

$$\begin{aligned}\mathrm{Var}(\widehat{\mathcal{O}}) = \frac{1}{s}\Big(&\sigma_{00}[M^\dagger M]_{00}\,\mathrm{Tr}[\rho_0 O^2] + \sigma_{11}[M^\dagger M]_{11}\,\mathrm{Tr}[\rho_1 O^2] \\ &+ \sigma_{01}[M^\dagger M]_{10}\,\mathrm{Tr}[\rho_0\rho_1 O^2] + \sigma_{10}[M^\dagger M]_{01}\,\mathrm{Tr}[\rho_1\rho_0 O^2] - |\,\mathrm{Tr}[\tau O]|^2\Big)\,.\end{aligned} \tag{C14}$$

Since we have $\alpha = \sigma \odot M^T \odot \gamma^{\mathrm{in}}$, it follows that

$$\begin{aligned}\mathrm{Var}(\widehat{\mathcal{O}}) = \frac{1}{s}\Bigg(&\sigma_{00}\left(\frac{|\alpha_{00}|^2}{|\sigma_{00}|^2} + \frac{|\alpha_{10}|^2}{|\sigma_{10}|^2}\right)\mathrm{Tr}[\rho_1]\,\mathrm{Tr}[\rho_0 O^2] + \sigma_{11}\left(\frac{|\alpha_{11}|^2}{|\sigma_{11}|^2} + \frac{|\alpha_{01}|^2}{|\sigma_{10}|^2}\right)\mathrm{Tr}[\rho_0]\,\mathrm{Tr}[\rho_1 O^2] \\ &+ \left(\frac{\alpha_{00}}{\sigma_{00}}\frac{\alpha_{01}^*}{\sigma_{01}^*} + \frac{\alpha_{11}^*}{\sigma_{11}^*}\frac{\alpha_{10}}{\sigma_{10}}\right)\sigma_{01}\,\mathrm{Tr}[\rho_0\rho_1 O^2] + \left(\frac{\alpha_{00}^*}{\sigma_{00}^*}\frac{\alpha_{01}}{\sigma_{01}} + \frac{\alpha_{11}}{\sigma_{11}}\frac{\alpha_{10}^*}{\sigma_{10}^*}\right)\sigma_{10}\,\mathrm{Tr}[\rho_1\rho_0 O^2] - |\,\mathrm{Tr}[\tau O]|^2\Bigg)\,.\end{aligned} \tag{C15}$$

We note that this expression is real since the off diagonal elements are complex conjugate of one another. We can use the operator independent bound, Eq. (C13), to bound this variance as

$$\begin{aligned}\mathrm{Var}(\widehat{\mathcal{O}}) \leq \frac{2}{s}\Bigg(&\left(\sigma_{00}\left(\frac{\alpha_{00}^2}{\sigma_{00}^2} + \frac{\alpha_{01}\alpha_{10}}{\sigma_{01}\sigma_{10}}\right) + \sigma_{11}\left(\frac{\alpha_{11}^2}{\sigma_{11}^2} + \frac{\alpha_{01}\alpha_{10}}{\sigma_{01}\sigma_{10}}\right)\right)\mathrm{Tr}[\rho_0]\,\mathrm{Tr}[\rho_1] \\ &+ \left(\frac{\alpha_{00}}{\sigma_{00}} + \frac{\alpha_{11}}{\sigma_{11}}\right)(\alpha_{01} + \alpha_{10})\,\mathrm{Tr}[\rho_0\rho_1]\Bigg)\,.\end{aligned} \tag{C16}$$

4. Linear combination of states via Quantum State Polynomial algorithm

For the special case of a linear combination of states we have $\alpha_{00} \to |\alpha_0|^2/\langle\psi_0|\psi_0\rangle$, $\alpha_{11} \to |\alpha_1|^2/\langle\psi_1|\psi_1\rangle$, $\alpha_{01} \to \alpha_0\alpha_1^*/\langle\psi_0|\psi_1\rangle$, $\alpha_{10} \to \alpha_1\alpha_0^*/\langle\psi_1|\psi_0\rangle$ and $\sigma_{00} \to |\beta_0|^2$, $\sigma_{11} \to |\beta_1|^2$, $\sigma_{01} \to \beta_0\beta_1^*$, $\sigma_{10} \to \beta_1\beta_0^*$. Therefore in this case we have

