

Preparing low-variance states using a distributed quantum algorithm

Xiaoyu Liu^{1,2}, Benjamin F. Schiffer³, and Jordi Tura^{1,2}

¹ $\langle aQa^L \rangle$ Applied Quantum Algorithms, Universiteit Leiden

²Instituut-Lorentz, Universiteit Leiden, P.O. Box 9506, 2300 RA Leiden, The Netherlands

³Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Str. 1, D-85748 Garching, Germany

Quantum computers are a highly promising tool for efficiently simulating quantum many-body systems. The preparation of their eigenstates is of particular interest and can be addressed, e.g., by quantum phase estimation algorithms. The routine then acts as an effective filtering operation, reducing the energy variance of the initial state. In this work, we present a distributed quantum algorithm inspired by iterative phase estimation to prepare low-variance states. Our method uses a single auxiliary qubit per quantum device, which controls its dynamics, and a postselection strategy for a joint quantum measurement on such auxiliary qubits. In the multi-device case, the result of this measurement heralds the successful runs of the protocol. This allows us to demonstrate that our distributed algorithm reduces the energy variance faster compared to single-device implementations, thereby highlighting the potential of distributed algorithms for near-term and early fault-tolerant devices.

1 Introduction

Preparing eigenstates of Hamiltonians plays an essential role in quantum simulation to investigate the properties of quantum many-body systems. Quantum computers are naturally well-suited for this task, with quantum phase estimation being the canonical method for eigenstate preparation [1, 2, 3, 4]. Obtaining precise

eigenstates then allows one to accurately compute their energies and measure other observables that provide insights into the behavior of complex many-body systems.

Quantum phase estimation (QPE) [1, 6] and filtering [7, 8] are closely related methods to prepare low-variance states on quantum computers. For a Hamiltonian of interest H , the algorithm relies on conditional dynamics e^{-iHt} that realizes phase kick-back to the controlling auxiliary qubit. Another interesting approach is to prepare the filtered state virtually by measuring Loschmidt echoes [9], or adiabatically by a parent Hamiltonian construction [10]. All approaches have in common that the system is initialized in a state that can be efficiently prepared. Typically, this initial state is a product state that will generally exhibit a Gaussian-like distribution in the eigenbasis of a local Hamiltonian [11]. A filter algorithm acts on the initial state by sharpening the energy distribution towards a delta function, as shown in Fig. 1(a). In the limit of an arbitrarily precise filter centered at an eigenenergy of the Hamiltonian, the operation prepares the micro-canonical ensemble by isolating the eigenstate(s) at that energy.

A particularly simple quantum routine for filtering is inspired by iterative quantum phase estimation (IQPE) [6, 12, 13, 14, 15, 16, 17, 18]. This algorithm applies a Hadamard-test circuit [19] for each iteration by implementing a time evolution operator controlled on a single auxiliary qubit. The auxiliary qubit can then be reused for the next iteration. Conversely, textbook QPE relies on a sequence of conditional dynamics (controlled unitary evolutions) with evolution times chosen such that one may obtain one additional digit of the phase with each auxiliary qubit. This precision requirement can be relaxed by choosing

Xiaoyu Liu: xiaoyu@lorentz.leidenuniv.nl

Benjamin F. Schiffer: benjamin.schiffer@mpq.mpg.de

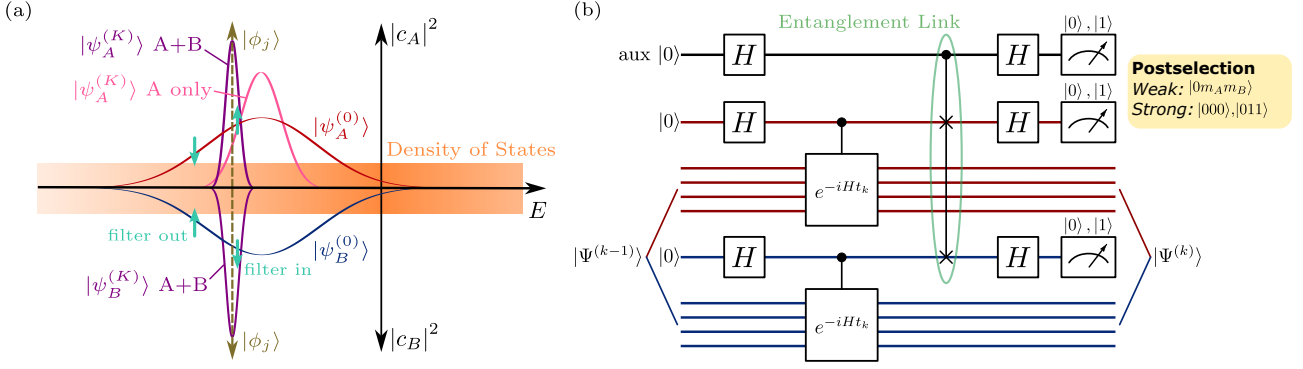


Figure 1: (a) Illustration of a filtering operation. Consider a local Hamiltonian H , which has an approximate Gaussian-shaped density of states centered at $\text{Tr}(H)$ [5]. The devices Alice (A) and Bob (B) start by preparing an initial state, with respective populations $|c_A|^2$ and $|c_B|^2$. After multiple iterations of the protocol, the state becomes increasingly concentrated. The distributed version of this algorithm uses an entanglement link between A and B to enhance the filtering process, resulting in an output state with a lower energy variance on average for the same number of iterations. Note that an energy bias may be introduced during the postselection routine. (b) Circuit for the distributed filtering algorithm. Both A and B are initialized in the same state $|\Psi^{(0)}\rangle = |\psi^{(0)}\rangle^{\otimes 2}$. In each iteration k , a random time evolution for a sufficiently large t_k is applied to both controlled unitaries, followed by a control-SWAP gate, realizing the entanglement link. We analyze two cases (*weak* and *strong*) for postselecting the measurement results of the auxiliary qubits.

the evolution times randomly from a sufficiently large time interval to shuffle the eigenphases, as we explain further below. For this protocol, it is known that the energy variance of the output state decreases with each iteration on average, generically converging to an eigenstate. However, reaching low-variance states may still require many iterations, demanding qubits with long coherence times.

A promising strategy to reduce the number of rounds — and consequently, the circuit depth — in this IQPE-based protocol is to leverage distributed quantum computing. Distributed quantum algorithms can address large-scale problems with shallower circuits by utilizing interconnected quantum devices that can share entanglement [20, 21, 22, 23]. In fact, such a distributed approach has been applied to eigenstate preparation, as demonstrated in [16]. In that work, an eigenstate broadcasting scenario was investigated with two quantum devices: a first device Alice, with an almost perfect eigenstate, and a second device Bob, with a rougher approximation to the target ground state. The distributed circuit allows Bob to speed up his preparation via implicit knowledge of the target state through a shared entanglement link with Alice.

In this work, we analyze a similar setup, focusing on its application to quantum state filtering: both Alice and Bob start with the same product state and iteratively converge to states

with a lower energy variance. We first propose a two-device distributed filtering algorithm where joint quantum measurements between the devices enable Alice and Bob to control the moments of the energy distribution of the initial state. We focus on the first and second moments, i.e. the energy and variance of the state. Concretely, we show that the distributed filtering algorithm reaches a low-variance state more rapidly on average than the single-device case. Postselection on the measurement outcome is an essential ingredient in the protocol and careful engineering of the postselection allows one to further optimize the state preparation process. The postselection introduces either a positive or negative energy bias. We provide a bound on the maximum bias introduced in the protocol. Finally, we generalize our algorithm to a larger number of connected quantum devices that jointly perform the filtering operation. Numerical evidence shows an enhanced eigenstate preparation as the number of devices is increased.

This paper is structured as follows: in Section 2, we introduce the filtering algorithm and provide a detailed explanation of the distributed algorithm, including its pseudocode. We analyze the performance of the two-device algorithm through numerical experiments on several product state instances, focusing on the behavior of the average variance and energy over iterations. The extension of the distributed architecture to

a setup with more than two devices is found in Section 3, including numerical experiments benchmarking the performance of two and three devices against a single device. Finally, in Section 4, we summarize and discuss our findings.

2 Distributed filtering with auxiliary qubits

The single-device setup was previously introduced under different names in [6, 13, 15, 14, 16], where Hadamard tests [19] are applied iteratively using the unitary evolution operator $U = e^{-iHt_k}$ for each iteration k . In each iteration, this unitary is controlled by a single local auxiliary qubit, which is initialized in $|+\rangle$ and measured in the $|\pm\rangle$ basis. The details of the single-device algorithm are restated in Appendix A.

To extend this setup to two devices, we use the configuration shown in Fig. 1(b). Here, two identical Hadamard test circuits are executed in parallel on both devices Alice and Bob. The two devices are connected through their respective control qubits to a single additional helper qubit “aux”. This allows quantum communication between the devices using a circuit that is effectively a swap test circuit on the two control qubits [24, 25], where the control-SWAP gate acts as the entanglement link, acting non-locally only on the spatially separated auxiliary qubits. For simplicity, we assume that Alice possesses the top auxiliary qubit (“aux” in Fig. 1). Then, for each iteration, only one Bell pair generated between Alice and Bob is required to teleport Bob’s single local auxiliary qubit and the control-SWAP gate can be applied locally. The three qubits are finally measured in the $|\pm\rangle$ -basis. We detail the realization of the distributed circuit with shared Bell pairs between the two devices in Appendix D.

In the following sections, we present a detailed description of the distributed filtering algorithm and demonstrate its ability to control the moments of quantum states. We focus our analysis on the first and second moment — energy E and variance σ^2 — defined as

$$E = \langle \psi | H | \psi \rangle, \quad (1)$$

$$\sigma^2 = \langle \psi | (H - E)^2 | \psi \rangle = \langle \psi | H^2 | \psi \rangle - E^2, \quad (2)$$

for a pure state $|\psi\rangle$. Note that for eigenstates of H we have $\sigma^2 = 0$.

2.1 Algorithm outline

Recall that the objective of the filtering algorithm is to prepare a state with lower variance in the eigenbasis of a given Hamiltonian H than the initial (product) state. In the distributed setup, both Alice and Bob start with identical product states: $|\psi_A\rangle \otimes |\psi_B\rangle = |\psi^{(0)}\rangle^{\otimes 2}$. They each input their state locally into the circuit shown in Fig. 1(b) and run the circuit iteratively. At each iteration, two identical controlled unitaries e^{-iHt_k} are applied locally. The time t_k for each iteration is randomly chosen from $(0, T]$, such that $\varphi_j^{(k)} = -t_k \lambda_j \pmod{2\pi} \sim \text{Uniform}(0, 2\pi)$ for every j , to effectively separate the eigenstates. Here λ_j is the j th eigenvalue of the Hamiltonian H . This randomization is needed because, without prior knowledge of the initial product state’s populations in the eigenbasis, each iteration requires a sufficiently large time evolution t_k to effectively separate the eigenstates from each other by randomizing the eigenphases $\varphi_j^{(k)} = -t_k \lambda_j$. Note that both Alice and Bob use the same t_k in each iteration to ensure that the output state remains symmetric across Alice and Bob. This condition is necessary for both parties to converge to the same eigenstate, heralded by postselection. This synchronization can be efficiently achieved by pre-sharing a list of random t_k values generated from a common random seed. Further details on this randomization of the phases are provided in Appendix A.2.

Additionally, the distributed filtering algorithm depends on postselection of the measured auxiliary qubits, which can be performed using either a *weak* postselection criterion (where we continue to the next iteration only when top auxiliary qubit outcome is $|0\rangle$), or a *strong* criterion (where we continue to the next iteration only when the outcome of the three auxiliary qubits is $|000\rangle$ or $|011\rangle$). These two types of postselection exhibit different properties and performance, which we explore later. If any iteration yields measurement outcomes that fall outside the postselection criteria, the protocol is restarted, as such outcomes on average broaden the state distribution, preventing the operation from effectively working as a filter. The pseudocode for the distributed filtering algorithm is shown in Algorithm 1.

For the *strong* postselection, $|\Psi_{AB}^{(K)}\rangle$ will have

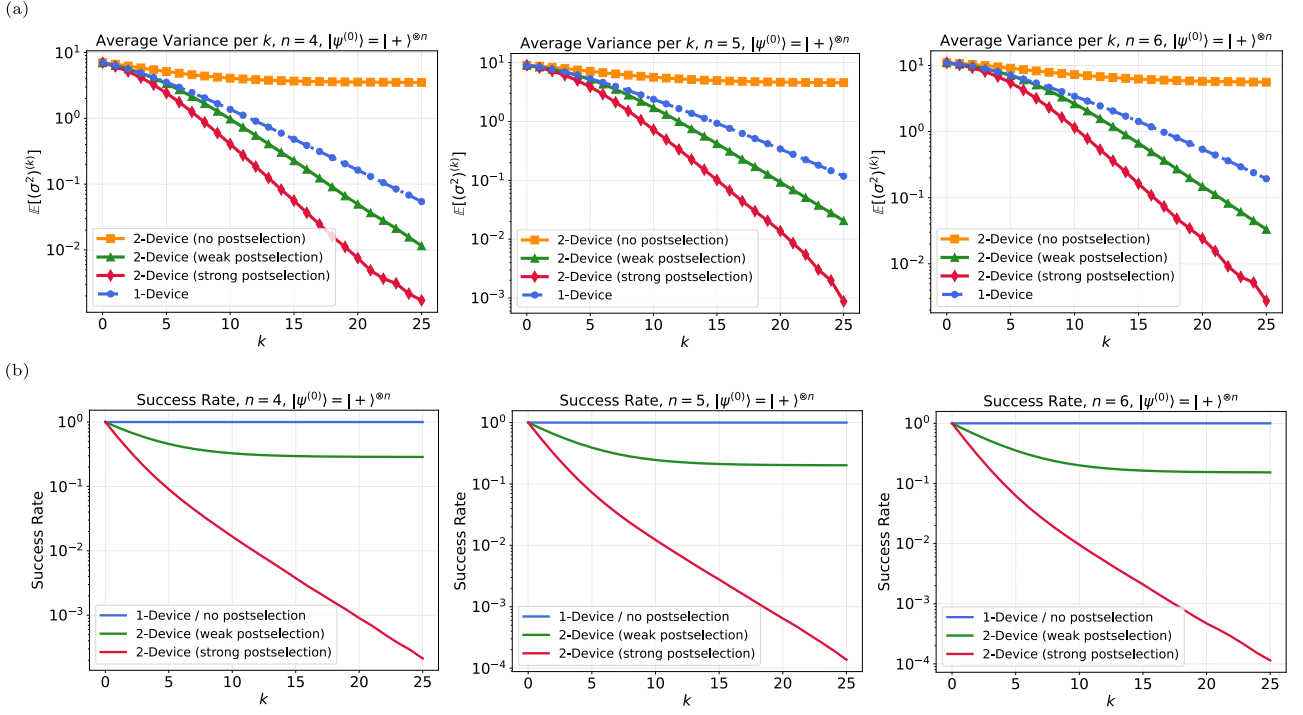


Figure 2: (a) Average variance of the pure state ensemble per iteration k up to 25. Three cases ($n = 4, 5, 6$, $|\psi^{(0)}\rangle = |+\rangle^{\otimes n}$) are shown. For cases requiring postselection, only states meeting the postselection criteria are retained at each iteration k . The cumulative success rate after postselection at each iteration are shown in (b). Note that the single-device algorithm does not require postselection, thus resulting in a 100% success rate. We use 10^8 trials from the beginning for *strong* postselection cases and 10^6 for the rest to acquire accurate enough expectation values. More numerical results starting with a fixed number of repetitions of the algorithm with complete error bars are shown in the Appendix E.

a tensor product structure across Alice and Bob. If we start with two copies of the same state, both Alice and Bob will hold the same final state. On the other hand, for the *weak* postselection, the state will be entangled in general, but for large values of K , it will approximate a product of identical states. This is because both postselection criteria select the top qubit to be measured in the $|0\rangle$ state, the state registers of both Alice and Bob are projected to their joint symmetric subspace in every iteration. As a result, tracing out either Alice or Bob's systems will result in the same states on both devices. If the minimal distance between eigenvalues of the Hamiltonian H is $\Omega(1/t_k)$, then the final states will be eigenstates of H . Our distributed protocol therefore allows one to prepare multiple identical copies of (approximate) eigenstates in a heralded way, and we include the details in Appendix B.2.

2.2 Faster reduction of the variance

We now analyze the performance of the distributed filtering algorithm. The variance of the

pure state is defined in Eq. (1). Due to the inherent randomness of the algorithm (random t_k and the measurement process), the output state after a fixed number of iterations can differ. Consequently, we compute the average variance of the pure state ensemble at iteration k , denoted $\mathbb{E}[(\sigma^2)^{(k)}]$, over a sufficiently large number of repeated runs. Analytically, we write

$$\begin{aligned} & \mathbb{E}[(\sigma^2)^{(k)}] \\ &= \mathbb{E}\left[\langle \psi^{(k)} | H^2 | \psi^{(k)} \rangle\right] - \mathbb{E}\left[\left(\langle \psi^{(k)} | H | \psi^{(k)} \rangle\right)^2\right] \\ &= \text{Tr}(\rho^{(k)} H^2) - \mathbb{E}\left[\left(\langle \psi^{(k)} | H | \psi^{(k)} \rangle\right)^2\right], \end{aligned} \quad (3)$$

where $|\psi^{(k)}\rangle$ denotes the output state at iteration k , and $\rho^{(k)}$ represents the density matrix describing a mixture of states $\{|\psi^{(k)}\rangle\}$ with the corresponding probabilities.