$$\begin{aligned}\mathrm{Var}(\widehat{\mathcal{O}}) &= \frac{1}{s}\Bigg(|\beta_0|^2 N_1\left(\frac{|\alpha_0|^4}{|\beta_0|^4 N_0} + \frac{|\alpha_0|^2|\alpha_1|^2}{|\beta_0|^2|\beta_1|^2 r}\right)\mathrm{Tr}[\rho_0 O^2] + |\beta_1|^2 N_0\left(\frac{|\alpha_1|^4}{|\beta_1|^4 N_1} + \frac{|\alpha_0|^2|\alpha_1|^2}{|\beta_0|^2|\beta_1|^2 r}\right)\mathrm{Tr}[\rho_1 O^2] \\ &\quad + 2\left(\frac{|\alpha_0|^2}{|\beta_0|^2\sqrt{N_0}} + \frac{|\alpha_1|^2}{|\beta_1|^2\sqrt{N_1}}\right)\Re\left(\alpha_0\alpha_1^*\langle\psi_0|O^2|\psi_1\rangle\right) - \mathrm{Tr}[\tau O]^2\Bigg) \\ &= \frac{1}{s}\Bigg(|\alpha_0|^2 N_1\left(\frac{|\alpha_0|^2}{|\beta_0|^2 N_0} + \frac{\alpha_1|^2}{|\beta_1|^2 r}\right)\mathrm{Tr}[\rho_0 O^2] + |\alpha_1|^2 N_0\left(\frac{|\alpha_1|^2}{|\beta_1|^2 N_1} + \frac{|\alpha_0|^2}{|\beta_0|^2 r}\right)\mathrm{Tr}[\rho_1 O^2] \\ &\quad + 2\left(\frac{|\alpha_0|^2}{|\beta_0|^2\sqrt{N_0}} + \frac{|\alpha_1|^2}{|\beta_1|^2\sqrt{N_1}}\right)\Re\left(\alpha_0\alpha_1^*\langle\psi_0|O^2|\psi_1\rangle\right) - \mathrm{Tr}[\tau O]^2\Bigg)\,,\end{aligned} \tag{C17}$$

where $N_0 = |\langle\psi_0|\psi_0\rangle|^2$, $N_1 = |\langle\psi_1|\psi_1\rangle|^2$, and $r = |\langle\psi_0|\psi_1\rangle|^2$.

With the weighted state method for generating linear combinations of states there is some freedom in how the final measurement operator is chosen depending on the state of the ancilla qubit. To assess the optimum strategy, i.e. the

optimum ancilla state and measurement pair that minimizes the sampling complexity, it is advantageous to use the operator independent bound on the variance Eq (C13). It follows from Eq (C13) and Eq. (C17) that

$$\langle M^2 \otimes \mathbb{I} \rangle = |\alpha_0|^2 \left(\frac{|\alpha_0|^2}{|\beta_0|^2} + \frac{|\alpha_1|^2}{|\beta_1|^2 r} \right) + |\alpha_1|^2 \left(\frac{|\alpha_0|^2}{|\beta_0|^2 r} + \frac{|\alpha_1|^2}{|\beta_1|^2} \right) . \tag{C18}$$

Setting $|\alpha_0|^2 = p$ and $|\beta_0|^2 = q$ this can be rewritten as

$$\langle M^2 \otimes \mathbb{I} \rangle = p \left(\frac{p}{q} + \frac{(1-p)}{(1-q)r} \right) + (1-p) \left(\frac{p}{qr} + \frac{(1-p)}{(1-q)} \right) = f(p,q,r) . \tag{C19}$$

Given $0 \leq p \leq 1$ and $0 \leq r \leq 1$, $f(p,q,r)$ is minimized when

$$q_{\text{opt}}(p,r) = \frac{2p - 2p^2 + 2p^2 r + 2r\sqrt{\frac{(-p+p^2)(-p+p^2-r+2pr-2p^2r-pr^2+p^2r^2)}{r^2}}}{2\left(2p - 2p^2 + r - 2pr + 2p^2 r + 2r\sqrt{\frac{(-p+p^2)(-p+p^2-r+2pr-2p^2r-pr^2+p^2r^2)}{r^2}}\right)} \tag{C20}$$

This expression thus gives us a way of picking the ancilla state specified by β_0 given a target state specified by α_0 and $r = |\langle \psi_0 | \psi_1 \rangle|^2$. In Fig. 8 we use (C20) to plot the optimum $|\beta_0|^2$ as a function of $|\alpha_0|^2$ and r. In the limit that $r \to 0$, we have that $q_{\text{opt}} =\to 1/2$, that is the optimum strategy is picking $|\beta_0| \approx 1/\sqrt{2}$. Conversely, in the limit that $r \to \infty$ we have $|\beta_0|^2 = q_{\text{opt}} \to \frac{p}{p+\sqrt{(1-p)p}}$.

5. Linear combination of states via incoherent post-processing method

Suppose instead we compute the expectation value of an observable O with respect to the state $|\psi\rangle$ using purely classical post-processing. That is, as discussed in Appendix B 2, we compute

$$\begin{aligned} \langle \psi | O | \psi \rangle &= |\alpha_0|^2 \langle \psi_0 | O | \psi_0 \rangle + |\alpha_1|^2 \langle \psi_1 | O | \psi_1 \rangle + \alpha_0^* \alpha_1 \langle \psi_0 | O | \psi_1 \rangle + \alpha_0 \alpha_1^* \langle \psi_1 | O | \psi_0 \rangle \\ &= |\alpha_0|^2 \langle \psi_0 | O | \psi_0 \rangle + |\alpha_1|^2 \langle \psi_1 | O | \psi_1 \rangle + 2\Re(\alpha_0^* \alpha_1)\Re(\langle \psi_0 | O | \psi_1 \rangle) - 2\Im(\alpha_0^* \alpha_1)\Im(\langle \psi_0 | O | \psi_1 \rangle) \end{aligned} \tag{C21}$$

by computing the terms $\langle \psi_0 | O | \psi_0 \rangle$, $\langle \psi_1 | O | \psi_1 \rangle$, $\Re(\langle \psi_0 | O | \psi_1 \rangle)$ and $\Im(\langle \psi_0 | O | \psi_1 \rangle)$ separately. To measure the off-diagonal terms $\Re(\langle \psi_0 | O | \psi_1 \rangle)$ and $\Im(\langle \psi_0 | O | \psi_1 \rangle)$ we can expand the measurement operator O as a sum of unitaries[5], i.e. $O = \sum_i \eta_i U_i$, and then use pairs of Hadamard tests to obtain the real and imaginary parts of the overlap terms $\langle \psi_0 | U_i | \psi_1 \rangle$. That is, in total one evaluates each of the terms in

$$\langle \psi | O | \psi \rangle = |\alpha_0|^2 \langle \psi_0 | O | \psi_0 \rangle + |\alpha_1|^2 \langle \psi_1 | O | \psi_1 \rangle + 2 \sum_i \eta_i \left(\Re(\alpha_0^* \alpha_1)\Re(\langle \psi_0 | U_i | \psi_1 \rangle) - \Im(\alpha_0^* \alpha_1)\Im(\langle \psi_0 | U_i | \psi_1 \rangle) \right) . \tag{C22}$$

Let use denote the estimators of the terms $\langle \psi_0 | O | \psi_0 \rangle$, $\langle \psi_1 | O | \psi_1 \rangle$, $\Re(\langle \psi_0 | U_i | \psi_1 \rangle)$, and $\Im(\langle \psi_1 | U_i | \psi_0 \rangle)$ in Eq. (C22) by $\widehat{\mathcal{P}}_0$, $\widehat{\mathcal{P}}_1$, $\widehat{\mathcal{P}}_{2,i}$ and $\widehat{\mathcal{P}}_{3,i}$ respectively. Following the approach of Ref. [21] we will suppose that our total shot quota is divided between the different circuits in proportion to the prefactor in the sum. Under this approach, for a random variable of the form $Q = \sum_i \mu_i Q_i$, we use $s_i \propto |\mu_i|$ shots to get the estimator $\widehat{\mathcal{Q}}_i$, resulting in a variance of

$$\text{Var}(\widehat{\mathcal{Q}}) = \sum_i \frac{|\mu_i|^2 \text{Var}(\widehat{\mathcal{Q}}_i)}{s_i} = \sum_i \frac{|\mu_i|^2 \text{Var}(\widehat{\mathcal{Q}}_i)}{\lfloor |\mu_i| \tilde{s} \rfloor} \approx \sum_i \frac{|\mu_i| \text{Var}(\widehat{\mathcal{Q}}_i)}{\tilde{s}} . \tag{C23}$$

Here the normalisation term $\tilde{s}$ is chosen such that $\sum_i s_i = \sum_i \lfloor |\mu_i| \tilde{s} \rfloor = s$ and the approximation at the last equality gets better with the number of shots. The variance for the incoherent post-processing method thus takes the form

$$\begin{aligned} \text{Var}(\widehat{\mathcal{O}}) \approx \frac{1}{\tilde{s}} \Big(& |\alpha_0|^2 \text{Var}(\widehat{\mathcal{P}}_0) + |\alpha_1|^2 \text{Var}(\widehat{\mathcal{P}}_1) + \\ & 2|\Re(\alpha_0^* \alpha_1)| \sum_i \eta_i \text{Var}(\widehat{\mathcal{P}}_{2,i}) + 2|\Im(\alpha_0^* \alpha_1)| \sum_i \eta_i \text{Var}(\widehat{\mathcal{P}}_{3,i}) \Big) , \end{aligned} \tag{C24}$$

[5] Without loss of generality we here assume that the η_i prefactors are real and positive, $\eta_i > 1$, by absorbing any phases into the corresponding unitary U_i.