Algorithm 1 Two-device distributed filtering algorithm

Require: Two product states $|\psi\rangle \otimes |\psi\rangle = |\psi^{(0)}\rangle^{\otimes 2}$; target iteration K ; postselection type $\mathcal{P}$; Hamiltonian H with a generic spectrum

Ensure: State $|\Psi_{AB}^{(K)}\rangle$.

- 1: Initialize: $|\Psi_{AB}^{(0)}\rangle \leftarrow |\psi^{(0)}\rangle^{\otimes 2}$;
 - 2: **for** $k = 1 : K$ **do**
 - 3: Generate random time t_k such that $\varphi_j^{(k)} = -t_k \lambda_j \bmod 2\pi \sim \text{Uniform}(0, 2\pi)$ for every j ;
 - 4: Input $|\Psi_{AB}^{(k-1)}\rangle$ and t_k to the circuit in Fig. 1(b);
 - 5: Measure and get the result $|m_k\rangle = |m_0^{(k)} m_A^{(k)} m_B^{(k)}\rangle$;
 - 6: **if** $\mathcal{P} = \text{weak}$ **and** $m_0^{(k)} = 0$ **then**
 - 7: Output $|\Psi_{AB}^{(k)}\rangle$ and **continue**;
 - 8: **else if** $\mathcal{P} = \text{strong}$ **and** $(m_k = 000 \text{ or } m_k = 011)$ **then**
 - 9: Output $|\Psi_{AB}^{(k)}\rangle$ and **continue**;
 - 10: **else**
 - 11: **break** and start over from initialization;
 - 12: **end if**
 - 13: **end for**
 - 14: **return** $|\Psi_{AB}^{(K)}\rangle$.
-

We numerically compute the variance of the pure state ensemble and consider the n -qubit initial input state $|\psi^{(0)}\rangle = \sum_j c_j^{(0)} |\phi_j\rangle$, where the quantities $|\phi_j\rangle$ and $c_j^{(0)}$ denote the j th eigenstate of the Hamiltonian H and its corresponding amplitude, respectively. Throughout this paper, we shall use the non-integrable, one-dimensional Ising model with both a longitudinal and a transverse field as a running example:

$$H = \sum_{j=1}^{n-1} \sigma_j^z \sigma_{j+1}^z + \sum_{j=1}^n (\sigma_j^x + \sigma_j^z). \quad (4)$$

In Fig. 2 we present numerical results for $n = 4, 5, 6$ and $|\psi^{(0)}\rangle = |+\rangle^{\otimes n}$. We compare the performance of the distributed filter with the single-device filter. For scenarios requiring postselection, states that do not meet the postselection criterion are discarded. The results show that the distributed filter with postselection achieves a lower average energy variance, outperforming the

single-device filter. In contrast, the distributed filter without postselection, where the algorithm continues regardless of measurement outcomes, performs worse than the single-device case. The *strong* criterion for postselection yields the fastest decrease in the average energy variance.

Nevertheless, the distributed filter incurs an overhead due to postselection. Fig. 2(b) illustrates how this overhead scales with the number of iterations k for both *weak* and *strong* postselection types. We calculate the cumulative success rate after each postselection, i.e., the average probability of the auxiliary qubit(s) result meeting the corresponding postselection criterion at each k . Notably, in the case of *weak* postselection, the cumulative success rate tends to stabilize after a few iterations. This implies that once the protocol succeeds in several consecutive rounds, it is highly likely to continue succeeding. Specifically, the average cumulative success rate of the *weak* postselection is lower bounded:

$$\begin{aligned} P_w &= \sum_j |c_j^{(0)}|^4 + \left(\frac{3}{4}\right)^k \sum_{j \neq j'} |c_j^{(0)}|^2 |c_{j'}^{(0)}|^2 \\ &\geq \sum_j |c_j^{(0)}|^4. \end{aligned} \quad (5)$$

Although this lower bound depends only on the the eigenbasis amplitudes of $|\psi^{(0)}\rangle$ which in principle can be any state, for n -qubit product state this lower bound indeed has a scaling of the form $A \exp(B/n)$ with A and B coefficients independent of n , when the Hamiltonian H is local. This lower bound decreases mildly with increasing n , which supports the practicality of the *weak* postselection protocol for larger system sizes. For *strong* postselection, however, this quantity decreases exponentially as:

$$P_s = \left(\frac{3}{4}\right)^k \sum_j |c_j^{(0)}|^4 + \left(\frac{1}{2}\right)^k \sum_{j \neq j'} |c_j^{(0)}|^2 |c_{j'}^{(0)}|^2. \quad (6)$$

Both derivations are included in Appendix B.2. Due to the overhead introduced by postselection, additional resource costs will be incurred, such as an increased number of time evolutions and greater consumption of Bell pairs. These costs can be evaluated from the cumulative success rate, and for *weak* postselection they scale linearly with k when the cumulative success rate plateaus, but the ones for *strong* postselection

scale exponentially. Thus, there is a trade-off between the two postselection types: *strong* postselection reduces the energy variance more quickly than *weak* postselection but suffers from an exponential postselection overhead.

2.3 Spreading of eigenstates

In the previous section, we considered the average energy variance of the pure state prepared in the algorithms, which asymptotically goes to zero in the limit of many iterations. The probability with which we converge to a specific eigenstate, though, is related to the initial populations $\{|c_j^{(0)}|^2\}$. This is another variance and can be interpreted as the variance of the mixed state $\rho^{(k)}$ that describes the expected physical system after k iterations. To avoid confusion between the two types of variances, we denote it as the *eigenstate spread*. Effectively, it is a measure over what energy range the eigenstates prepared in the algorithm are distributed. We can also see the difference between the variance of the pure states and the eigenstate spread, by looking at the analytic description of the eigenstate spread

$$\begin{aligned} V^{(k)} &= \text{Tr}(\rho^{(k)} H^2) - \mathbb{E}[E^{(k)}]^2 \\ &= \text{Tr}(\rho^{(k)} H^2) - \left(\text{Tr}(\rho^{(k)} H)\right)^2, \end{aligned} \quad (7)$$

where the second term differs from Eq. (3).

We show the eigenvalue spread in Fig. 3 and observe that the over several iterations of the protocol the spread decreases. For both of the postselection scenarios, the eigenvalue spread converges to a fixed value:

$$V_{k \rightarrow \infty} = \frac{\sum_j \lambda_j^2 |c_j^{(0)}|^4}{\sum_j |c_j^{(0)}|^4} - \left(\frac{\sum_j \lambda_j |c_j^{(0)}|^4}{\sum_j |c_j^{(0)}|^4} \right)^2. \quad (8)$$

This corresponds to the smallest populations having been suppressed. We observe the eigenvalue spread converge faster towards this value for the *strong* postselection criterion.

We can find an analytical expression for the asymptotic behavior of the eigenstate spread by considering a smooth distribution for the amplitudes of the Hamiltonian eigenvalues and also for the initial populations of the state. Concretely, we assume H is traceless, local and has a Gaussian density of states centered at zero with variance σ^2 (cf. [5]). The input product state

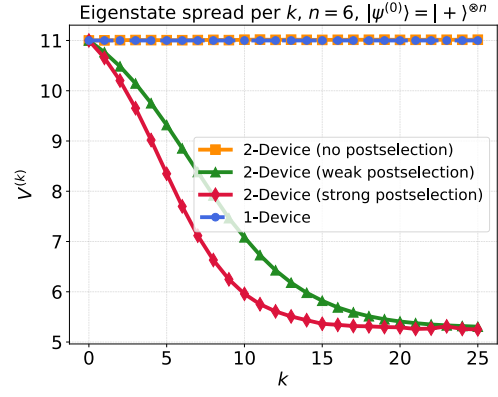


Figure 3: Eigenvalue spread $V^{(k)}$ for $n = 6$ and $|\psi^{(0)}\rangle = |+\rangle^{\otimes n}$. Note that the eigenvalue spread goes to a fixed lower bound under the postselection circumstances.

$|\psi^{(0)}\rangle$ is assumed to have eigenstate populations with Gaussian shape centered at energy μ and with variance ξ^2 (cf. [11]). The eigenstate spread is a function of the initial distributions, the difference in the limit of many iterations $k \rightarrow \infty$ from the initial eigenvalue spread is described by the expression

$$V^{(0)} - V_{k \rightarrow \infty}^{(k)} = \frac{1}{\left(\frac{\xi^2}{\sigma^2} + 1\right) \left(\frac{\xi^2}{\sigma^2} + 2\right)} \xi^2. \quad (9)$$

In the limit of an extremely narrow energy distribution of the eigenstates ($\xi \rightarrow 0$), the eigenvalue spread does not change anymore during the protocol. We provide a rigorous derivation of the expressions Eq. (8) and Eq. (9) in the Appendix B.3 for the *weak* postselection criterion. For the *strong* postselection criterion, we numerically observe that the same expression holds.

We remark that the ensemble $\rho^{(k)}$ is special in the sense that we know, in the large number of rounds limit, it is generated as a mixture of eigenstates of the Hamiltonian, but in a smaller energy window than that of the initial product state. The smaller window is a structural property that depends on the density of states and the initial state distribution.

2.4 Energy bias analysis

We now turn to the energy behavior of the resulting states, focusing on the ensemble average of the pure state energy at iteration k , given by

$$\mathbb{E}[E^{(k)}] = \mathbb{E}[\langle \psi^{(k)} | H | \psi^{(k)} \rangle] = \text{Tr}(\rho^{(k)} H). \quad (10)$$

Using the same setup as above, we have shown the average energy obtained in numerical simulations in Fig. 4. It follows from the Born rule [26], that for both the single-device and two-device algorithms without postselection, the average energy remains constant, matching the energy of the input state $|\psi^{(0)}\rangle$:

$$\mathbb{E}[E^{(k)}] = \sum_j \lambda_j |c_j^{(0)}|^2, \quad (11)$$

where λ_j is the energy of the j th eigenstate of H . However, for the two distributed cases with postselection, energy biases are introduced in this process. Specifically, with the postselection, one can show that in the limit of many iterations k

$$\mathbb{E}[E^{(k)}]_{k \rightarrow \infty} \rightarrow \left(\sum_j \lambda_j |c_j^{(0)}|^4 \right) / \left(\sum_j |c_j^{(0)}|^4 \right), \quad (12)$$

effectively bounding the energy bias introduced. Hence, the eigenenergies of the Hamiltonian and the initial state distribution fully determine the expected converged energy. The effective effect can either drive the system towards an effective higher or lower average energy. Again, similar to the energy spread, $\mathbb{E}[E^{(k)}]$ with the *strong* postselection criterion shows faster convergence.

We provide an additional analysis of the energy bias for the continuous case in Appendix B.1, where both the density of states of the local H and the populations of $|\psi^{(0)}\rangle$ have Gaussian shapes. Again, we assume H is traceless, local and has a Gaussian density of states centered at zero with variance σ^2 . As in the previous section, the input product state $|\psi^{(0)}\rangle$ is assumed to have eigenstate populations with Gaussian shape centered at energy μ and with variance ξ^2 . Then, we show that the largest bias we can introduce is

$$\begin{aligned} |E^{(0)} - \mathbb{E}[E^{(k)}]_{k \rightarrow \infty}| &= \left| \frac{\mu}{\frac{\xi^2}{2\sigma^2} + 1} - \frac{\mu}{\frac{\xi^2}{\sigma^2} + 1} \right| \\ &= |\mu| \left(\frac{1}{\frac{\xi^2}{2\sigma^2} + 1} - \frac{1}{\frac{\xi^2}{\sigma^2} + 1} \right) \\ &\leq |\mu| \frac{1}{2\sqrt{2} + 3} \end{aligned} \quad (13)$$

The full derivation is included in Appendix B.1.2.

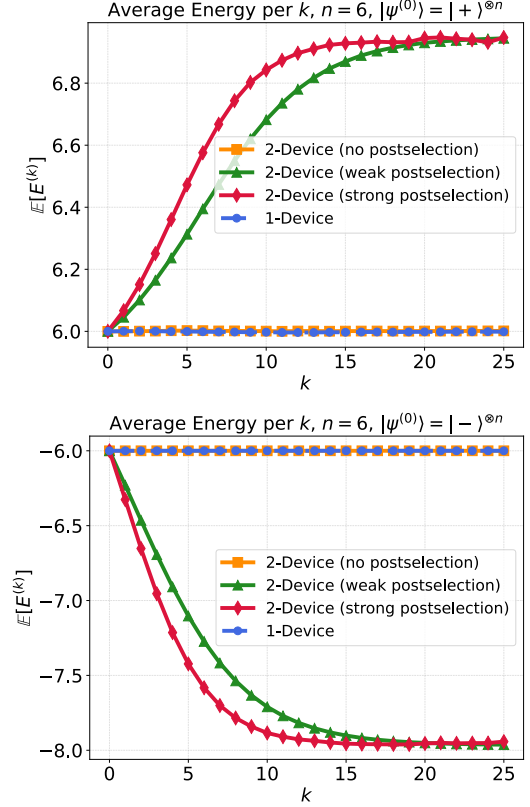
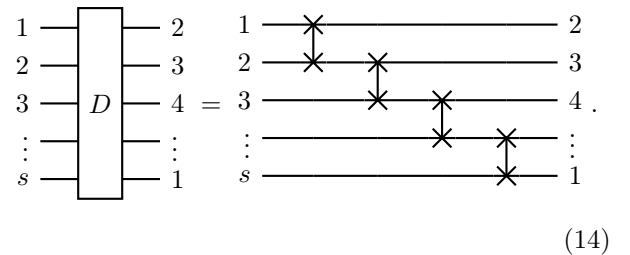


Figure 4: Average energy of the pure state ensemble per iteration k for $n = 6$, $|\psi^{(0)}\rangle = |+\rangle^{\otimes n}$ and $|-\rangle^{\otimes n}$. Energy biases are introduced in cases with postselection. Note that the energy bias can be either positive or negative, depending on the sign of $\mathbb{E}[E^{(k)}]_{k \rightarrow \infty}$. Additional numerical simulations with complete error bars are shown in the Appendix E.

3 Multi-device extension

We now show how to extend the distributed filter method to an arbitrary number of devices. A natural approach is to generalize the swap operator S to a derangement operator D . One way to achieve this is shown by the following circuit:



However, this generalization does not act as an effective filtering operation, as briefly mentioned in [16]. Instead, we consider generalizing the swap test circuit to a cyclic permutation test [27, 28, 29], depicted in Fig. 5(a). To apply the filter

across s ($s \geq 2$) devices, the top auxiliary qudit must have s levels. The operator F performs the discrete Fourier transform, defined as: $|z\rangle \rightarrow \frac{1}{\sqrt{s}} \sum_{q=0}^{s-1} \omega^{zq} |q\rangle$ where $\omega = e^{2\pi i/s}$. The multi-level controlled- D operator $\sum_{q=0}^{s-1} |q\rangle \langle q| \otimes D^q$ is then applied as follows:

$$\sum_{q=0}^{s-1} a_q |q\rangle \text{ --- } \begin{array}{c} \bullet \\ \text{---} \text{---} \text{---} \\ \boxed{D} \end{array} \text{ --- } = \sum_{q=0}^{s-1} a_q |q\rangle \otimes D^q |\Psi\rangle, \quad (15)$$

where $a_q \in \mathbb{C}$ and $\sum_{q=0}^{s-1} |a_q|^2 = 1$. An inverse discrete Fourier transform $F^\dagger$ is then applied before measuring the top qudit. Similar to the two-device filter, we extend the postselection criteria correspondingly, and one can see that *strong* selection can still make the states of each party remain product. In addition, *weak* postselection makes the state in general entangled but it evolves asymptotically to a product of identical eigenstates when K is large. The pseudocode for the s -device filter is shown in Algorithm 2.

Algorithm 2 s -device distributed filtering algorithm

Require: s product states $|\psi_1\rangle \otimes \cdots \otimes |\psi_s\rangle = |\psi^{(0)}\rangle^{\otimes s}$; target iteration K ; postselection type $\mathcal{P}$.

Ensure: State $|\Psi_{12\dots s}^{(K)}\rangle$.

- 1: Initialize: $|\Psi^{(0)}\rangle \leftarrow |\psi^{(0)}\rangle^{\otimes s}$;
 - 2: **for** $k = 1$ to K **do**
 - 3: Generate random time t_k such that $\varphi_j^{(k)} = -t_k \lambda_j \bmod 2\pi \sim \text{Uniform}(0, 2\pi)$ for every j ;
 - 4: Input $|\Psi^{(k-1)}\rangle$ and t_k to the circuit in Fig. 5(a);
 - 5: Measure and get the result $|m_k\rangle = |m_0^{(k)} m_1^{(k)} \cdots m_s^{(k)}\rangle$;
 - 6: **if** $\mathcal{P} = \text{weak}$ **and** $m_0^{(k)} = 0$ **then**
 - 7: Output $|\Psi^{(k)}\rangle$ and **continue**;
 - 8: **else if** $\mathcal{P} = \text{strong}$ **and** $(m_k = 00 \cdots 0 \text{ or } m_k = 01 \cdots 1)$ **then**
 - 9: Output $|\Psi^{(k)}\rangle$ and **continue**;
 - 10: **else**
 - 11: **break** and start over from initialization;
 - 12: **end if**
 - 13: **end for**
 - 14: **return** $|\Psi_{12\dots s}^{(K)}\rangle$.
-

We numerically benchmark the performance of a distributed filter algorithm on a single device, two or three devices in Fig. 5(b,c,d). We choose $n = 4$ and $|\psi^{(0)}\rangle = |+\rangle^{\otimes n}$ per device. Notably, the three-device filter (with either *weak* or *strong* postselection) consistently yields states with lower average energy variance. It also introduces greater energy biases. Regarding the cumulative success rate, similar to the two-device case, *strong* postselection incurs exponential overhead of the success rate with respect to the number of iterations k , while the overhead for *weak* postselection stabilizes after several iterations, which is lower-bounded by $\sum_j |c_j^{(0)}|^2 s$ for general s . Additionally, the three-device algorithm always incurs more overhead, as the number of possible measurement results that are discarded under postselection increases. Note that when the number of devices s is prime, each D^q for $q \neq 0$ constitutes a full-cycle permutation on the registers [29]. However, this does not hold for non-prime s , which might affect the performance of the multi-device distributed filtering algorithm under *weak* postselection. We provide further analysis on this generalization in Appendix C.