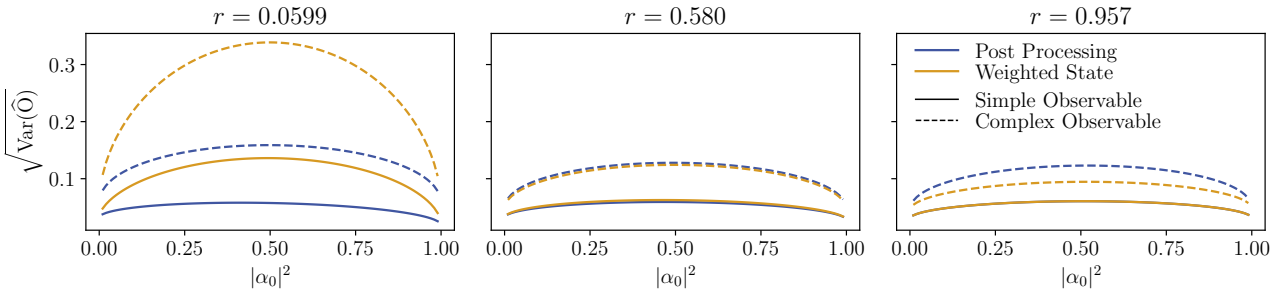

FIG. 12. **Comparison of weighted state and post-processing methods.** Here we plot the standard deviation of the estimator, $\sqrt{\mathrm{Var}(\widehat{\mathcal{O}})}$, for the weighted state method (yellow) and the post-processing method (blue) for implementing the linear combination of two $n = 4$ qubit states as a function of α_0. In both cases we suppose that a total of $s = 100$ shots are used. We plot the results for pairs of randomly generated states with a small overlap ($r = 0.067$, left), median overlap ($r = 0.58$, middle) and maximum overlap ($r = 0.95$, right). We compare the case of measuring a simple measurement (solid) composed of a single Pauli term and a complex measurement (dashed) composed of n^2 Pauli terms.

where $\tilde{s} \approx s\left(|\alpha_0|^2 + |\alpha_1|^2 + 2\left(|\Re(\alpha_0^*\alpha_1)| + |\Im(\alpha_0^*\alpha_1)|\right)\sum_i r_i\right)^{-1}$.

Since the Hadamard test only requires measuring a Pauli operator, the errors for computing the overlap terms take the form

$$\mathrm{Var}(P_{2i}) = 1 - \Re(\langle\psi_0| U_i |\psi_1\rangle)^2 \tag{C25}$$

$$\mathrm{Var}(P_{3i}) = 1 - \Im(\langle\psi_0| U_i |\psi_1\rangle)^2 . \tag{C26}$$

Thus the total error takes the form

$$\begin{aligned}\mathrm{Var}(\widehat{\mathcal{O}}) = \frac{1}{\tilde{s}}\Bigg(&|\alpha_0|^2\left(\langle\psi_0| O^2 |\psi_0\rangle - \langle\psi_0| O |\psi_0\rangle^2\right)\\ &+ |\alpha_1|^2\left(\langle\psi_1| O^2 |\psi_1\rangle - \langle\psi_1| O |\psi_1\rangle^2\right)\\ &+ 2\Re(\alpha_0^*\alpha_1)\sum_i r_i\left(1 - \Re(\langle\psi_0| U_i |\psi_1\rangle)^2\right)\\ &+ 2\Im(\alpha_0^*\alpha_1)\sum_i r_i\left(1 - \Im(\langle\psi_0| U_i |\psi_1\rangle)^2\right)\Bigg).\end{aligned} \tag{C27}$$

In Fig. 12 we compare the convergence of the weighted state and post-processing methods for taking the linear combination of states. We find that with the exception of taking the linear combination of nearly orthogonal states the convergence is comparable. It is unsurprising that the convergence for nearly orthogonal states is poor for the weighted state method since, as discussed in the main text, this method cannot be applied to perfectly orthogonal states. In contrast for observables composed of a sum of many unitaries (i.e. 'complex' observables in Fig. 12) the weighted state approach has smaller variance than the post-processing method because a large number of distinct circuits need to be run for the latter.

6. Powers of states via the generalized quantum transpose algorithm

The Generalized Quantum Transpose algorithm may be used to prepare the Quantum Hadamard Product of a pair of states $\rho^{(0)}$ and $\rho^{(1)}$ if one knows how to prepare the transpose of either $\rho^{(0)}$ or $\rho^{(1)}$. Moreover, the procedure for preparing the transpose of a state may be straightforwardly computed if the procedure for preparing the original state is known in some detail. Thus GQT, under the right circumstances, provides an alternative means of implementing the QHP and so powers of states. In this section, we include a sampling analysis of this alternative means of implementing powers of states and compare it with the standard method of QHP.

When the GQT circuit is repeatedly applied to implement the power of a state, the variance is given by

$$\mathrm{Var}_{GQT}(\widehat{\mathcal{O}}) = \mathrm{Tr}[D(|\psi\rangle\langle\psi|)O^2] - \left(\langle\psi^k| O |\psi^k\rangle\right)^2 \tag{C28}$$

for any power k. Comparing this to the variance $\mathrm{Var}_{\mathrm{QHP}}[\hat{\mathcal{O}}]$ obtained for the QHP algorithm we have

$$\begin{aligned} D &:= \mathrm{Var}_{\mathrm{QHP}}[\hat{\mathcal{O}}] - \mathrm{Var}_{\mathrm{GQT}}[\hat{\mathcal{O}}] = \langle\psi^k|O^2|\psi^k\rangle - \mathrm{Tr}(D(|\psi\rangle\langle\psi|)O^2) \\ &= \sum_{ij} (\psi_j^*)^k (O^2)_{ji} \psi_i^k - \sum_i |\psi_i|^2 (O^2)_{ii}\,. \end{aligned} \tag{C29}$$

To get a handle on this quantity it is helpful to first consider the case where $O^2 = 1$, as is the case, for example if O is a Pauli operator. In this case, we have

$$D = \langle\psi^k|\psi^k\rangle - 1\,. \tag{C30}$$

Since $\langle\psi^k|\psi^k\rangle \leq 1$ it follows that $\mathrm{Var}_{\mathrm{QHP}}[\hat{\mathcal{O}}] \leq \mathrm{Var}_{\mathrm{GQT}}[\hat{\mathcal{O}}]$, i.e. the convergence of QHP is better than that of GQT.

It is also straightforward to show that when the input states are pure single qubit states we also are guaranteed that $\mathrm{Var}_{\mathrm{QHP}}[\hat{\mathcal{O}}] \leq \mathrm{Var}_{\mathrm{GQT}}[\hat{\mathcal{O}}]$. To see this, let us suppose $|\psi\rangle = \sqrt{1-\epsilon}|0\rangle + \sqrt{\epsilon}|1\rangle$, where we drop a relative phase between $|0\rangle$ and $|1\rangle$ without loss of generality since this phase does not contribute to the variance. For the case of a single application $k = 2$ we have

$$\begin{aligned} D &= \sum_{ij} (\psi_j^*)^2 (O^2)_{ji} \psi_i^2 - \sum_i |\psi_i|^2 (O^2)_{ii} \\ &= (1-\epsilon)^2 (O^2)_{00} + \epsilon^2 (O^2)_{11} + \epsilon(1-\epsilon)((O^2)_{01} + (O^2)_{10}) - (1-\epsilon)(O^2)_{00} - \epsilon(O^2)_{11} \\ &= -\epsilon(1-\epsilon)\left((O^2)_{00} + (O^2)_{11} - (O^2)_{01} - (O^2)_{10}\right)\,. \end{aligned} \tag{C31}$$

Without loss of generality we can write $O = \alpha_I \mathbb{I} + \alpha_x X + \alpha_y Y + \alpha_z Z$ such that $O^2 = (\alpha_I^2 + \alpha_x^2 + \alpha_y^2 + \alpha_z^2)\mathbb{I} + \alpha_I\alpha_x X + \alpha_I\alpha_y Y + \alpha_I\alpha_z Z$. Thus we end up with

$$\begin{aligned} D &= -\epsilon(1-\epsilon)\left(\alpha_I^2 + \alpha_x^2 + \alpha_y^2 + \alpha_z^2 - 2\alpha_I\alpha_x\right) \\ &= -\epsilon(1-\epsilon)\left((\alpha_I + \alpha_x)^2 + \alpha_y^2 + \alpha_z^2\right) \leq 0\,. \end{aligned} \tag{C32}$$

Thus the variance of QHP is less than that of GQT in the case that $k = 2$. It is manifestly clear from Eq. (C29) that increasing k for a given O and $|\phi\rangle$, decreases D and hence D is negative for all k for the case of single qubit states.

From these two examples, the case of a single qubit system and/or Pauli operators O, we suggest that in most cases the QHP converges quicker than the GQT. This is supported by the numerical results shown in Fig. 13. However, of course, we are not quite comparing like for like here since GQT can be used to implement the transpose operation.

7. Concatenation can be inefficient

Consider the task of preparing the weighted state $\tau = \rho^{(0)} \odot \rho^{(1)T}$ from the input state $\rho^{(0)} \otimes \rho^{(1)}$. This can be achieved in two ways: (i) use the generalized quantum transpose circuit, but now treating the ancilla register as part of the input, or (ii) first apply the quantum transpose on $\rho^{(1)}$ (by using the GQT circuit with input $\sigma = |+\rangle\langle+|$ and measurement operator $d\times$SWAP) followed by the quantum Hadamard product with $\rho^{(0)}$. Thus we have two different quantum instruments that implement the same transformation on the inputs: (i) achieves the task directly, and (ii) concatenates two primitives. It is informative to study the sampling complexity of both approaches. We want to estimate the expectation value of some operator O in the weighted state τ given by $\mathrm{Tr}(\tau O)$.