Numerical simulations for more than three devices are not feasible due to the exponential growth of Hilbert space. However, already from the data available for up to three devices and four qubits we can attempt an extrapolation of the protocol performance. We perform an exponential fit $\propto \exp(-\eta k)$ of the average variance shown in Fig. 5(c) in the interval $k \in [4, 12]$, i.e., after the initial roll-off and before the data for the *strong* postselection case has some visible fluctuations due to a smaller number of samples remaining after postselection. For the single device case, we find $\eta = 0.178$. Using the *weak* postselection criterion, we have $\eta = 0.277$ for two devices and $\eta = 0.386$ for three devices. For the *strong* postselection criterion, we find $\eta = 0.376$ for two devices and $\eta = 0.490$ for three devices.

With only limited data available due to the prohibitive simulation cost with the system size and the number of devices, an extrapolation to more than two devices is only indicative. However, for the *strong* postselection case there is a related analytical result in the limit of many devices [16]: the relative suppression between a dominant and a subdominant amplitude converges asymptoti-

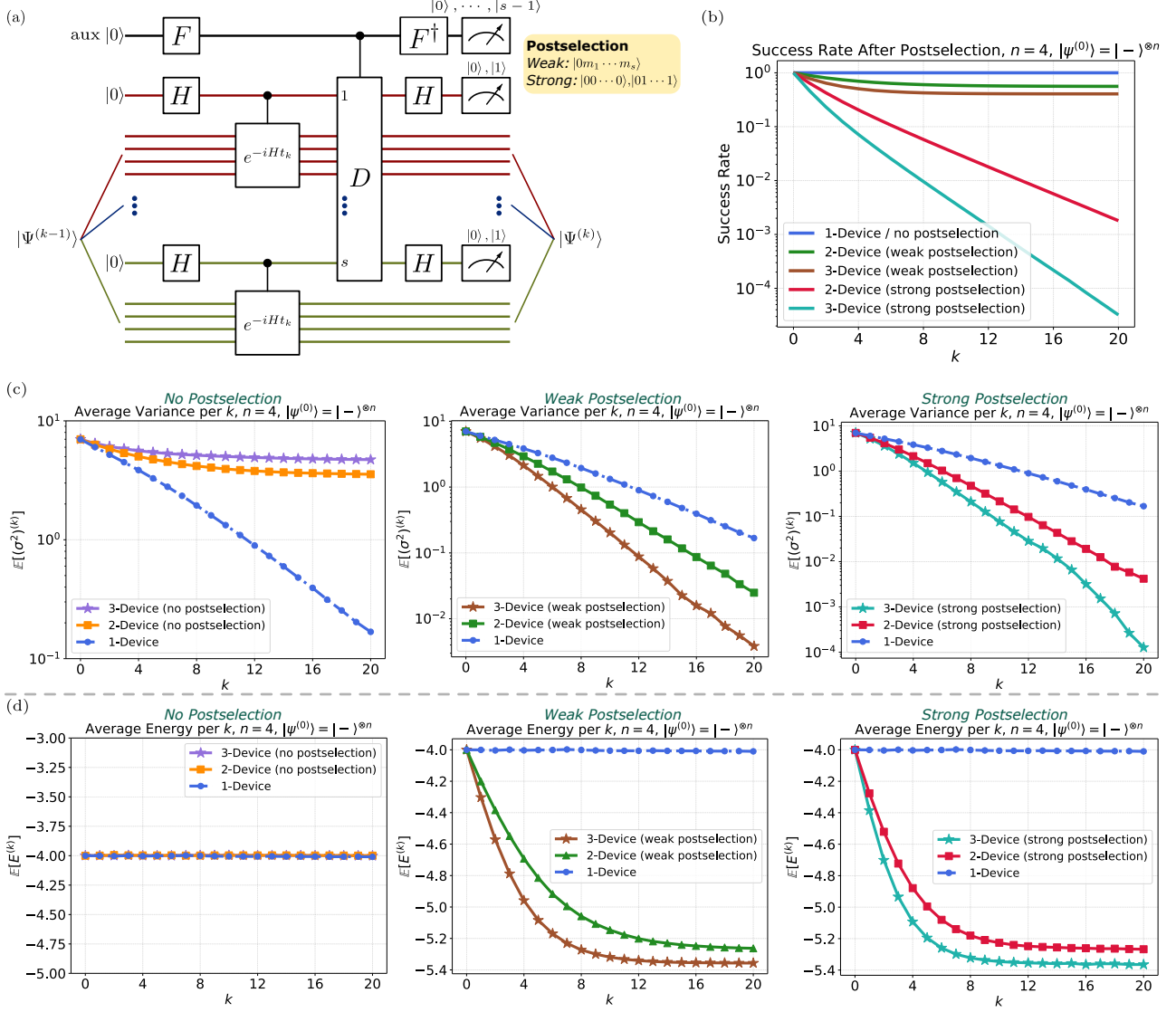


Figure 5: (a) Proposed distributed filtering algorithm across $s \geq 2$ devices. Similar to the two-device algorithm, the input is a product state $|\Psi^{(0)}\rangle = |\psi^{(0)}\rangle^{\otimes s}$ with sufficiently large random values t_k for each iteration k . Additionally, a top auxiliary qudit with s levels is introduced, and a cyclic permutation test [27, 28, 29] is applied across the top auxiliary qudit and the Hadamard test auxiliary qubits of each device. Postselection, as defined above, is still required. Numerics for $n = 4$ and $|\psi^{(0)}\rangle = |-\rangle^{\otimes n}$ for single-, two-, and three-device cases are shown in (b), (c), and (d): (b) cumulative success rate after postselection; (c) average variance; and (d) average energy of the pure state ensemble per iteration k under no postselection, *weak* postselection, and *strong* postselection, respectively. We use 10^7 trials from the beginning for *strong* postselection cases and 10^5 for the rest to acquire accurate enough expectation values. Plots with complete error bars are presented in Appendix E.

cally to a fixed value. However, the data in Fig. 5(c) does not consider the ratio between amplitudes, but takes all amplitudes of the state into consideration. While an immediate quantitative comparison with the analytical result is therefore not possible, we conjecture that the scaling of the average variance of the *strong* postselection result could also converge, implying that additional devices result in diminishing returns.

4 Discussion and Outlook

In this work, we have presented a distributed filtering algorithm for eigenstate preparation that employs multiple quantum devices with postselection. By coordinating operations across two devices, this approach allows one to prepare low-variance states with a lower circuit depth than in a single-device setup. Our results demonstrate that postselection can play a critical role in accelerating convergence towards the eigenstates. We emphasize that only the overhead for the *strong* postselection grows exponentially in the number of rounds. For *weak* postselection, the overhead due to postselection converges to a fixed value that is independent of the number of rounds of the protocol.

The protocol we present is a distributed algorithm inspired by methods from iterative quantum phase estimation (IQPE). Our technique inherits the minimal requirement of a single auxiliary qubit for the conditional dynamics from IQPE. Textbook QPE, however, proceeds by including an inverse quantum Fourier transformation on the auxiliary register. One can devise a similar distributed algorithm for textbook QPE, although this would require additional teleportation operations per circuit between the two devices due to the additional auxiliary qubits than the ones considered in this work.

The two-device circuit was previously presented in [16] with an eigenstate broadcasting application, so we briefly highlight the novel contributions in this work. First, we no longer assume the initial state to be in a superposition of a dominant eigenstate and other subdominant state. Rather, we consider an energy distribution described by its mean and variance, corresponding to the typical scenario of an initial product states [11]. Second, we introduced novel technical analysis tools based on the moments of

the distribution. Next, we have included a new numerical analysis on the different use (which we named *weak* vs. *strong*) of postselection in the distributed quantum algorithm. Finally, we present a new approach to realizing a symmetric projection of the auxiliary qubits that uses a qudit-controlled derangement operator.

The extension of the distributed algorithms to arbitrarily many connected quantum devices is a main result of our paper. We also show numerically that a third device allows for an even larger reduction of the energy variance for a fixed number of iterations of the protocol than with two devices. This naturally leads to an increased reset overhead, which is also bounded for the *weak* postselection criterion, and increases exponentially with each iteration for the *strong* postselection criterion. However, when successful, *strong* postselection converges exponentially faster than the corresponding *weak* postselection scheme.

Our findings also suggest that a hybrid protocol that starts with *strong* postselection and transitions to *weak* postselection would optimize resources better, as the *weak* postselection variance plateaus when the devices hold a sufficiently good approximation to identical eigenstates.

Overall, distributed quantum algorithms offer promising strategies to accelerate eigenstate preparation by reducing the circuit depth required. Practical realizations of these algorithms in noisy quantum devices may shed light on optimal engineering of projections onto symmetric subspaces and further improvements of distributed techniques more generally.

Acknowledgments

The numerical experiments of this work were performed using the compute resources from the Academic Leiden Interdisciplinary Cluster Environment (ALICE) provided by Leiden University. X.L. thanks Kshiti Sneh Rai, Johannes Knörzer, Zherui Jerry Wang and Yuning Zhang for the helpful discussions. B.F.S. acknowledges funding from the the Federal Ministry of Education and Research Germany (BMBF) via the project FermiQP (13N15889). Work in Munich was part of the Munich Quantum Valley, which is

supported by the Bavarian state government with funds from the Hightech Agenda Bayern Plus. J.T. acknowledges the support received from the European Union’s Horizon Europe research and innovation programme through the ERC StG FINETEA-SQUAD (Grant No. 101040729). J.T. also acknowledges the support received by the Dutch National Growth Fund (NGF), as part of the Quantum Delta NL programme. This publication is part of the ‘Quantum Inspire – the Dutch Quantum Computer in the Cloud’ project (with project number [NWA.1292.19.194]) of the NWA research program ‘Research on Routes by Consortia (ORC)’, which is funded by the Netherlands Organization for Scientific Research (NWO). The views and opinions expressed here are solely those of the authors and do not necessarily reflect those of the funding institutions. Neither of the funding institutions can be held responsible for them.

References

- [1] Michael A. Nielsen and Isaac L. Chuang. “Quantum Computation and Quantum Information: 10th Anniversary Edition”. Cambridge University Press. (2010). url: <https://doi.org/10.1017/CB09780511976667>.
- [2] A. Yu Kitaev. “Quantum measurements and the Abelian Stabilizer Problem” (1995). [arXiv:quant-ph/9511026](https://arxiv.org/abs/quant-ph/9511026).
- [3] Thomas E. O’Brien, Brian Tarasinski, and Barbara M. Terhal. “Quantum phase estimation of multiple eigenvalues for small-scale (noisy) experiments”. *New Journal of Physics* **21**, 023022 (2019).
- [4] Alicja Dutkiewicz, Barbara M. Terhal, and Thomas E. O’Brien. “Heisenberg-limited quantum phase estimation of multiple eigenvalues with few control qubits”. *Quantum* **6**, 830 (2022).
- [5] Michael Hartmann, Günter Mahler, and Orwin Hess. “Spectral Densities and Partition Functions of Modular Quantum Systems as Derived from a Central Limit Theorem”. *Journal of Statistical Physics* **119**, 1139–1151 (2005).
- [6] Daniel S. Abrams and Seth Lloyd. “Quantum Algorithm Providing Exponential Speed Increase for Finding Eigenvalues and Eigenvectors”. *Physical Review Letters* **83**, 5162–5165 (1999).
- [7] David Poulin and Pawel Wocjan. “Preparing Ground States of Quantum Many-Body Systems on a Quantum Computer”. *Physical Review Letters* **102**, 130503 (2009).
- [8] Yimin Ge, Jordi Tura, and J. Ignacio Cirac. “Faster ground state preparation and high-precision ground energy estimation with fewer qubits”. *Journal of Mathematical Physics* **60**, 022202 (2019).
- [9] Sirui Lu, Mari Carmen Bañuls, and J. Ignacio Cirac. “Algorithms for Quantum Simulation at Finite Energies”. *PRX Quantum* **2**, 020321 (2021).
- [10] Reinis Irmejs, Mari Carmen Bañuls, and J. Ignacio Cirac. “Efficient Quantum Algorithm for Filtering Product States”. *Quantum* **8**, 1389 (2024).
- [11] Kshiti Sneh Rai, J. Ignacio Cirac, and Álvaro M. Alhambra. “Matrix product state approximations to quantum states of low energy variance”. *Quantum* **8**, 1401 (2024).
- [12] Miroslav Dobšíček, Göran Johansson, Vitaly Shumeiko, and Göran Wendin. “Arbitrary accuracy iterative quantum phase estimation algorithm using a single ancillary qubit: A two-qubit benchmark”. *Physical Review A* **76**, 030306 (2007).
- [13] Jin-Shi Xu, Man-Hong Yung, Xiao-Ye Xu, Sergio Boixo, Zheng-Wei Zhou, Chuan-Feng Li, Alán Aspuru-Guzik, and Guang-Can Guo. “Demon-like algorithmic quantum cooling and its realization with quantum optics”. *Nature Photonics* **8**, 113–118 (2014).
- [14] Richard Meister and Simon C. Benjamin. “Resource-frugal Hamiltonian eigenstate preparation via repeated quantum phase estimation measurements” (2022). [arXiv:2212.00846](https://arxiv.org/abs/2212.00846).
- [15] Yanzhu Chen and Tzu-Chieh Wei. “Quantum algorithm for spectral projection by measuring an ancilla iteratively”. *Physical Review A* **101**, 032339 (2020).
- [16] Benjamin F. Schiffer and Jordi Tura. “Quantum eigenstate preparation assisted by a coherent link”. *Physical Review A* **111**, 012445 (2025).

- [17] Zhengrong Qian, Jacob Watkins, Gabriel Given, Joey Bonitati, Kenneth Choi, and Dean Lee. “Demonstration of the rodeo algorithm on a quantum computer”. *The European Physical Journal A* **60**, 151 (2024).
- [18] Kenneth Choi, Dean Lee, Joey Bonitati, Zhengrong Qian, and Jacob Watkins. “Rodeo Algorithm for Quantum Computing”. *Physical Review Letters* **127**, 040505 (2021).
- [19] Dorit Aharonov, Vaughan Jones, and Zeph Landau. “A polynomial quantum algorithm for approximating the Jones polynomial”. In Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing. Pages 427–436. STOC ’06 New York, NY, USA (2006). Association for Computing Machinery.
- [20] P. Magnard, S. Storz, P. Kurpiers, J. Schär, F. Marxer, J. Lütolf, T. Walter, J.-C. Besse, M. Gabureac, K. Reuer, A. Akin, B. Royer, A. Blais, and A. Wallraff. “Microwave Quantum Link between Superconducting Circuits Housed in Spatially Separated Cryogenic Systems”. *Physical Review Letters* **125**, 260502 (2020).
- [21] Patrick Harvey-Collard, Jurgen Dijkema, Guoji Zheng, Amir Sammak, Giordano Scappucci, and Lieven M. K. Vandersypen. “Coherent Spin-Spin Coupling Mediated by Virtual Microwave Photons”. *Physical Review X* **12**, 021026 (2022).
- [22] J. I. Cirac, A. K. Ekert, S. F. Huelga, and C. Macchiavello. “Distributed quantum computation over noisy channels”. *Physical Review A* **59**, 4249–4254 (1999).
- [23] Harry Buhrman and Hein Röhrig. “Distributed Quantum Computing”. In Branislav Rován and Peter Vojtáš, editors, *Mathematical Foundations of Computer Science 2003*. Pages 1–20. Berlin, Heidelberg (2003). Springer.
- [24] Adriano Barenco, André Berthiaume, David Deutsch, Artur Ekert, Richard Jozsa, and Chiara Macchiavello. “Stabilization of Quantum Computations by Symmetrization”. *SIAM Journal on Computing* **26**, 1541–1557 (1997).
- [25] Harry Buhrman, Richard Cleve, John Watrous, and Ronald de Wolf. “Quantum Fingerprinting”. *Physical Review Letters* **87**, 167902 (2001).
- [26] Max Born. “Zur Quantenmechanik der Stoßvorgänge”. *Zeitschrift für Physik* **37**, 863–867 (1926).
- [27] Masaru Kada, Harumichi Nishimura, and Tomoyuki Yamakami. “The efficiency of quantum identity testing of multiple states”. *Journal of Physics A: Mathematical and Theoretical* **41**, 395309 (2008).
- [28] Harry Buhrman, Dmitry Grinko, Philip Verduyn Lunel, and Jordi Weggemans. “Permutation tests for quantum state identity” (2024). [arXiv:2405.09626](https://arxiv.org/abs/2405.09626).
- [29] Xiaoyu Liu, Johannes Knörzer, Zherui Jerry Wang, and Jordi Tura. “Generalized concentratable entanglement via parallelized permutation tests”. *Physical Review Research* **7**, L032022 (2025).
- [30] Hermann Weyl. “Über die Gleichverteilung von Zahlen mod. Eins”. *Mathematische Annalen* **77**, 313–352 (1916).

A Single-device filter

We start from stating the single-device filter algorithm [6, 13, 15, 14, 16] and the pseudocode is shown below:

Algorithm 3 Single-device distributed ancilla-mediated filtering algorithms.

Require: A product state $|\psi^{(0)}\rangle$; target iteration K .

Ensure: State $|\psi^{(K)}\rangle$.

- 1: Initialize: $|\Psi^{(0)}\rangle \leftarrow |\psi^{(0)}\rangle$;
 - 2: **for** $k = 1$ to K **do**
 - 3: Generate random time t_k such that $\varphi_j^{(k)} = -t_k \lambda_j \mod 2\pi \sim \text{Uniform}(0, 2\pi)$ for every j ;
 - 4: Input $|\Psi^{(k-1)}\rangle$ and t_k to the circuit in Fig. 6;
 - 5: Measure and output $|\Psi^{(k)}\rangle$;
 - 6: **continue**;
 - 7: **end for**
 - 8: $|\psi^{(K)}\rangle \leftarrow |\Psi^{(K)}\rangle$;
 - 9: **return** $|\psi^{(K)}\rangle$.
-

Similarly, we would like to find out the behavior of the output states' average E and σ^2 .

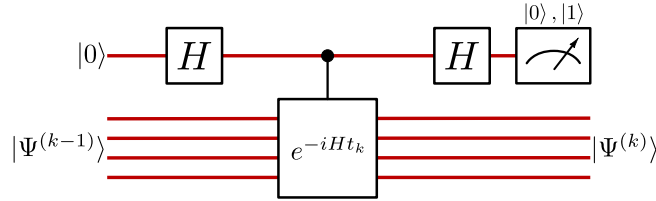


Figure 6: Circuit for single-device filtering algorithm.

A.1 Energy

After $K \in \mathbb{Z}^+$ iterations, the output state $|\psi^{(K)}\rangle$ has the energy:

$$E^{(K)} = \langle \psi^{(K)} | H | \psi^{(K)} \rangle. \quad (16)$$

As mentioned in the main text, for each repetition, $|\psi^{(K)}\rangle$ can vary. Suppose we have enough repetitions r which corresponds to each state $|\psi_r^{(K)}\rangle$ with probability p_r . Therefore, assume the sequence of time evolution $\mathbf{t} = \{t_1, t_2, \dots, t_K\}$ on average, we have:

$$\mathbb{E} [E^{(K)} | \mathbf{t}] = \sum_r p_r \langle \psi_r^{(K)} | H | \psi_r^{(K)} \rangle = \text{Tr} \left(H \sum_r p_r |\psi_r^{(K)}\rangle \langle \psi_r^{(K)}| \right) = \text{Tr} (H \rho_{\mathbf{t}}^{(K)}), \quad (17)$$

where we denote $\rho_{\mathbf{t}}^{(K)}$ as the density matrix at iteration K under the fixed time evolution sequence $\mathbf{t}$. Note that for the Hadamard test operation at each iteration k , we have a channel:

$$\rho_{\mathbf{t}}^{(k)} = \mathcal{E}^{(k)}(\rho^{(k-1)}) = \sum_{m=0}^1 \mathbf{E}_m^{(k)} \rho^{(k-1)} \mathbf{E}_m^{(k)\dagger}, \quad (18)$$

with Kraus operators:

$$\begin{aligned}\mathbf{E}_0^{(k)} &= \frac{1}{2}(I + U(t_k)) \quad \rightarrow |0\rangle \text{ on the auxiliary qubit;} \\ \mathbf{E}_1^{(k)} &= \frac{1}{2}(I - U(t_k)) \quad \rightarrow |1\rangle \text{ on the auxiliary qubit.}\end{aligned}\tag{19}$$

Therefore,

$$\rho_{\mathbf{t}}^{(K)} = \mathcal{E}^{(K)} \circ \mathcal{E}^{(K-1)} \circ \dots \circ \mathcal{E}^{(1)} \left(|\psi^{(0)}\rangle \langle \psi^{(0)}| \right).\tag{20}$$

Note that:

$$\rho_{\mathbf{t}}^{(k)} = \mathcal{E}^{(k)} \left(\rho^{(k-1)} \right) = \frac{1}{2} \left(\rho^{(k-1)} + U(t_k) \rho^{(k-1)} U(t_k)^\dagger \right),\tag{21}$$

including the input state, controlled unitaries and Hamiltonian into the eigenbasis, i.e.,

$$|\psi^{(0)}\rangle = \sum_j c_j^{(0)} |\phi_j\rangle;\tag{22}$$

$$U(t_k) = e^{-it_k H} = \sum_j e^{-it_k \lambda_j} |\phi_j\rangle \langle \phi_j| = \sum_j e^{i\varphi_j^{(k)}} |\phi_j\rangle \langle \phi_j|;\tag{23}$$

$$H = \sum_j \lambda_j |\phi_j\rangle \langle \phi_j|,\tag{24}$$

we have (we omit superscript $^{(0)}$ of $c_j^{(0)}$ for simplicity):

$$\begin{aligned}\rho_{\mathbf{t}}^{(K)} &= \mathcal{E}^{(K)} \circ \mathcal{E}^{(K-1)} \circ \dots \circ \mathcal{E}^{(1)} \left(|\psi^{(0)}\rangle \langle \psi^{(0)}| \right) \\ &= \frac{1}{2^K} \sum_{\alpha \in \mathcal{P}(\{1,2,\dots,K\})} \left(\prod_{a \in \alpha} U(t_a) \right) |\psi^{(0)}\rangle \langle \psi^{(0)}| \left(\prod_{a' \in \alpha} U(t_{a'}) \right)^\dagger \\ &= \frac{1}{2^K} \sum_{\alpha \in \mathcal{P}(\{1,2,\dots,K\})} \sum_{l,l'} \exp \left(i \sum_{a \in \alpha} \varphi_l^{(a)} - i \sum_{a' \in \alpha} \varphi_{l'}^{(a')} \right) c_l c_{l'}^* |\phi_l\rangle \langle \phi_{l'}|,\end{aligned}\tag{25}$$

where $\mathcal{P}(\{1,2,\dots,K\})$ denotes all the subsets of $\{1,2,\dots,K\}$. Therefore:

$$\begin{aligned}\mathbb{E} \left[E^{(K)} | \mathbf{t} \right] &= \text{Tr} \left(H \rho_{\mathbf{t}}^{(K)} \right) \\ &= \frac{1}{2^K} \sum_{\alpha \in \mathcal{P}(\{1,2,\dots,K\})} \text{Tr} \left(\sum_j \lambda_j |\phi_j\rangle \langle \phi_j| \sum_{l,l'} \exp \left(i \sum_{a \in \alpha} \varphi_l^{(a)} - i \sum_{a' \in \alpha} \varphi_{l'}^{(a')} \right) c_l c_{l'}^* |\phi_l\rangle \langle \phi_{l'}| \right) \\ &= \text{Tr} \left(\sum_{l,l'} \lambda_l c_l c_{l'}^* |\phi_l\rangle \langle \phi_{l'}| \right) = \sum_l \lambda_l |c_l|^2.\end{aligned}\tag{26}$$

As $\mathbb{E} \left[E^{(K)} | \mathbf{t} \right]$ does not depend on $\mathbf{t}$, therefore:

$$\mathbb{E} \left[E^{(k)} \right] = \sum_j \lambda_j |c_j|^2.\tag{27}$$

which equals to the energy of the initial state $|\psi^{(0)}\rangle$.

A.2 Energy Variance

The variance of the output state $|\psi^{(K)}\rangle$ is:

$$(\sigma^2)^{(K)} = \langle \psi^{(K)} | H^2 | \psi^{(K)} \rangle - \left(\langle \psi^{(K)} | H | \psi^{(K)} \rangle \right)^2, \quad (28)$$

and on average:

$$\mathbb{E} \left[(\sigma^2)^{(K)} \right] = \mathbb{E} \left[\langle \psi^{(K)} | H^2 | \psi^{(K)} \rangle \right] - \mathbb{E} \left[\left(\langle \psi^{(K)} | H | \psi^{(K)} \rangle \right)^2 \right]. \quad (29)$$

Similar to the previous derivations, we have the first term as:

$$\mathbb{E} \left[\langle \psi^{(K)} | H^2 | \psi^{(K)} \rangle \right] = \text{Tr} \left(H^2 \rho^{(K)} \right) = \sum_j \lambda_j^2 |c_j|^2. \quad (30)$$

The second term is complicated. First, we again fix the time sequence $\mathbf{t}$ and find:

$$\mathbb{E} \left[\left(\langle \psi^{(K)} | H | \psi^{(K)} \rangle \right)^2 | \mathbf{t} \right] = \sum_r p_r \left(\langle \psi_r^{(K)} | H | \psi_r^{(K)} \rangle \right)^2. \quad (31)$$

Note that each $|\psi_r^{(K)}\rangle$ can be denoted as:

$$|\psi_r^{(K)}\rangle = \frac{1}{\sqrt{P_{\mathbf{m}}}} \frac{1}{2^K} \prod_{k=1}^K (I + (-1)^{m_k} U(t_k)) |\psi^{(0)}\rangle = \frac{1}{\sqrt{P_{\mathbf{m}}}} \frac{1}{2^K} \sum_j c_j \prod_{k=1}^K \left(1 + (-1)^{m_k} e^{i\varphi_j^{(k)}} \right) |\phi_j\rangle, \quad (32)$$

where $P_{\mathbf{m}}$ denotes the probability of getting the measurement result $\mathbf{m} = \{m_1, m_2, \dots, m_K\}$ on the auxiliary qubit of each iteration k at repetition r . Then we have:

$$\langle \psi_r^{(K)} | H | \psi_r^{(K)} \rangle = \frac{1}{P_{\mathbf{m}}} \frac{1}{2^K} \sum_j |c_j|^2 \prod_{k=1}^K \left(1 + (-1)^{m_k} \cos \varphi_j^{(k)} \right) \lambda_j, \quad (33)$$

and:

$$P_{\mathbf{m}} = \frac{1}{2^K} \sum_j |c_j|^2 \prod_{k=1}^K \left(1 + (-1)^{m_k} \cos \varphi_j^{(k)} \right). \quad (34)$$

Therefore:

$$\begin{aligned} \mathbb{E} \left[\left(\langle \psi^{(K)} | H | \psi^{(K)} \rangle \right)^2 | \mathbf{t} \right] &= \sum_r p_r \left(\langle \psi_r^{(K)} | H | \psi_r^{(K)} \rangle \right)^2 \\ &= \frac{1}{2^K} \sum_{\mathbf{m}} \frac{\left(\sum_j |c_j|^2 \lambda_j \prod_{k=1}^K \left(1 + (-1)^{m_k} \cos \varphi_j^{(k)} \right) \right)^2}{\sum_j |c_j|^2 \prod_{k=1}^K \left(1 + (-1)^{m_k} \cos \varphi_j^{(k)} \right)} \\ &= \frac{1}{2^K} \sum_{\mathbf{m}} \frac{\left(\sum_j |c_j|^2 \lambda_j \prod_{k=1}^K \left(1 + (-1)^{m_k} \cos \lambda_j t_k \right) \right)^2}{\sum_j |c_j|^2 \prod_{k=1}^K \left(1 + (-1)^{m_k} \cos \lambda_j t_k \right)}. \end{aligned} \quad (35)$$

Now, suppose $\mathbf{t}$ are uniformly distributed within a time interval T . Then, we have:

$$\mathbb{E} \left[\left(\langle \psi^{(K)} | H | \psi^{(K)} \rangle \right)^2 \right] = \frac{1}{T^K} \frac{1}{2^K} \sum_{\mathbf{m}} \int_0^T \cdots \int_0^T \frac{\left(\sum_j |c_j|^2 \lambda_j \prod_{k=1}^K \left(1 + (-1)^{m_k} \cos \lambda_j t_k \right) \right)^2}{\sum_j |c_j|^2 \prod_{k=1}^K \left(1 + (-1)^{m_k} \cos \lambda_j t_k \right)} dt_1 \cdots dt_K. \quad (36)$$

Next, we show why we need large T to maximize the quantity $\mathbb{E} \left[\left(\langle \psi^{(K)} | H | \psi^{(K)} \rangle \right)^2 \right]$, thus minimize $\mathbb{E} \left[(\sigma^2)^{(K)} \right]$. Since we lack prior knowledge of the populations of $|\psi^{(0)}\rangle$ in the eigenbasis, one of the optimal approaches is to completely randomize the phases $\varphi_j^{(k)}$. Specifically, from Eq. (34) and (32), we can observe that: suppose one of the populations $|c_j^{(k-1)}|^2$ is already sufficiently larger than the other populations at $(k-1)$ -th iteration. And we assume for the k -th iteration, the extra factor $(1 + (-1)^{m_k} \cos \varphi_j^{(k)}) > 1$. This, in general, magnifies the probability $P_{\mathbf{m}}$ of achieving $|c_j^{(k)}|^2 = |c_j^{(k-1)}|^2 (1 + (-1)^{m_k} \cos \varphi_j^{(k)}) > |c_j^{(k-1)}|^2$, amplifying this population even more. In order to achieve this, we need to let $\varphi_j^{(k)}$ be independent with each other, which requires large time interval T . Otherwise, the correlations among different $\varphi_j^{(k)}$ may magnify several populations instead of only one, turning into superpositions of eigenstates.

Proposition 1. *Under the assumptions that $\frac{\lambda_{j'}}{\lambda_{j''}}$ is (almost) irrational for any $j' \neq j'' \in \{1, 2, \dots, 2^n\}$, for the time interval T in large limit and t_k uniformly sampled within the interval, $\varphi_j^{(k)} = -t_k \lambda_j \pmod{2\pi}$ can be approximated as i.i.d. in $\text{Uniform}(0, 2\pi)$ for all j and k .*

Proof. Firstly, we consider the Weyl's criterion of equidistribution [30], showing that $\varphi_j^{(k)}$ for specific j and random t_k distributed in large interval T can be considered as $\text{Uniform}(0, 2\pi)$. Weyl's criterion states that $\frac{\varphi_j^{(k)}}{2\pi} \pmod{1}$ is uniformly distributed if and only if for all non-zero integers l :

$$\lim_{R \rightarrow \infty} \frac{1}{R} \sum_{r=1}^R e^{i2\pi l \frac{\varphi_j^{(k)}}{2\pi}} = \lim_{R \rightarrow \infty} \frac{1}{R} \sum_{r=1}^R e^{-ilt_k \lambda_j} = \frac{1}{T} \int_{T_{\min}}^{T_{\max}} e^{-ilt_k \lambda_j} dt_k = 0. \quad (37)$$

Here we suppose $t_k \sim \text{Uniform}(T_{\min}, T_{\max})$, $T_{\max} > T_{\min} > 0$ and $T_{\max} - T_{\min} = T$. By integrating the above integral, we have:

$$\left| \frac{1}{T} \int_{T_{\min}}^{T_{\max}} e^{-ilt_k \lambda_j} dt_k \right|^2 = \frac{1}{l^2} \text{sinc}^2 \left(\frac{T \lambda_j}{2} \right) \leq \text{sinc}^2 \left(\frac{T \lambda_j}{2} \right) \rightarrow 0, \quad (38)$$

for large interval T .

Secondly, we show that each $\varphi_j^{(k)}$ is independent with each other for all j and k , under the assumptions that $\frac{\lambda_{j'}}{\lambda_{j''}}$ is irrational for any $j' \neq j'' \in \{1, 2, \dots, 2^n\}$. If $\frac{\lambda_{j'}}{\lambda_{j''}}$ is rational, i.e., $\frac{\lambda_{j'}}{\lambda_{j''}} = \frac{p}{q}$ where $p, q \in \mathbb{Z}$, then there exists a period t within $T_{\min} \sim T_{\max}$ such that $\frac{2\pi p}{\lambda_{j'}} = \frac{2\pi q}{\lambda_{j''}} = t$. Because of this, the points with coordinates $(\varphi_{j'}^{(k)}, \varphi_{j''}^{(k)})$ cannot cover the whole area of $(0, 2\pi)^2$. However, if $\frac{\lambda_{j'}}{\lambda_{j''}}$ is irrational, there will be no such t that exists. Thus, $(\varphi_{j'}^{(k)}, \varphi_{j''}^{(k)})$ can cover all $(0, 2\pi)^2$ if t_k is sufficiently large. The independence can also be found for rational $\frac{\lambda_{j'}}{\lambda_{j''}}$ but with extremely large period t . In this case, $(\varphi_{j'}^{(k)}, \varphi_{j''}^{(k)})$ still nearly cover $(0, 2\pi)^2$, before any periodic behavior emerges. This is where we name almost-irrational $\frac{\lambda_{j'}}{\lambda_{j''}}$.