The variance of the estimator for method (i) is already calculated in Eq. (34) and is given by $\mathrm{Tr}(D(\rho^{(0)})O^2) - |\mathrm{Tr}(\tau O)|^2$. The variance of the estimator for method (ii) can be expressed in terms of Eq. (29) if we treat the two concatenated instrument as a single instrument with measurment operator $d(|0\rangle\langle 0| \otimes SWAP)$ and is given by $d^2\,\mathrm{Tr}(D(\rho^{(0)})O^2) - |\mathrm{Tr}(\tau O)|^2$. The additional factor of d^2 drastically favors the first method. This demonstrates that concatenation of primitives can be very costly and it might be advantageous to design instruments that implement a desired transformation directly.

A similar observation can be made with regards to the LCS algorithm. Given a concatenation scheme to prepare the superposition of $L > 2$ states, one can in general find a more efficient way of implementing the same task directly using a single instrument that takes all states as input as described in Section B 1.

Appendix D: Quantum state polynomials realizable without randomization of instruments

In Section. A 4 we have described how quantum instruments with nonnormal measurement operators can thereby be emulated by physical quantum instruments that have normal measurement operators and can be realized in a

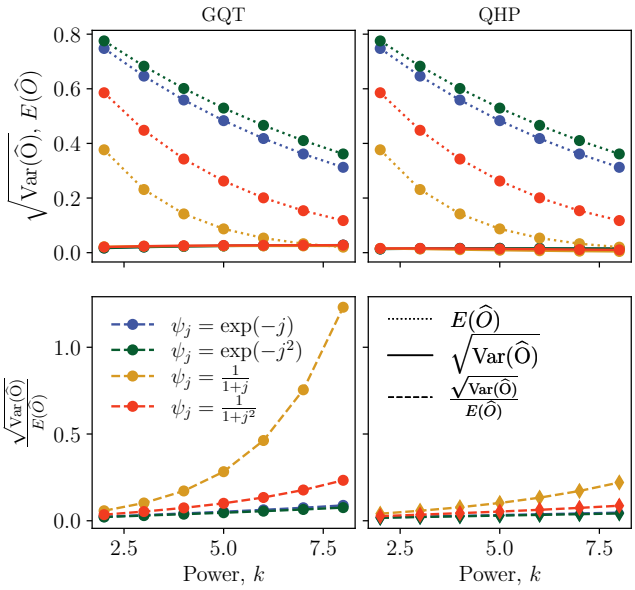

FIG. 13. **Comparing Sampling Error for Powers of States via QHP and GQT.** Here we plot the standard deviation of the estimator $\sqrt{\mathrm{Var}(\widehat{\mathcal{O}})}$ (solid), the actual measurement outcome $\mathrm{E}(\widehat{\mathcal{O}})$ (dotted) and the relative error $\sqrt{\mathrm{Var}(\widehat{\mathcal{O}})}/\mathrm{E}(\widehat{\mathcal{O}})$ (dashed) for the state $|\psi\rangle \propto \sum_j \psi_j |j\rangle$, with the functions ψ_j indicated in the legend, as a function of power k. The GQT (QHP) algorithm is used in the left (right) column. In all cases we consider an $n = 6$ qubit circuit, we measure the all zero projector $O = |0\rangle\langle 0|$ and suppose 1000 shots are used.

quantum computer. This is done by randomizing over quantum instruments. This additional step, however, can increase the sampling complexity, as it adds yet another component to the variance of the estimator. However, in practice there might be other considerations than minimizing the sampling cost. For instance if M can be expressed as the sum of Pauli operators, the circuits for measuring each term can be much shorter than the one for measuring M directly. In such cases the randomized instrument may be preferred in a noisy implementation. Below we classify the class of all transformations we can implement with the QSP algorithm using a single instrument shown in Fig. 6 (i.e. without the additional ancilla as in Fig. 10).

Suppose $\alpha \in \mathbb{C}^{2\times 2}$ contains the coefficients for which that we wish to realize a weighted state. α can be written as $\alpha = H + iS$ where $H = \frac{\alpha+\alpha^\dagger}{2}$ and $S = \frac{\alpha-\alpha^\dagger}{2i}$ are both Hermitian matrices. Since the Pauli matrices, $\mathcal{P} = \{I, X, Y, Z\}$, span the set of Hermitian matrices, we can write $H = \sum_{P\in\mathcal{P}} h_P P$ and $S = \sum_{P\in\mathcal{P}} s_P P$. Note that the coefficients, h_P, s_P, can be found easily from the entries of α.