A simple counter-example for the independence is that we choose the Hamiltonian as $H = \sum_j Z_j Z_{j+1}$ where it has symmetric eigenvalues like $\lambda_{j'} = -\lambda_{j''}$. In this case, $\varphi_{j'}^{(k)}$ and $\varphi_{j''}^{(k)}$ are strongly correlated. The points $(\varphi_{j'}^{(k)}, \varphi_{j''}^{(k)})$ form a diagonal line on $(0, 2\pi)^2$. Consequently, this algorithm fails as it cannot result in $|c_{j'}|^2$ being magnified while $|c_{j''}|^2$ is suppressed, or vice versa. Shifting the Hamiltonian to break the symmetries can remedy this, i.e., $H' = H + \epsilon I$ (cf. Appendix in [16]). $\square$

Now, by proving Proposition 1, we have:

$$\begin{aligned}
& \mathbb{E} \left[\left(\langle \psi^{(K)} | H | \psi^{(K)} \rangle \right)^2 \right] \\
&= \frac{1}{2^K} \sum_{\mathbf{m}} \left(\frac{1}{2\pi} \right)^{NK} \int_0^{2\pi} \cdots \int_0^{2\pi} \frac{\left(\sum_j |c_j|^2 \lambda_j \prod_{k=1}^K \left(1 + (-1)^{m_k} \cos \varphi_j^{(k)} \right) \right)^2}{\sum_j |c_j|^2 \prod_{k=1}^K \left(1 + (-1)^{m_k} \cos \varphi_j^{(k)} \right)} d\varphi_1^{(1)} \cdots d\varphi_N^{(K)} \\
&= \left(\frac{1}{2\pi} \right)^{NK} \int_0^{2\pi} \cdots \int_0^{2\pi} \frac{\left(\sum_j |c_j|^2 \lambda_j \prod_{k=1}^K \left(1 - \cos \varphi_j^{(k)} \right) \right)^2}{\sum_j |c_j|^2 \prod_{k=1}^K \left(1 - \cos \varphi_j^{(k)} \right)} d\varphi_1^{(1)} \cdots d\varphi_N^{(K)},
\end{aligned} \tag{39}$$

where $N = 2^n$. The second equality is due to the symmetries of $\cos \varphi_j^{(k)}$ within $0 \sim 2\pi$. Note that this integral cannot be computed analytically. However, as shown in [15], $\mathbb{E} \left[\left(\langle \psi^{(K)} | H | \psi^{(K)} \rangle \right)^2 \right]$ increases with K , leading to a decrease in the average variance $\mathbb{E} \left[(\sigma^2)^{(K)} \right]$. Therefore, finally we have:

$$\mathbb{E} \left[(\sigma^2)^{(K)} \right] = \sum_l \lambda_l^2 |c_l^{(0)}|^2 - \left(\frac{1}{2\pi} \right)^{NK} \int_0^{2\pi} \cdots \int_0^{2\pi} \frac{\left(\sum_j |c_j|^2 \lambda_j \prod_{k=1}^K \left(1 - \cos \varphi_j^{(k)} \right) \right)^2}{\sum_j |c_j|^2 \prod_{k=1}^K \left(1 - \cos \varphi_j^{(k)} \right)} d\varphi_1^{(1)} \cdots d\varphi_N^{(K)}. \tag{40}$$

B Two-device filter

The setup of two-device filter has been shown in Fig. 1(b) and Algorithm 1. Below we analyze the behavior of two-device filter under no, *weak* and *strong* postselection, respectively.

B.1 Energy

B.1.1 No postselection

In the case where we do not postselect anything and let the two-device protocol run until K iterations, at each iteration k , we have multiple possible operations (Kraus operators) that can be applied on the state (we omit the iteration notation $^{(k)}$ for now):

$$\begin{aligned}
\mathbf{E}_0^{00} &= \frac{1}{4} (I + U) \otimes (I + U) = \frac{1}{4} \sum_{j,j'} (1 + e^{i\varphi_j})(1 + e^{i\varphi_{j'}}) |\phi_j \phi_{j'}\rangle \langle \phi_j \phi_{j'}|; \\
\mathbf{E}_0^{01} &= \frac{1}{4} (I \otimes I - U \otimes U) = \frac{1}{4} \sum_{j,j'} (1 - e^{i(\varphi_j + \varphi_{j'})}) |\phi_j \phi_{j'}\rangle \langle \phi_j \phi_{j'}| = \mathbf{E}_0^{10}; \\
\mathbf{E}_0^{11} &= \frac{1}{4} (I - U) \otimes (I - U) = \frac{1}{4} \sum_{j,j'} (1 - e^{i\varphi_j})(1 - e^{i\varphi_{j'}}) |\phi_j \phi_{j'}\rangle \langle \phi_j \phi_{j'}|; \\
\mathbf{E}_1^{01} &= \frac{1}{4} (U \otimes I - I \otimes U) = \frac{1}{4} \sum_{j,j'} (e^{i\varphi_j} - e^{i\varphi_{j'}}) |\phi_j \phi_{j'}\rangle \langle \phi_j \phi_{j'}| = -\mathbf{E}_1^{10}; \\
\mathbf{E}_1^{00} &= 0 = \mathbf{E}_1^{11}.
\end{aligned} \tag{41}$$

where the subscript is the measurement result of top auxiliary qubit and the superscript denotes the other two. One can easily check that:

$$\sum_{j,k,l=0}^1 \mathbf{E}_j^{kl\dagger} \mathbf{E}_j^{kl} = I. \tag{42}$$

For energy, similarly, we consider the quantum channel. Initially, the state is:

$$\rho^{(0)} = |\psi^{(0)}\psi^{(0)}\rangle\langle\psi^{(0)}\psi^{(0)}|. \quad (43)$$

Once we apply the quantum channel with Kraus operators $\mathcal{E}$, we have the state after one round of the protocol as:

$$\begin{aligned} \rho^{(1)} &= \sum_{j,k,l=0}^1 \mathbf{E}_j^{kl} \rho^{(0)} \mathbf{E}_j^{kl\dagger} \\ &= \frac{1}{4} \left(\rho^{(0)} + (I \otimes U) \rho^{(0)} (I \otimes U^\dagger) + (U \otimes I) \rho^{(0)} (U^\dagger \otimes I) + (U \otimes U) \rho^{(0)} (U^\dagger \otimes U^\dagger) \right). \end{aligned} \quad (44)$$

Then we can define our new equivalent channel with Kraus operators:

$$\begin{aligned} \mathcal{K}_1 &= I \otimes I = \sum_{j,j'} |\phi_j \phi_{j'}\rangle \langle \phi_j \phi_{j'}|; \\ \mathcal{K}_2 &= I \otimes U = \sum_{j,j'} e^{i\varphi_j} |\phi_j \phi_{j'}\rangle \langle \phi_j \phi_{j'}|; \\ \mathcal{K}_3 &= U \otimes I = \sum_{j,j'} e^{i\varphi_{j'}} |\phi_j \phi_{j'}\rangle \langle \phi_j \phi_{j'}|; \\ \mathcal{K}_4 &= U \otimes U = \sum_{j,j'} e^{i(\varphi_j + \varphi_{j'})} |\phi_j \phi_{j'}\rangle \langle \phi_j \phi_{j'}|. \end{aligned} \quad (45)$$

Therefore, we have:

$$\rho^{(1)} = \frac{1}{4} \sum_{l,l',h,h'} \left(1 + e^{i(\varphi_l - \varphi_h)}\right) \left(1 + e^{i(\varphi_{l'} - \varphi_{h'})}\right) c_l c_{l'} c_h^* c_{h'}^* |\phi_l \phi_{l'}\rangle \langle \phi_h \phi_{h'}|. \quad (46)$$

Here we also denote c as the input population and omit the iteration notation $^{(k)}$. Therefore, after K rounds, we have (adding the superscripts on φ to denote the different time evolution for each iteration k):

$$\rho^{(K)} = \frac{1}{4^K} \sum_{l,l',h,h'} \prod_{k=1}^K \left(1 + e^{i(\varphi_l^{(k)} - \varphi_h^{(k)})}\right) \left(1 + e^{i(\varphi_{l'}^{(k)} - \varphi_{h'}^{(k)})}\right) c_l c_{l'} c_h^* c_{h'}^* |\phi_l \phi_{l'}\rangle \langle \phi_h \phi_{h'}|. \quad (47)$$

Now, we trace out one of the system (e.g., Bob) and preserve one (e.g., Alice), we have:

$$\rho_A^{(K)} = \text{Tr}_B \left(\rho^{(K)} \right) = \frac{1}{2^K} \sum_{l,h} \sum_{\alpha \in \mathcal{P}(\{1, \dots, K\})} e^{i \sum_{a \in \alpha} (\varphi_l^{(a)} - \varphi_h^{(a)})} c_l c_h^* |\varphi_l\rangle \langle \varphi_h|. \quad (48)$$

Then, the average energy becomes:

$$\begin{aligned} \mathbb{E} \left[E^{(K)} \right] &= \mathbb{E} \left[E^{(K)} | \mathbf{t} \right] = \text{Tr} \left(H \rho_A^{(K)} \right) \\ &= \frac{1}{2^K} \text{Tr} \left(\sum_j \lambda_j |\phi_j\rangle \langle \phi_j| \sum_{l,h} \sum_{\alpha \in \mathcal{P}(1, \dots, K)} e^{i \sum_{a \in \alpha} (\varphi_l^{(a)} - \varphi_h^{(a)})} c_l c_h^* |\varphi_l\rangle \langle \varphi_h| \right) \\ &= \frac{1}{2^K} \text{Tr} \left(\sum_{j,l,h} \sum_{\alpha \in \mathcal{P}(1, \dots, K)} \lambda_j e^{i \sum_{a \in \alpha} (\varphi_l^{(a)} - \varphi_h^{(a)})} c_l c_h^* \delta(h, j) \delta(l, j) \right) \\ &= \frac{1}{2^K} \text{Tr} \left(\sum_j \lambda_j 2^K |c_j|^2 \right) = \sum_j \lambda_j |c_j|^2. \end{aligned} \quad (49)$$

This concludes that in the case where there is no postselection, the average energy will stay the same as the energy of the initial input state $|\psi^{(0)}\rangle$.

B.1.2 Weak postselection

Now we throw away the cases with $|1\rangle$ on the top auxiliary qubit. In this case, after one iteration of the protocol, we have the density matrix:

$$\rho^{(1)} = \frac{1}{\text{Tr}(\cdot)} \left(\rho^{(0)} + \frac{I \otimes U + U \otimes I}{\sqrt{2}} \rho^{(0)} \frac{I \otimes U^\dagger + U^\dagger \otimes I}{\sqrt{2}} + U \otimes U \rho^{(0)} U^\dagger \otimes U^\dagger \right), \quad (50)$$

where $\text{Tr}(\cdot)$ is the trace normalization term to make sure $\text{Tr}(\rho) = 1$.

We define:

$$\begin{aligned} \mathcal{K}_1 &= I \otimes I; \\ \mathcal{K}_2 &= \frac{I \otimes U + U \otimes I}{\sqrt{2}}; \\ \mathcal{K}_3 &= U \otimes U. \end{aligned} \quad (51)$$

Therefore, after K rounds, we have (note that $\mathcal{K}$ changes for each iteration due to different time evolution t_k):

$$\begin{aligned} \rho^{(K)} &= \frac{1}{\text{Tr}(\cdot)} \sum_{j,j',h,h'} \prod_{k=1}^K \left(1 + \frac{\left(e^{i\varphi_j^{(k)}} + e^{i\varphi_{j'}^{(k)}} \right) \left(e^{-i\varphi_h^{(k)}} + e^{-i\varphi_{h'}^{(k)}} \right)}{2} + e^{i\varphi_j^{(k)}} e^{i\varphi_{j'}^{(k)}} e^{-i\varphi_h^{(k)}} e^{-i\varphi_{h'}^{(k)}} \right) \\ &\quad c_j c_{j'} c_h^* c_{h'}^* |\phi_j \phi_{j'}\rangle \langle \phi_h \phi_{h'}|. \end{aligned} \quad (52)$$

Similarly, we trace out one of the parties and then we have:

$$\rho_A^{(K)} = \frac{\sum_{j,j',h} \prod_{k=1}^K \left(1 + \frac{\left(e^{i\varphi_j^{(k)}} + e^{i\varphi_{j'}^{(k)}} \right) \left(e^{-i\varphi_h^{(k)}} + e^{-i\varphi_{j'}^{(k)}} \right)}{2} + e^{i\varphi_j^{(k)}} e^{-i\varphi_h^{(k)}} \right) c_j |c_{j'}|^2 c_h^* |\phi_j\rangle \langle \phi_h|}{\sum_{j,j'} \prod_{k=1}^K (3 + \cos t_k (\lambda_j - \lambda_{j'})) |c_j|^2 |c_{j'}|^2}. \quad (53)$$

Therefore:

$$E^{(K)} = \text{Tr} \left(H \rho_A^{(K)} \right) = \frac{\sum_{j,j'} \lambda_j \prod_{k=1}^K \left(1 + \frac{1}{3} \cos(\varphi_j^{(k)} - \varphi_{j'}^{(k)}) \right) |c_j|^2 |c_{j'}|^2}{\sum_{j,j'} \prod_{k=1}^K \left(1 + \frac{1}{3} \cos(\varphi_j^{(k)} - \varphi_{j'}^{(k)}) \right) |c_j|^2 |c_{j'}|^2}. \quad (54)$$

and:

$$\mathbb{E} [E^{(K)}] = \left(\frac{1}{2\pi} \right)^{NK} \int_0^{2\pi} \dots \int_0^{2\pi} \frac{\sum_{j,j'} \lambda_j \prod_{k=1}^K \left(1 + \frac{1}{3} \cos(\varphi_j^{(k)} - \varphi_{j'}^{(k)}) \right) |c_j|^2 |c_{j'}|^2}{\sum_{j,j'} \prod_{k=1}^K \left(1 + \frac{1}{3} \cos(\varphi_j^{(k)} - \varphi_{j'}^{(k)}) \right) |c_j|^2 |c_{j'}|^2} d\varphi_1^{(1)} \dots d\varphi_N^{(K)}. \quad (55)$$

This integral is also hard to be solved analytically. However, let us make the following observation:

Proposition 2. Eq. (55) can be approximated as:

$$\begin{aligned}
& \left(\frac{1}{2\pi}\right)^{NK} \int_0^{2\pi} \cdots \int_0^{2\pi} \frac{\sum_{j,j'} \lambda_j \prod_{k=1}^K \left(1 + \frac{1}{3} \cos(\varphi_j^{(k)} - \varphi_{j'}^{(k)})\right) |c_j|^2 |c_{j'}|^2}{\sum_{j,j'} \prod_{k=1}^K \left(1 + \frac{1}{3} \cos(\varphi_j^{(k)} - \varphi_{j'}^{(k)})\right) |c_j|^2 |c_{j'}|^2} d\varphi_1^{(1)} \cdots d\varphi_N^{(K)} \\
& \approx \frac{\left(\frac{1}{2\pi}\right)^{NK} \int_0^{2\pi} \cdots \int_0^{2\pi} \sum_{j,j'} \lambda_j \prod_{k=1}^K \left(1 + \frac{1}{3} \cos(\varphi_j^{(k)} - \varphi_{j'}^{(k)})\right) |c_j|^2 |c_{j'}|^2 d\varphi_1^{(1)} \cdots d\varphi_N^{(K)}}{\left(\frac{1}{2\pi}\right)^{NK} \int_0^{2\pi} \cdots \int_0^{2\pi} \sum_{j,j'} \prod_{k=1}^K \left(1 + \frac{1}{3} \cos(\varphi_j^{(k)} - \varphi_{j'}^{(k)})\right) |c_j|^2 |c_{j'}|^2 d\varphi_1^{(1)} \cdots d\varphi_N^{(K)}} \\
& = \frac{\mathbb{E} \left[\sum_{j,j'} \lambda_j \prod_{k=1}^K \left(1 + \frac{1}{3} \cos(\varphi_j^{(k)} - \varphi_{j'}^{(k)})\right) |c_j|^2 |c_{j'}|^2 \right]}{\mathbb{E} \left[\sum_{j,j'} \prod_{k=1}^K \left(1 + \frac{1}{3} \cos(\varphi_j^{(k)} - \varphi_{j'}^{(k)})\right) |c_j|^2 |c_{j'}|^2 \right]} \\
& = \frac{\left(\left(\frac{4}{3}\right)^K - 1\right) \sum_j \lambda_j |c_j|^4 + \sum_j \lambda_j |c_j|^2}{\left(\left(\frac{4}{3}\right)^K - 1\right) \sum_j |c_j|^4 + 1}.
\end{aligned} \tag{56}$$

Proof. Let's denote:

$$g(\varphi) = \sum_{j,j'} \prod_{k=1}^K \left(1 + \frac{1}{3} \cos(\varphi_j^{(k)} - \varphi_{j'}^{(k)})\right) |c_j|^2 |c_{j'}|^2. \tag{57}$$

In order to make the approximation accurate enough, we are going to show that the variance of $g(\varphi)$:

$$\sigma^2[g(\varphi)] = \mathbb{E}[g^2(\varphi)] - (\mathbb{E}[g(\varphi)])^2, \tag{58}$$

is sufficiently small compared to $\mathbb{E}[g(\varphi)]$. Since:

$$\mathbb{E}[g(\varphi)] = \left(\left(\frac{4}{3}\right)^K - 1\right) \sum_j |c_j|^4 + 1, \tag{59}$$

and:

$$\begin{aligned}
\mathbb{E}[g^2(\varphi)] &= \left(\left(\frac{4}{3}\right)^K \sum_j |c_j^{(0)}|^4\right)^2 + 2 \left(\frac{4}{3}\right)^K \left(\sum_j |c_j^{(0)}|^4\right) \times \left(\sum_{j \neq j'} |c_j|^2 |c_{j'}|^2\right) + \\
&\quad \sum_{\substack{\{j\}, \{j'\}, \{l\}, \{l'\} \text{ and} \\ \{j, l\}, \{j'\}, \{l'\} \text{ and} \\ \{j, l'\}, \{j'\}, \{l\} \text{ and} \\ \{j', l\}, \{j\}, \{l'\} \text{ and} \\ \{j', l'\}, \{j\}, \{l\}}} |c_j|^2 |c_{j'}|^2 |c_l|^2 |c_{l'}|^2 + \left(\frac{19}{18}\right)^K \sum_{\substack{\{j, l\}, \{j', l'\} \text{ and} \\ \{j, l'\}, \{j', l\}}} |c_j|^2 |c_{j'}|^2 |c_l|^2 |c_{l'}|^2,
\end{aligned} \tag{60}$$

where the indices inside each braces are equal to each other, and are completely distinct with indices in other braces. For example, $\{j, l\}, \{j'\}, \{l'\}$ denotes $j = l \neq j' \neq l' \neq j = l$. Then, we have:

$$\begin{aligned}
\sigma^2[g(\varphi)] &= \mathbb{E}[g^2(\varphi)] - \mathbb{E}^2[g(\varphi)] \\
&= 2 \left(\left(\frac{19}{18}\right)^K - 1\right) \left(\left(\sum_j |c_j|^4\right)^2 - \sum_j |c_j|^8\right) \\
&\ll \left(\left(\frac{4}{3}\right)^K - 1\right) \sum_j |c_j|^4 + 1 = \mathbb{E}[g(\varphi)].
\end{aligned} \tag{61}$$

Therefore, we approximate this integral of quotient into quotient of integrals, which does not hold in general but holds for this specific case (or any other similar integral forms). $\square$

Therefore, the average energy becomes:

$$\mathbb{E}[E^{(K)}] \approx \frac{\left(\left(\frac{4}{3}\right)^K - 1\right) \sum_j \lambda_j |c_j|^4 + \sum_j \lambda_j |c_j|^2}{\left(\left(\frac{4}{3}\right)^K - 1\right) \sum_j |c_j|^4 + 1}, \quad (62)$$

which introduces energy biases to the system. Moreover, the biases are bounded, as $K \rightarrow \infty$, we have:

$$\mathbb{E}[E^{(\infty)}] \rightarrow \left(\sum_j \lambda_j |c_j^{(0)}|^4\right) / \left(\sum_j |c_j^{(0)}|^4\right). \quad (63)$$

We can further analyze cases where the energy distribution of the Hamiltonian H is continuous, extending the applicability to a broader range of scenarios. We assume H is traceless and local. This kind of H has Gaussian density of states centered at zero [5]. Also, the input product state $|\psi^{(0)}\rangle$ has populations with Gaussian shape in the eigenbasis [11]. Therefore, we have the density of states:

$$D(\lambda) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\lambda^2}{2\sigma^2}\right), \quad (64)$$

and the populations:

$$A(\lambda) = \mathbf{A} \exp\left(-\frac{(\lambda - \mu)^2}{2\xi^2}\right), \quad (65)$$

where σ^2 and ξ^2 denote the variance of density of states and populations, respectively. μ is the mean of the Gaussian shape (not the mean of the overall populations) and $\mathbf{A}$ is the normalization factor. Firstly, we need to find out $\mathbf{A}$, as:

$$\int_{-\infty}^{\infty} D(\lambda) A(\lambda) d\lambda = \mathbf{A} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\lambda^2}{2\sigma^2}\right) \exp\left(-\frac{(\lambda - \mu)^2}{2\xi^2}\right) d\lambda = 1. \quad (66)$$

Therefore:

$$\mathbf{A} = \frac{\exp\left(\frac{\mu^2}{2(\xi^2 + \sigma^2)}\right) \sqrt{\xi^2 + \sigma^2}}{\xi}. \quad (67)$$

Then:

$$\sum_j |c_j^{(0)}|^4 \approx \int_{-\infty}^{\infty} A^2(\lambda) D(\lambda) d\lambda = \frac{\exp\left(\frac{\mu^2}{\xi^2 + \sigma^2} - \frac{\mu^2}{\xi^2 + 2\sigma^2}\right) (\xi^2 + \sigma^2)}{\xi \sqrt{\xi^2 + 2\sigma^2}}, \quad (68)$$

$$\sum_j \lambda_j |c_j^{(0)}|^4 \approx \int_{-\infty}^{\infty} \lambda A^2(\lambda) D(\lambda) d\lambda = \frac{2 \exp\left(\frac{\mu^2}{\xi^2 + \sigma^2} - \frac{\mu^2}{\xi^2 + 2\sigma^2}\right) \mu \sigma^2 (\xi^2 + \sigma^2)}{\xi (\xi^2 + 2\sigma^2)^{\frac{3}{2}}}, \quad (69)$$

$$E^{(0)} = \sum_j \lambda_j |c_j^{(0)}|^2 \approx \int_{-\infty}^{\infty} \lambda A(\lambda) D(\lambda) d\lambda = \frac{\sigma^2}{\xi^2 + \sigma^2} \mu. \quad (70)$$

Therefore:

$$\mathbb{E}[E^{(K)}] = \mu \frac{\sigma^2 \left(1 + \frac{2 \left(-1 + \left(\frac{4}{3}\right)^K\right) e^{\frac{\mu^2 \sigma^2}{\xi^4 + 3\xi^2 \sigma^2 + 2\sigma^4}} (\xi^2 + \sigma^2)^2}{\xi (\xi^2 + 2\sigma^2)^{3/2}}\right)}{(\xi^2 + \sigma^2) \left(1 + \frac{\left(-1 + \left(\frac{4}{3}\right)^K\right) e^{\frac{\mu^2 \sigma^2}{\xi^4 + 3\xi^2 \sigma^2 + 2\sigma^4}} (\xi^2 + \sigma^2)}{\xi \sqrt{\xi^2 + 2\sigma^2}}\right)}. \quad (71)$$

The energy goes up for positive μ and goes down for negative μ . Then, for $K \rightarrow \infty$, we have:

$$\mathbb{E} [E^{(\infty)}] = \frac{\mu}{\frac{\xi^2}{2\sigma^2} + 1}. \quad (72)$$

Then the largest bias we can introduce is approximately:

$$\left| E^{(0)} - \mathbb{E} [E^{(\infty)}] \right| = \left| \frac{\mu}{\frac{\xi^2}{2\sigma^2} + 1} - \frac{\mu}{\frac{\xi^2}{\sigma^2} + 1} \right| = |\mu| \left(\frac{1}{\frac{\xi^2}{2\sigma^2} + 1} - \frac{1}{\frac{\xi^2}{\sigma^2} + 1} \right) \leq |\mu| \frac{1}{2\sqrt{2} + 3} \approx 0.1716|\mu|. \quad (73)$$

And there is no energy bias if $\mu = 0$.

B.1.3 Strong postselection

Now, if we have *strong* postselection, i.e., we only preserve the one with $|000\rangle$ and $|011\rangle$ on all three auxiliary qubits. We have:

$$\rho^{(1)} = \frac{1}{\text{Tr}(\cdot)} \left((I \otimes U + U \otimes I) \rho^{(0)} (I \otimes U^\dagger + U^\dagger \otimes I) + (I \otimes I + U \otimes U) \rho^{(0)} (I \otimes I + U^\dagger \otimes U^\dagger) \right). \quad (74)$$

We define:

$$\begin{aligned} \mathcal{K}_1 &= I \otimes U + U \otimes I; \\ \mathcal{K}_2 &= I \otimes I + U \otimes U. \end{aligned} \quad (75)$$

Therefore, after K rounds, we have (note that $\mathcal{K}$ changes for each round due to different time evolution t_k):

$$\rho^{(K)} = \frac{1}{\text{Tr}(\cdot)} \sum_{j,j',h,h'} \prod_{k=1}^K \left[\left(e^{i\varphi_j^{(k)}} + e^{i\varphi_{j'}^{(k)}} \right) \left(e^{-i\varphi_h^{(k)}} + e^{-i\varphi_{h'}^{(k)}} \right) + (1 + e^{i(\varphi_j^{(k)} + \varphi_{j'}^{(k)})})(1 + e^{-i(\varphi_h^{(k)} + \varphi_{h'}^{(k)})}) \right] c_j c_{j'} c_h^* c_{h'}^* |\phi_j \phi_{j'}\rangle \langle \phi_h \phi_{h'}| \quad (76)$$

Once we trace out one of the systems, we have:

$$\begin{aligned} \rho_A^{(K)} &= \frac{\sum_{j,j',h} \prod_{k=1}^K \left[\left(e^{i\varphi_j^{(k)}} + e^{i\varphi_{j'}^{(k)}} \right) \left(e^{-i\varphi_h^{(k)}} + e^{-i\varphi_{j'}^{(k)}} \right) + (1 + e^{i(\varphi_j^{(k)} + \varphi_{j'}^{(k)})})(1 + e^{-i(\varphi_h^{(k)} + \varphi_{j'}^{(k)})}) \right] c_j |c_{j'}|^2 c_h^* |\phi_j\rangle \langle \phi_h|}{\sum_{j,j'} \prod_{k=1}^K \left[\left(e^{i\varphi_j^{(k)}} + e^{i\varphi_{j'}^{(k)}} \right) \left(e^{-i\varphi_j^{(k)}} + e^{-i\varphi_{j'}^{(k)}} \right) + (1 + e^{i(\varphi_j^{(k)} + \varphi_{j'}^{(k)})})(1 + e^{-i(\varphi_j^{(k)} + \varphi_{j'}^{(k)})}) \right] |c_j|^2 |c_{j'}|^2} \end{aligned} \quad (77)$$

Therefore:

$$E^{(K)} = \text{Tr} (H \rho_A^{(K)}) = \frac{\sum_{j,j'} \lambda_j \prod_{k=1}^K (1 + \cos \varphi_j^{(k)} \cos \varphi_{j'}^{(k)}) |c_j|^2 |c_{j'}|^2}{\sum_{j,j'} \prod_{k=1}^K (1 + \cos \varphi_j^{(k)} \cos \varphi_{j'}^{(k)}) |c_j|^2 |c_{j'}|^2} \quad (78)$$

and:

$$\mathbb{E} [E^{(K)}] = \left(\frac{1}{2\pi} \right)^{NK} \int_0^{2\pi} \cdots \int_0^{2\pi} \frac{\sum_{j,j'} \lambda_j \prod_{k=1}^K (1 + \cos \varphi_j^{(k)} \cos \varphi_{j'}^{(k)}) |c_j|^2 |c_{j'}|^2}{\sum_{j,j'} \prod_{k=1}^K (1 + \cos \varphi_j^{(k)} \cos \varphi_{j'}^{(k)}) |c_j|^2 |c_{j'}|^2} d\varphi_1^{(1)} \cdots d\varphi_N^{(K)} \quad (79)$$

This is still not analytically solvable. However, our numerical studies show that the approximation similar to Proposition 2 still holds for all $K \in \mathbb{Z}^+$:

Conjecture 3. Eq. (79) can be approximated as:

$$\begin{aligned}
& \left(\frac{1}{2\pi}\right)^{NK} \int_0^{2\pi} \cdots \int_0^{2\pi} \frac{\sum_{j,j'} \lambda_j \prod_{k=1}^K \left(1 + \cos \varphi_j^{(k)} \cos \varphi_{j'}^{(k)}\right) |c_j|^2 |c_{j'}|^2}{\sum_{j,j'} \prod_{k=1}^K \left(1 + \cos \varphi_j^{(k)} \cos \varphi_{j'}^{(k)}\right) |c_j|^2 |c_{j'}|^2} d\varphi_1^{(1)} \cdots d\varphi_N^{(K)} \\
& \approx \frac{\left(\frac{1}{2\pi}\right)^{NK} \int_0^{2\pi} \cdots \int_0^{2\pi} \sum_{j,j'} \lambda_j \prod_{k=1}^K \left(1 + \cos \varphi_j^{(k)} \cos \varphi_{j'}^{(k)}\right) |c_j|^2 |c_{j'}|^2 d\varphi_1^{(1)} \cdots d\varphi_N^{(K)}}{\left(\frac{1}{2\pi}\right)^{NK} \int_0^{2\pi} \cdots \int_0^{2\pi} \sum_{j,j'} \prod_{k=1}^K \left(1 + \cos \varphi_j^{(k)} \cos \varphi_{j'}^{(k)}\right) |c_j|^2 |c_{j'}|^2 d\varphi_1^{(1)} \cdots d\varphi_N^{(K)}} \\
& = \frac{\mathbb{E} \left[\sum_{j,j'} \lambda_j \prod_{k=1}^K \left(1 + \cos \varphi_j^{(k)} \cos \varphi_{j'}^{(k)}\right) |c_j|^2 |c_{j'}|^2 \right]}{\mathbb{E} \left[\sum_{j,j'} \prod_{k=1}^K \left(1 + \cos \varphi_j^{(k)} \cos \varphi_{j'}^{(k)}\right) |c_j|^2 |c_{j'}|^2 \right]} \\
& = \frac{\left(\left(\frac{3}{2}\right)^K - 1\right) \sum_j \lambda_j |c_j|^4 + \sum_j \lambda_j |c_j|^2}{\left(\left(\frac{3}{2}\right)^K - 1\right) \sum_j |c_j|^4 + 1}.
\end{aligned} \tag{80}$$

This conjecture is numerically tested and the approximation matches with the data shown in Fig. 4. However, the similar proof of Proposition 2 does not apply here. The reason is as follows. We let:

$$g(\varphi) = \left(\frac{1}{2\pi}\right)^{NK} \int_0^{2\pi} \cdots \int_0^{2\pi} \sum_{j,j'} \prod_{k=1}^K \left(1 + \cos \varphi_j^{(k)} \cos \varphi_{j'}^{(k)}\right) |c_j|^2 |c_{j'}|^2 d\varphi_1^{(1)} \cdots d\varphi_N^{(K)}. \tag{81}$$

Since:

$$\mathbb{E}[g(\varphi)] = \left(\left(\frac{3}{2}\right)^K - 1\right) \sum_j |c_j|^4 + 1, \tag{82}$$

and:

$$\begin{aligned}
\mathbb{E}[g^2(\varphi)] &= \left(\frac{19}{8}\right)^K \sum_j |c_j|^8 + \left(\frac{9}{4}\right)^K \sum_{j \neq j'} |c_j|^4 |c_{j'}|^4 + 2 \left(\frac{3}{2}\right)^K \sum_{j,l \neq l'} |c_j|^4 |c_l|^2 |c_{l'}|^2 + \\
&\quad \sum_{\substack{\{j\}, \{j'\}, \{l\}, \{l'\} \text{ and} \\ \{j,l\}, \{j'\}, \{l'\} \text{ and} \\ \{j,l'\}, \{j'\}, \{l\} \text{ and} \\ \{j',l\}, \{j\}, \{l'\} \text{ and} \\ \{j',l'\}, \{j\}, \{l\}}} |c_j|^2 |c_{j'}|^2 |c_l|^2 |c_{l'}|^2 + \left(\frac{5}{4}\right)^K \sum_{\substack{\{j,l\}, \{j',l'\} \text{ and} \\ \{j,l'\}, \{j',l\}}} |c_j|^2 |c_{j'}|^2 |c_l|^2 |c_{l'}|^2.
\end{aligned} \tag{83}$$

Therefore:

$$\begin{aligned}
\sigma^2[g(\varphi)] &= \mathbb{E}[g^2(\varphi)] - \mathbb{E}^2[g(\varphi)] \\
&= 2 \left(\left(\frac{5}{4}\right)^K - 1 \right) \left(\sum_j |c_j|^4 \right)^2 + \left(\left(\frac{19}{8}\right)^K - \left(\frac{9}{4}\right)^K - 2 \left(\frac{5}{4}\right)^K + 2 \right) \sum_j |c_j|^8.
\end{aligned} \tag{84}$$

Note that $\sigma^2[g(\varphi)] \ll \mathbb{E}[g^2(\varphi)]$ only holds for small K . For large K , $\sigma^2[g(\varphi)] \gg \mathbb{E}[g^2(\varphi)]$. This is why a similar analysis of Proposition 2 does not apply here. Moreover, direct samples of Eq. (79) to compute the integral may have precision problem, as the number of random φ to sample is 2^{NK} , scaled exponentially for both N and K . Again, if this approximation holds, we can proceed to a similar analysis for continuous case, and get the same energy bias bound in the end.