Let us assume for now that the input states are normalized. We would like to find a density operator σ and normal matrix M such that $\alpha = \sigma \odot M^T$. Since σ is a density operator it can be written as $\sigma = I/2 + (r/2)\boldsymbol{u}\cdot\boldsymbol{\sigma}$, where $\boldsymbol{\sigma} \equiv (X, Y, Z)$, $\|\boldsymbol{u}\| = \|(u_x, u_y, u_z)\| = 1$, and $r \in [0, 1]$. Similarly, by normality of M we can write $M = (aI + \boldsymbol{v}\cdot\boldsymbol{\sigma}) + i(bI + c\boldsymbol{v}\cdot\boldsymbol{\sigma})$ where $a, b, c \in \mathbb{R}$ and $\boldsymbol{v} \in \mathbb{R}^3$. Without loss of generality we let $u_y = 0$, so that $\boldsymbol{u} = (\sin\theta, 0, \cos\theta)$.

Comparing this choice of coefficients with the requirement $\alpha = \sigma \odot M^T$, we find:

$$H = h_I I + h_X X + h_Y Y + h_Z Z = (a + r\cos\theta v_z)I + (r\sin\theta v_x)X - (r\sin\theta v_y)Y + (ar\cos\theta + v_z)Z \tag{D1}$$

$$S = s_I I + s_X X + s_Y Y + s_Z Z = (b + cr\cos\theta v_z)I + (cr\sin\theta v_x)X - (cr\sin\theta v_y)Y + (br\cos\theta + cv_z)Z \tag{D2}$$

It can easily be shown that choosing σ to be pure is more general than choosing a mixed state, so we will assume $r = 1$. We also only consider the case where α is not diagonal, since this can be dealt with easily. Note that this condition will necessarily require $\sin\theta$ and $\cos\theta$ to be nonzero.

Equations (D1) and (D2) represent 8 equations in 7 variables, a, b, c, v_x, v_y, v_z, and θ. By comparing the equations for h_X, h_Y, s_X, and s_Y, we immediately see that $s_x = ch_x$ and $s_y = ch_y$. These imply that $c = s_x/h_x = s_y/h_y$, which is easily shown to be equivalent to $|\alpha_{01}| = |\alpha_{10}|$. We can satisfy these four equations by setting

$$v_x = \frac{2h_x}{\sin\theta}, \qquad v_y = \frac{-2h_y}{\sin\theta}. \tag{D3}$$

We now consider two cases.

Case 1: $\alpha = e^{i\phi}\hat{H}$ **for some Hermitian** $\hat{H}$. By Euler's formula, we know that $\alpha = \cos\phi\hat{H} + i\sin\phi\hat{H} = H + iS$. Thus, for every $i \in \{I, X, Y, Z\}$ it is implied that $s_i = \tan\phi \cdot h_i$ (in fact, the assumption is equivalent to the existence of a scalar, k, such that $s_i = k \cdot h_i$ for every i). By setting $b = a\tan\phi$ and $c = \tan\phi$ we can satisfy the s_I and s_Z equations. By simple substitutions in the h_I and h_Z equations, we find that

$$a = 2h_0 - \cos\theta v_z, \quad v_z = \frac{2h_z - 2h_0\cos\theta}{1 - \cos^2\theta}. \tag{D4}$$

We have solutions for a, b, c, v_x, v_y, and v_z, and all equations are satisfied, but θ is still a free parameter. Thus in the case $\alpha = e^{i\phi} \cdot \hat{H}$ for some Hermitian $\hat{H}$, we have an infinite choice of ancilla states by setting θ to be any non-multiple of π.

Note: the assumption in this case is a generalization of both Hermitian and skew-Hermitian matrices. Suppose $A \in \mathbb{C}^{2\times 2}$ satisfies $A = e^{i\phi}\hat{H}$. This occurs if and only if $A^\dagger = e^{-i\phi}\hat{H} = e^{-2i\phi}e^{i\phi}\hat{H} = e^{-2i\phi}A$. Note that if $\phi = 0$ then A is Hermitian, and if $\phi = \pi/2$ then A is skew-Hermitian.

Case 2: $\alpha \neq e^{i\phi} \cdot \hat{H}$ **for any Hermitian** $\hat{H}$. By solving the equations for h_I and s_I for a and b and substituting these values into the equations for h_Z and s_Z, we find that

$$\cos\theta = R(\alpha) \equiv \frac{s_z - ch_z}{s_0 - ch_0} \tag{D5}$$

which is well defined since $c = s_0/h_0$ would imply a satisfying assignment such that $c = s_z/h_z$, which contradicts the assumption. Since $\cos\theta$ must be strictly between -1 and 1 (α is not diagonal), if $R(\alpha) \notin (-1, 1)$ then we will not be able to satisfy (D5), implying that no ancilla states exist for $\alpha = \sigma \odot M^T$. However, in the case $R(\alpha) \in (-1, 1)$ we can set $\theta = \pm\arccos[R(\alpha)]$, fixing θ and satisfying (D5).