B.2 Energy Variance

In two-device case, we trace out one of the systems in the output state $|\Psi^{(K)}\rangle$ at iteration K and we have:

$$\mathbb{E} \left[\left(\sigma^2 \right)^{(K)} \right] = \mathbb{E} \left[\langle \Psi^{(K)} | I \otimes H^2 | \Psi^{(K)} \rangle \right] - \mathbb{E} \left[\left(\langle \Psi^{(K)} | I \otimes H | \Psi^{(K)} \rangle \right)^2 \right]. \tag{85}$$

Let's analyze the first term first. We denote $\mathbb{E}_1$, $\mathbb{E}_n$, $\mathbb{E}_w$ and $\mathbb{E}_s$ as the expected values for cases of single-device, two-device without postselection, with *weak* postselection, and with *strong* postselection, respectively. Therefore, from the previous analysis, and also the Proposition 2 and 3, we have:

$$\mathbb{E}_1 \left[\left\langle \psi^{(K)} \right| H^2 \left| \psi^{(K)} \right\rangle \right] = \mathbb{E}_n \left[\left\langle \Psi^{(K)} \right| H^2 \left| \Psi^{(K)} \right\rangle \right] = \sum_j \lambda_j^2 |c_j|^2, \quad (86)$$

$$\mathbb{E}_w \left[\left\langle \Psi^{(K)} \right| I \otimes H^2 \left| \Psi^{(K)} \right\rangle \right] = \frac{\left(\left(\frac{4}{3} \right)^K - 1 \right) \sum_j \lambda_j^2 |c_j|^4 + \sum_j \lambda_j^2 |c_j|^2}{\left(\left(\frac{4}{3} \right)^K - 1 \right) \sum_j |c_j|^4 + 1}, \quad (87)$$

$$\mathbb{E}_s \left[\left\langle \Psi^{(K)} \right| I \otimes H^2 \left| \Psi^{(K)} \right\rangle \right] = \frac{\left(\left(\frac{3}{2} \right)^K - 1 \right) \sum_j \lambda_j^2 |c_j|^4 + \sum_j \lambda_j^2 |c_j|^2}{\left(\left(\frac{3}{2} \right)^K - 1 \right) \sum_j |c_j|^4 + 1}. \quad (88)$$

For the second term, we remind that:

$$\mathbb{E}_1 \left[\left(\left\langle \psi^{(K)} \right| H \left| \psi^{(K)} \right\rangle \right)^2 \right] = \left(\frac{1}{2\pi} \right)^N \int_0^{2\pi} \cdots \int_0^{2\pi} \frac{\left(\sum_j |c_j|^2 \lambda_j (1 - \cos \varphi_j) \right)^2}{\sum_j |c_j|^2 (1 - \cos \varphi_j)} d\varphi_1 \cdots d\varphi_N. \quad (89)$$

For the two-device case, remind the Kraus operators we have in the no-postselection case:

$$\begin{aligned} \mathbf{E}_0^{00} &= \frac{1}{4} (I + U) \otimes (I + U) = \frac{1}{4} \sum_{j,j'} (1 + e^{i\varphi_j})(1 + e^{i\varphi_{j'}}) |\phi_j \phi_{j'}\rangle \langle \phi_j \phi_{j'}|; \\ \mathbf{E}_0^{01} &= \frac{1}{4} (I \otimes I - U \otimes U) = \frac{1}{4} \sum_{j,j'} (1 - e^{i(\varphi_j + \varphi_{j'})}) |\phi_j \phi_{j'}\rangle \langle \phi_j \phi_{j'}| = \mathbf{E}_0^{10}; \\ \mathbf{E}_0^{11} &= \frac{1}{4} (I - U) \otimes (I - U) = \frac{1}{4} \sum_{j,j'} (1 - e^{i\varphi_j})(1 - e^{i\varphi_{j'}}) |\phi_j \phi_{j'}\rangle \langle \phi_j \phi_{j'}|; \\ \mathbf{E}_1^{01} &= \frac{1}{4} (U \otimes I - I \otimes U) = \frac{1}{4} \sum_{j,j'} (e^{i\varphi_j} - e^{i\varphi_{j'}}) |\phi_j \phi_{j'}\rangle \langle \phi_j \phi_{j'}| = -\mathbf{E}_1^{10}; \\ \mathbf{E}_1^{00} &= 0 = \mathbf{E}_1^{11}. \end{aligned} \quad (90)$$

We can actually write them in a closed form:

$$\mathbf{E}_s^{ab} = \frac{1}{8} \sum_{j,j'} (1 + (-1)^a e^{i\varphi_j})(1 + (-1)^b e^{i\varphi_{j'}}) + (-1)^s (1 + (-1)^b e^{i\varphi_j})(1 + (-1)^a e^{i\varphi_{j'}}) |\phi_j \phi_{j'}\rangle \langle \phi_j \phi_{j'}|, \quad (91)$$

where a, b, s can be either 0 or 1, denoting the auxiliary qubit result $|sab\rangle$ for each round. Now, we apply $\mathbf{E}$ on the pure state $|\psi^{(0)}\psi^{(0)}\rangle$, the state becomes:

$$\begin{aligned} |\Psi^{(1)}\rangle &= \frac{1}{\sqrt{\langle \psi^{(0)}\psi^{(0)} | (\mathbf{E}_s^{ab})^\dagger \mathbf{E}_s^{ab} | \psi^{(0)}\psi^{(0)} \rangle}} \mathbf{E}_s^{ab} |\psi^{(0)}\psi^{(0)}\rangle = \frac{1}{\sqrt{P_s^{ab}}} \mathbf{E}_s^{ab} |\psi^{(0)}\psi^{(0)}\rangle \\ &= \frac{1}{\sqrt{P_s^{ab}}} \times \frac{1}{8} \sum_{j,j'} (1 + (-1)^a e^{i\varphi_j})(1 + (-1)^b e^{i\varphi_{j'}}) + (-1)^s (1 + (-1)^b e^{i\varphi_j})(1 + (-1)^a e^{i\varphi_{j'}}) \\ &\quad c_j c_{j'} |\phi_j \phi_{j'}\rangle, \end{aligned} \quad (92)$$

where P_s^{ab} denotes the probability of applying operation $\mathbf{E}_s^{ab}$ on the state $|\psi^{(0)}\psi^{(0)}\rangle$. Therefore:

$$\langle \Psi^{(1)} | I \otimes H | \Psi^{(1)} \rangle = \frac{1}{P_s^{ab}} \frac{1}{64} \sum_{j,j'} |c_j|^2 |c_{j'}|^2 \lambda_{j'} \Theta_s^{ab}, \quad (93)$$

where:

$$\begin{aligned}
\Theta_0^{00} &= 16(1 + \cos \varphi_j)(1 + \cos \varphi_{j'}) = 64 \cos^2 \frac{\varphi_j}{2} \cos^2 \frac{\varphi_{j'}}{2}; \\
\Theta_0^{01} &= 8(1 - \cos(\varphi_j + \varphi_{j'})) = 16 \sin^2 \frac{\varphi_j + \varphi_{j'}}{2} = \Theta_0^{10}; \\
\Theta_0^{11} &= 16(1 - \cos \varphi_j)(1 - \cos \varphi_{j'}) = 64 \sin^2 \frac{\varphi_j}{2} \sin^2 \frac{\varphi_{j'}}{2}; \\
\Theta_1^{01} &= 8(1 - \cos(\varphi_j - \varphi_{j'})) = 16 \sin^2 \frac{\varphi_j - \varphi_{j'}}{2} = \Theta_1^{10}; \\
\Theta_1^{00} &= 0 = \Theta_1^{11},
\end{aligned} \tag{94}$$

and:

$$P_s^{ab} = \frac{1}{64} \sum_{j,j'} |c_j|^2 |c_{j'}|^2 \Theta_s^{ab}. \tag{95}$$

Extending it to K rounds, we have:

$$\langle \Psi^{(K)} | I \otimes H | \Psi^{(K)} \rangle = \frac{1}{P_s^{ab}} \frac{1}{64^K} \sum_{j,j'} |c_j|^2 |c_{j'}|^2 \lambda_{j'} \prod_{k=1}^K \Theta_{s_k}^{a_k b_k}, \tag{96}$$

and:

$$P_s^{ab} = \frac{1}{64^K} \sum_{j,j'} |c_j|^2 |c_{j'}|^2 \prod_{k=1}^K \Theta_{s_k}^{a_k b_k}, \tag{97}$$

with the populations of $|\phi_j \phi_{j'}\rangle$:

$$|c_{jj'}^{(K)}|^2 = \frac{1}{P_s^{ab}} \frac{1}{64^K} |c_j|^2 |c_{j'}|^2 \prod_{k=1}^K \Theta_{s_k}^{a_k b_k}. \tag{98}$$

Notably, when we pick $|sab\rangle = |000\rangle$ or $|011\rangle$ only (in the *strong* postselection criterion), the states on the devices keep the tensor product structure; that is, $c_{jj'}^{(k)}$ can be always written in $c_j^{(k)} c_{j'}^{(k)}$. This implies that the states of Alice and Bob are pure product states after each iteration of operation. Then, under the no postselection case:

$$\begin{aligned}
\mathbb{E}_n \left[\left(\langle \Psi^{(K)} | I \otimes H | \Psi^{(K)} \rangle \right)^2 \right] &= \sum_{\mathbf{a}, \mathbf{b}, \mathbf{s}} \mathbb{E} \left[P_s^{ab} \left(\frac{1}{P_s^{ab}} \frac{1}{64^K} \sum_{j,j'} |c_j|^2 |c_{j'}|^2 \lambda_{j'} \prod_{k=1}^K \Theta_{s_k}^{a_k b_k} \right)^2 \right] \\
&= \frac{1}{64^K} \sum_{\mathbf{a}, \mathbf{b}, \mathbf{s}} \mathbb{E} \left[\frac{\left(\sum_{j,j'} |c_j|^2 |c_{j'}|^2 \lambda_{j'} \prod_{k=1}^K \Theta_{s_k}^{a_k b_k} \right)^2}{\sum_{j,j'} |c_j|^2 |c_{j'}|^2 \prod_{k=1}^K \Theta_{s_k}^{a_k b_k}} \right].
\end{aligned} \tag{99}$$

Then, for the *weak* postselection case, a normalization on P_s^{ab} is required. Therefore:

$$W^{\mathbf{a}'\mathbf{b}'} = \frac{P_0^{\mathbf{a}'\mathbf{b}'}}{\sum_{\mathbf{a}, \mathbf{b}} P_0^{\mathbf{ab}}} = \frac{P_0^{\mathbf{a}'\mathbf{b}'}}{\frac{1}{4^K} \sum_{j,j'} |c_j|^2 |c_{j'}|^2 \prod_{k=1}^K (3 + \cos(\varphi_j^{(k)} - \varphi_{j'}^{(k)}))}, \tag{100}$$

and:

$$\begin{aligned}
&\mathbb{E}_w \left[\left(\langle \Psi^{(K)} | I \otimes H | \Psi^{(K)} \rangle \right)^2 \right] \\
&= \sum_{\mathbf{a}, \mathbf{b}} \mathbb{E} \left[W^{\mathbf{ab}} \left(\frac{1}{P_0^{\mathbf{ab}}} \frac{1}{64^K} \sum_{j,j'} |c_j|^2 |c_{j'}|^2 \lambda_{j'} \prod_{k=1}^K \Theta_0^{a_k b_k} \right)^2 \right] \\
&= \frac{1}{16^K} \sum_{\mathbf{a}, \mathbf{b}} \mathbb{E} \left[\frac{\left(\sum_{j,j'} |c_j|^2 |c_{j'}|^2 \lambda_{j'} \prod_{k=1}^K \Theta_0^{a_k b_k} \right)^2}{\sum_{j,j'} |c_j|^2 |c_{j'}|^2 \prod_{k=1}^K \Theta_0^{a_k b_k}} \left(\sum_{j,j'} |c_j|^2 |c_{j'}|^2 \prod_{k=1}^K (3 + \cos(\varphi_j^{(k)} - \varphi_{j'}^{(k)})) \right)^{-1} \right].
\end{aligned} \tag{101}$$

Similarly, for the *strong* postselection, the normalization on P_s^{ab} is:

$$S^{a'_k=b'_k} = \frac{P_0^{a'_k=b'_k}}{\sum_{a_k=b_k} P_0^{a_k=b_k}} = \frac{P_0^{a'_k=b'_k}}{\frac{1}{2^K} \sum_{j,j'} |c_j|^2 |c_{j'}|^2 \prod_{k=1}^K (1 + \cos(\varphi_j^{(k)}) \cos(\varphi_{j'}^{(k)}))} \quad (102)$$

and:

$$\begin{aligned} & \mathbb{E}_s \left[\left(\langle \Psi^{(K)} | I \otimes H | \Psi^{(K)} \rangle \right)^2 \right] \\ &= \sum_{a_k=b_k} \mathbb{E} \left[S^{a_k=b_k} \left(\frac{1}{P_0^{a_k=b_k}} \frac{1}{64^K} \sum_{j,j'} |c_j|^2 |c_{j'}|^2 \lambda_{j'} \prod_{k=1}^K \Theta_0^{a_k=b_k} \right)^2 \right] \\ &= \frac{1}{32^K} \sum_{a_k=b_k} \mathbb{E} \left[\frac{\left(\sum_{j,j'} |c_j|^2 |c_{j'}|^2 \lambda_{j'} \prod_{k=1}^K \Theta_0^{a_k=b_k} \right)^2}{\sum_{j,j'} |c_j|^2 |c_{j'}|^2 \prod_{k=1}^K \Theta_0^{a_k=b_k}} \left(\sum_{j,j'} |c_j|^2 |c_{j'}|^2 \prod_{k=1}^K (1 + \cos(\varphi_j^{(k)}) \cos(\varphi_{j'}^{(k)})) \right)^{-1} \right]. \end{aligned} \quad (103)$$

These are the analytical expressions for the expectation value of $\left(\langle \Psi^{(K)} | I \otimes H | \Psi^{(K)} \rangle \right)^2$, which are extremely difficult to compute explicitly. However, a similar line of analysis can be carried out by analogy with the single-device filtering process. As shown in Eq. (92) and (94), the overall state shared between Alice and Bob remains symmetric throughout the protocol; that is, the state is invariant under the exchange of indices j and j' . Consequently, the reduced states of Alice and Bob are always identical. This implies that, asymptotically, the two-device eigenstate preparation protocol prepares two identical eigenstates, provided that cases with ancillary outcome $|s\rangle = |1\rangle$ are excluded. If such outcome occurs, the populations corresponding to the state $|\phi_j\rangle \otimes |\phi_j\rangle$ may vanish, possibly resulting in a final state composed of superpositions of $|\phi_j\rangle \otimes |\phi_{j'}\rangle$ and $|\phi_{j'}\rangle \otimes |\phi_j\rangle$ for $j \neq j'$.

To build intuition, we first focus on the *strong* postselection case. In this scenario, we apply the same filtering factor $(1 \pm \cos \varphi_j^{(k)})$ to the populations of each party, which is the same the single-device case. Also, the global state remains separable, and the time evolutions on both devices are identical at each iteration. As a result, the dominant (unnormalized) population on two identical eigenstates is amplified as:

$$|c_j^{(k)}|^2 \cdot |c_{j'}^{(k)}|^2 = |c_j^{(k-1)}|^2 \times (1 \pm \cos \varphi_j^{(k)}) \cdot |c_{j'}^{(k-1)}|^2 \times (1 \pm \cos \varphi_{j'}^{(k)}), \quad (104)$$

with the larger associated measurement probability given by:

$$\begin{aligned} \frac{P_0^{00 \text{ or } 11}}{P_0^{00} + P_0^{11}} &= \frac{\frac{1}{4} \sum_{j,j'} |c_j^{(k-1)}|^2 |c_{j'}^{(k-1)}|^2 (1 \pm \cos \varphi_j^{(k)}) (1 \pm \cos \varphi_{j'}^{(k)})}{\frac{1}{2} \sum_{j,j'} |c_j^{(k-1)}|^2 |c_{j'}^{(k-1)}|^2 (1 + \cos \varphi_j^{(k)} \cos \varphi_{j'}^{(k)})} \\ &= \frac{\left(1 \pm \sum_j |c_j^{(k-1)}|^2 \cos \varphi_j^{(k)} \right)^2}{2 + 2 \left(\sum_j |c_j^{(k-1)}|^2 \cos \varphi_j^{(k)} \right)^2}, \end{aligned} \quad (105)$$

when $\pm \cos \varphi_j^{(k)} > 0$. The signs $+$ and $-$ correspond to the postselection outcomes P_0^{00} and P_0^{11} , respectively. Note that the resulting probability distribution differs from the single-device case. Since $|c_j^{(k-1)}|^2$ is the dominant population, it is likely that $\pm \sum_j |c_j^{(k-1)}|^2 \cos \varphi_j^{(k)} > 0$. In such cases, it is straightforward to verify that Eq. (105) yields a value larger than or equal to the single-device filtering probability,

$$P_{0 \text{ or } 1} = \frac{1}{2} \pm \frac{1}{2} \sum_j |c_j^{(k-1)}|^2 \cos \varphi_j^{(k)},$$

meaning that the probability of amplifying the dominant $|c_j^{(k-1)}|^2$ is higher in the two-device setting. This demonstrates that the two-device protocol with *strong* postselection can outperform the single-device case, albeit at the cost of the exponential postselection overhead.

In the case of *weak* postselection, the accepted outcomes include not only those from the *strong* postselection (i.e., $|sab\rangle = |000\rangle$ and $|011\rangle$), but also additional measurement results such as $|001\rangle$ and $|010\rangle$. These outcomes can generally induce entanglement between Alice and Bob's subsystems. Nevertheless, similar analytical arguments apply, and numerical evidence suggests that the *weak* postselection still outperforms the single-device algorithm, while maintaining a bounded postselection overhead. As the energy variance decreases with successive successful iterations, the entanglement between the two parties gradually diminishes, and the overall state increasingly approximates a product of two identical eigenstates shared between Alice and Bob.

Note that the P_s^{ab} is analytically calculable for both *weak* and *strong* postselection scenarios. In the *weak* postselection, the cumulative success rate:

$$\begin{aligned} P_w &= \frac{1}{64^K} \sum_{j,j'} |c_j^{(0)}|^2 |c_{j'}^{(0)}|^2 \prod_{k=1}^K \frac{1}{4\pi^2} \int_0^{2\pi} \int_0^{2\pi} (64 - 32 \sin^2 \frac{\varphi_j^{(k)} - \varphi_{j'}^{(k)}}{2}) d\varphi_j^{(k)} d\varphi_{j'}^{(k)} \\ &= \sum_j |c_j^{(0)}|^4 + \left(\frac{3}{4}\right)^K \sum_{j \neq j'} |c_j^{(0)}|^2 |c_{j'}^{(0)}|^2, \end{aligned} \quad (106)$$

which is lower-bounded by $\sum_j |c_j|^4$ with respect to K . This matches with the numerical result where after several iterations the cumulative success rate plateaus. Therefore, the lower-bound of P_w only depends on the populations of the initial state in the eigenbasis of the Hamiltonian H . In the case of product state as the initial state and local Hamiltonian, $\sum_j |c_j|^4$ can be then approximated in Eq. (68) with $\sigma, \xi \sim O(\sqrt{n})$ where n is the number of qubits of the prepared eigenstate [5, 11]. Therefore, this lower-bound has scaling of $A \exp(B/n)$ where A and B are coefficients independent from n .