Since θ is now fixed, the equations for h_I, s_I, h_z, and s_Z represent a linear system of 4 equations in three variables a, b, and v_z, which can be written as

$$\begin{bmatrix} 1 & 0 & \cos\theta \\ 0 & 1 & c\cos\theta \\ \cos\theta & 0 & 1 \\ 0 & \cos\theta & c \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ v_z \end{bmatrix} = \begin{bmatrix} h_0 \\ s_0 \\ h_z \\ s_z \end{bmatrix}, \tag{D6}$$

which we will abbreviate as $A\boldsymbol{x} = \boldsymbol{y}$. Note that given $\boldsymbol{y}$ this equation will have a unique solution if and only if A is non-singular, i.e. $\det(A^T A) \neq 0$. By direct calculation $\det(A^T A) = (1 + c^2)(1 - \cos^2\theta)^4 \neq 0$, thus we will be able to determine satisfying assignments for a, b, and v_z by Gaussian elimination.

For this case we see that exactly two pure ancilla states exist for $\alpha = \sigma \odot M^T$ whenever $|\alpha_{01}| = |\alpha_{10}|$ and $R(\alpha) \in (-1, 1)$, and no ancilla state (pure or mixed) exists otherwise.

Appendix E: Hardware implementation

We implemented our algorithms for performing nonlinear operations on IBMQ-Bogota. In Fig. 14 we plot the results of the hardware implementation (dotted) as compared to the ideal output of the algorithm in the absence of hardware noise (dashed) and the ideal output computed classically (solid). The powers of state algorithm (via QHP) works reasonably well for squaring a state, i.e. $k = 2$, but performs more poorly for higher powers. We suggest that this is because qubit resets are currently rather noisy. We find that on current hardware the post-processing method for implementing a linear combination of states substantially outperforms the weighted state method. This is perhaps unsurprising given that a controlled swap operation, requiring 18 CNOTs, is required for the weighted states algorithm and the CNOT error on IBMQ-Bogota is of the order 10^{-2}. More generally, while the hardware results broadly reproduce the expected trends, moderate errors are observed. Thus we expect these algorithms to prove more useful as we approach the fault tolerant era.

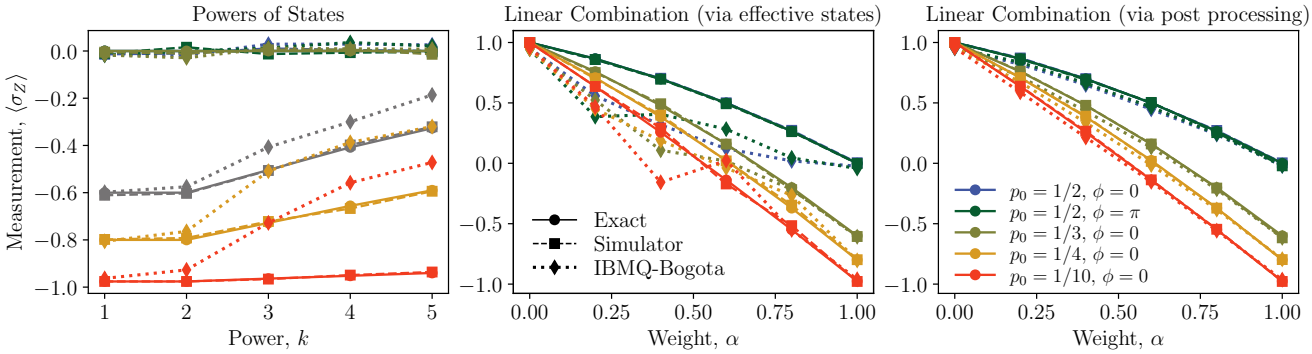

FIG. 14. **Hardware Implementation.** On the left we plot $\langle \psi(p_0,\phi)^k | \sigma_z | \psi(p_0,\phi)^k \rangle$ where $|\psi(p_0,\phi)\rangle := \sqrt{p_0}\,|0\rangle + \exp(-i\phi)\sqrt{1-p_0^2}\,|1\rangle$ and ψ^k is computed using QHP on a simulator (dashed line) and on IBMQ-Bogota (dotted line). The solid line indicates the exact expectation value computed classically. In the centre (right) plot we plot $\langle \Psi_\alpha(p_0,\phi) | \sigma_z | \Psi_\alpha(p_0,\phi) \rangle$ where the linear combination of states $|\Psi_\alpha(p_0,\phi)\rangle := \alpha\,|0\rangle + (1-\alpha)\,|\psi(p_0,\phi)\rangle$ is computed using the weighted states method (post-processing method). In both the hardware and noiseless simulations $s = 8000$ shots are used.