On the other hand, for the *strong* postselection case:

$$\begin{aligned} P_s &= \frac{1}{64^K} \sum_{j,j'} |c_j^{(0)}|^2 |c_{j'}^{(0)}|^2 \prod_{k=1}^K \frac{1}{4\pi^2} \int_0^{2\pi} \int_0^{2\pi} (64 \cos^2 \frac{\varphi_j^{(k)}}{2} \cos^2 \frac{\varphi_{j'}^{(k)}}{2} + 64 \sin^2 \frac{\varphi_j^{(k)}}{2} \sin^2 \frac{\varphi_{j'}^{(k)}}{2}) d\varphi_j^{(k)} d\varphi_{j'}^{(k)} \\ &= \left(\frac{3}{4}\right)^K \sum_j |c_j^{(0)}|^4 + \left(\frac{1}{2}\right)^K \sum_{j \neq j'} |c_j^{(0)}|^2 |c_{j'}^{(0)}|^2, \end{aligned} \quad (107)$$

which decreases exponentially with K .

B.3 Eigenvalue spreading (variance of the mixed state)

Throughout this work, we primarily discuss the average variance of the pure state ensemble. Specifically, for each repetition, we let the algorithm to run for several iterations, obtaining the output pure state (provided the algorithm does not restart due to postselection). We then compute the energy variance of this output state and store the result. This process is repeated, and the energy variance values are averaged across all repetitions. This averaged value is what we refer to as the average variance of the pure state ensemble, i.e.,

$$\mathbb{E}[(\sigma^2)^{(k)}] = \mathbb{E}[\langle \psi^{(k)} | H^2 | \psi^{(k)} \rangle] - \mathbb{E}[\left(\langle \psi^{(k)} | H | \psi^{(k)} \rangle\right)^2] = \text{Tr}(\rho^{(k)} H^2) - \mathbb{E}[E^{(k)2}], \quad (108)$$

where $E^{(k)} = \langle \psi^{(k)} | H | \psi^{(k)} \rangle$ denotes the energy of pure state $\psi^{(k)}$ at k th iteration.

Alternatively, we can consider an expected ensemble of pure states returned by the algorithm over many iterations, and consider the ensemble as a mixed state $\rho^{(k)}$. This mixed state has an associated energy $E^{(k)} = \text{Tr}(\rho^{(k)} H)$. We can then define another variance, corresponding to the mixed state $\rho^{(k)}$, denoted as $V^{(k)}$:

$$V^{(k)} = \text{Tr}(\rho^{(k)} H^2) - \mathbb{E}[E^{(k)2}] = \text{Tr}(\rho^{(k)} H^2) - \left(\text{Tr}(\rho^{(k)} H)\right)^2. \quad (109)$$

We highlight that this is a different quantity from $\mathbb{E}[(\sigma^2)^{(k)}]$. We refer to $V^{(k)}$ as the *eigenvalue spread* to avoid using the term variance for both concepts. The eigenvalue spread is the variance of mixed state ρ .

For the analysis below, we distinguish different cases for the respective postselection criterion.

B.3.1 No postselection

In this case, there is no energy bias. Therefore, $V^{(K)}$ stays the same as the energy variance of the input state and does not vary for any $K \in \mathbb{Z}^+$, i.e.,

$$V^{(K)} = \sum_j \lambda_j^2 |c_j|^2 - \left(\sum_j \lambda_j |c_j|^2 \right)^2. \quad (110)$$

B.3.2 Weak postselection

For the first term, we use the approximation of Proposition 2 but only change λ_j into λ_j^2 , which still makes the approximation (integral of quotient $\rightarrow$ quotient of integrals) valid. And the second term has been introduced in the section discussing the energy biases of *weak* postselection cases. Therefore, here:

$$V^{(K)} \approx \frac{\left(\left(\frac{4}{3} \right)^K - 1 \right) \sum_j \lambda_j^2 |c_j|^4 + \sum_j \lambda_j^2 |c_j|^2}{\left(\left(\frac{4}{3} \right)^K - 1 \right) \sum_j |c_j|^4 + 1} - \left(\frac{\left(\left(\frac{4}{3} \right)^K - 1 \right) \sum_j \lambda_j |c_j|^4 + \sum_j \lambda_j |c_j|^2}{\left(\left(\frac{4}{3} \right)^K - 1 \right) \sum_j |c_j|^4 + 1} \right)^2, \quad (111)$$

and the bound is:

$$V^{(\infty)} = \frac{\sum_j \lambda_j^2 |c_j|^4}{\sum_j |c_j|^4} - \left(\frac{\sum_j \lambda_j |c_j|^4}{\sum_j |c_j|^4} \right)^2. \quad (112)$$

Similarly, we can also extend this into continuous scenarios. As:

$$\sum_j \lambda_j^2 |c_j^{(0)}|^4 \approx \int_{-\infty}^{\infty} \lambda^2 A^2(\lambda) D(\lambda) d\lambda = \frac{\exp\left(\frac{\mu^2}{\xi^2 + \sigma^2} - \frac{\mu^2}{\xi^2 + 2\sigma^2}\right) \sigma^2 (\xi^2 + \sigma^2) (\xi^4 + 4\mu^2 \sigma^2 + 2\xi^2 \sigma^2)}{\xi (\xi^2 + 2\sigma^2)^{\frac{5}{2}}}; \quad (113)$$

$$\sum_j \lambda_j^2 |c_j^{(0)}|^2 \approx \int_{-\infty}^{\infty} \lambda^2 A(\lambda) D(\lambda) d\lambda = \frac{\sigma^2 (\xi^4 + \mu^2 \sigma^2 + \xi^2 \sigma^2)}{(\xi^2 + \sigma^2)^2}, \quad (114)$$

then we have:

$$V^{(K)} = \frac{\sigma^2}{(\xi^2 + \sigma^2)^2} \cdot \frac{\Gamma + \Xi}{\Theta}, \quad (115)$$

where:

$$\Theta = \left(1 + \frac{\left(-1 + \left(\frac{4}{3} \right)^K \right) e^{\frac{\mu^2 \sigma^2}{\xi^4 + 3\xi^2 \sigma^2 + 2\sigma^4}} (\xi^2 + \sigma^2)}{\xi \sqrt{\xi^2 + 2\sigma^2}} \right)^2, \quad (116)$$

$$\Gamma = -\sigma^2 \left(\mu + \frac{2 \left(-1 + \left(\frac{4}{3} \right)^K \right) e^{\frac{\mu^2 \sigma^2}{\xi^4 + 3\xi^2 \sigma^2 + 2\sigma^4}} \mu (\xi^2 + \sigma^2)^2}{\xi (\xi^2 + 2\sigma^2)^{3/2}} \right)^2, \quad (117)$$

$$\Xi = \left(1 + \frac{\left(-1 + \left(\frac{4}{3}\right)^K\right) e^{\frac{\mu^2 \sigma^2}{\xi^4 + 3\xi^2 \sigma^2 + 2\sigma^4}} (\xi^2 + \sigma^2)}{\xi \sqrt{\xi^2 + 2\sigma^2}} \right) \times \left(\xi^4 + (\mu^2 + \xi^2)\sigma^2 + \frac{\left(-1 + \left(\frac{4}{3}\right)^K\right) e^{\frac{\mu^2 \sigma^2}{\xi^4 + 3\xi^2 \sigma^2 + 2\sigma^4}} (\xi^2 + \sigma^2)^3 (\xi^4 + 4\mu^2 \sigma^2 + 2\xi^2 \sigma^2)}{\xi (\xi^2 + 2\sigma^2)^{5/2}} \right). \quad (118)$$

One can find that $V^{(K)}$ decreases with K and does not depend on whether the input state energy is above or below 0. For the cases when $K \rightarrow \infty$, we have the eigenstate spread decrease:

$$V^{(0)} - V^{(\infty)} = \xi^2 \frac{1}{\frac{\xi^2}{\sigma^2} + 1} - \xi^2 \frac{1}{\frac{\xi^2}{\sigma^2} + 2} = \frac{1}{\left(\frac{\xi^2}{\sigma^2} + 1\right) \left(\frac{\xi^2}{\sigma^2} + 2\right)} \xi^2. \quad (119)$$

B.3.3 Strong postselection

As shown in Proposition 3, the approximation:

$$V^{(K)} \approx \frac{\left(\left(\frac{3}{2}\right)^K - 1\right) \sum_j \lambda_j^2 |c_j|^4 + \sum_j \lambda_j^2 |c_j|^2}{\left(\left(\frac{3}{2}\right)^K - 1\right) \sum_j |c_j|^4 + 1} - \left(\frac{\left(\left(\frac{3}{2}\right)^K - 1\right) \sum_j \lambda_j |c_j|^4 + \sum_j \lambda_j |c_j|^2}{\left(\left(\frac{3}{2}\right)^K - 1\right) \sum_j |c_j|^4 + 1} \right)^2, \quad (120)$$

is conjectured to be true. Moreover, this is also supported by the numerics as in Fig. 3. Therefore, we can also analyze *strong* postselection for continuous case and finally get the same bound of the eigenstate spread.

C Generalization to s devices

In the main text, we discuss a possible way to generalize our algorithm to multiple devices, where we use the cycle permutations to replace the SWAP operations. Here we briefly give more analytical details on the extensions.

For s -device filter, we write the corresponding state in Fig. 5 step by step. Assuming the input state is $\sum_{j_1, \dots, j_s} c_{j_1 \dots j_s} |\phi_{j_1} \dots \phi_{j_s}\rangle$. Before the controlled- D operator, the state is:

$$\frac{1}{2^{\frac{s}{2}} s^{\frac{1}{2}}} \sum_{\alpha=0}^{s-1} |\alpha\rangle \otimes \sum_{a_1, \dots, a_s=0}^1 |a_1 \dots a_s\rangle \otimes \sum_{j_1, \dots, j_s} \exp(i(a_1 \varphi_{j_1} + \dots + a_s \varphi_{j_s})) c_{j_1 \dots j_s} |\phi_{j_1} \dots \phi_{j_s}\rangle, \quad (121)$$

where $|\alpha\rangle$ and $|a_1 \dots a_s\rangle$ denote the top and s -device local auxiliary qudit/qubits, respectively. Then we apply a multi-level controlled- D operator $\sum_{q=0}^{s-1} |q\rangle \langle q| \otimes D^q$, and apply Fourier transform on every auxiliary qudit/qubit. Before the measurement, the state becomes:

$$\begin{aligned} & \frac{1}{2^s s} \sum_{\alpha=0}^{s-1} |\alpha\rangle \otimes \sum_{a_1, \dots, a_s=0}^1 \sum_{b_1, \dots, b_s=0}^1 (-1)^{a_1 b_1 + \dots + a_s b_s} |b_1 \dots b_s\rangle \\ & \otimes \sum_{j_1, \dots, j_s} \left[\exp(i(a_1 \varphi_{j_1} + \dots + a_s \varphi_{j_s})) + \omega^\alpha \exp(i(a_1 \varphi_{j_2} + \dots + a_s \varphi_{j_1})) + \dots + \right. \\ & \left. \omega^{(s-1)\alpha} \exp(i(a_1 \varphi_{j_s} + \dots + a_s \varphi_{j_{s-1}})) \right] c_{j_1 \dots j_s} |\phi_{j_1} \dots \phi_{j_s}\rangle. \end{aligned} \quad (122)$$

One can see that if the top auxiliary s -level qudit clicks at $\alpha \neq 0$, the populations for $|\phi_j \dots \phi_j\rangle$ become 0. That is why we still need to postselect $|\alpha\rangle = |0\rangle$.

By measuring all the auxiliary qudit and qubits with outcome $|0\rangle \otimes |b_1 \cdots b_s\rangle$, we have the state and the corresponding measurement probability as:

$$\frac{1}{2^s s \sqrt{P_0^{b_1 \cdots b_s}}} \sum_{a_1, \dots, a_s=0}^1 (-1)^{a_1 b_1 + \cdots + a_s b_s} \sum_{j_1, \dots, j_s} [\exp(i(a_1 \varphi_{j_1} + \cdots + a_s \varphi_{j_s})) + \omega^\alpha \exp(i(a_1 \varphi_{j_2} + \cdots + a_s \varphi_{j_1})) + \cdots + \omega^{(s-1)\alpha} \exp(i(a_1 \varphi_{j_s} + \cdots + a_s \varphi_{j_{s-1}}))] c_{j_1 \dots j_s} |\phi_{j_1} \cdots \phi_{j_s}\rangle, \quad (123)$$

and,

$$P_0^{b_1 \cdots b_s} = \frac{1}{s^2} \sum_{j_1, \dots, j_s} |c_{j_1 \dots j_s}|^2 \left(\sum_{cyc \varphi_j} \prod_{l=1}^s \left(b_l \sin \frac{\varphi_{j_l}}{2} + (1 - b_l) \cos \frac{\varphi_{j_l}}{2} \right) \right)^2, \quad (124)$$

respectively. From Eq. (123), one can see that due to the symmetry of cyclic permutation, the state for each reduced system will still be identical.

Still, the *strong* postselection case can be analyzed similarly. In this case, $|c_{j_1 \dots j_s}|^2 = |c_{j_1}|^2 \cdots |c_{j_s}|^2$. Once $|b_1 \cdots b_s\rangle = |0 \cdots 0\rangle$ or $|b_1 \cdots b_s\rangle = |1 \cdots 1\rangle$, the dominant population $|c_j|^2$ for each corresponding index j will also be multiplied by the same factor $(1 \pm \cos \varphi_j)$. When $\pm \cos \varphi_j > 0$, the dominant population will be further amplified with increased probability:

$$\frac{(1 \pm \sum_j |c_j|^2 \cos \varphi_j)^s}{(1 + \sum_j |c_j|^2 \cos \varphi_j)^s + (1 - \sum_j |c_j|^2 \cos \varphi_j)^s} \quad (125)$$

as $\pm \sum_j |c_j|^2 \cos \varphi_j$ is more likely to be larger than 0. Note that in this circumstance, this probability increases with number of devices s , which means that increasing s will yield faster eignestate convergence, but with more communication costs between parties. Also, one can find that on average, $P_0^{0 \cdots 0}$ and $P_1^{1 \cdots 1}$ still exponentially decrease with number of iterations K . For *weak* postselection cases, the analytics become more complicated but numerical evidences suggest that 3-device filter still works in the similar way and we expect that it also works for general s . However, notice that D^q ($q \neq 0$) for non-prime s may not constitute full-cycle permutation, which might affect the convergence rate. This has been briefly mentioned in the main text and also discussed in [29]. Finally, for *weak* postselection, via calculations, the success probability on average is bounded by:

$$\sum_j |c_j^{(0)}|^{2s} \quad (126)$$

D Resource cost in the distributed settings

In this appendix, we discuss the resource costs associated with the distributed setting. In our distributed filtering protocol, non-local operations are required. specifically, a control-SWAP gate per iteration that acts across the single auxiliary qubits on each device, facilitated by shared Bell pairs. Also, due to the overhead introduced by postselection, additional resource consumption arises, such as an increased number of controlled time evolutions and greater use of Bell pairs.

We begin by emphasizing that the single-device filter requires only one control- e^{-iHt} operation per iteration and no Bell pairs. Moreover, since there is no postselection involved in the single-device protocol, it always succeeds. Consequently, the total number of control- e^{-iHt} operations grows linearly with the number of iterations k .

In contrast, for the multi-device filter, each party must apply a local control- e^{-iHt} operation per iteration. Thus, the total number of control time evolutions per iteration equals the number of

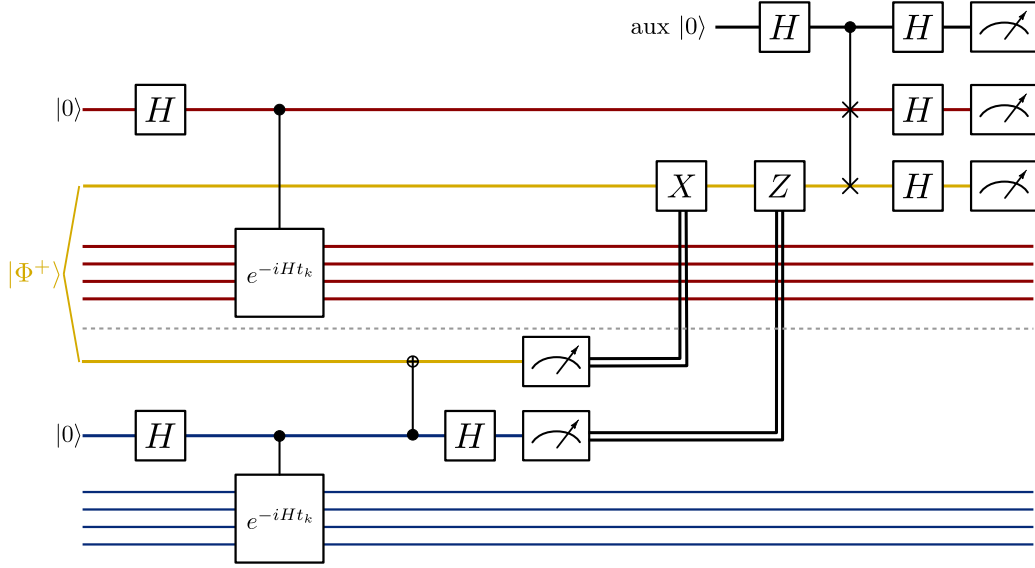


Figure 7: The quantum circuit per iteration to realize the distributed filtering protocol with shared Bell pairs and local operations and classical communication. The qubits above and below the dashed line are held by Alice and Bob, respectively.

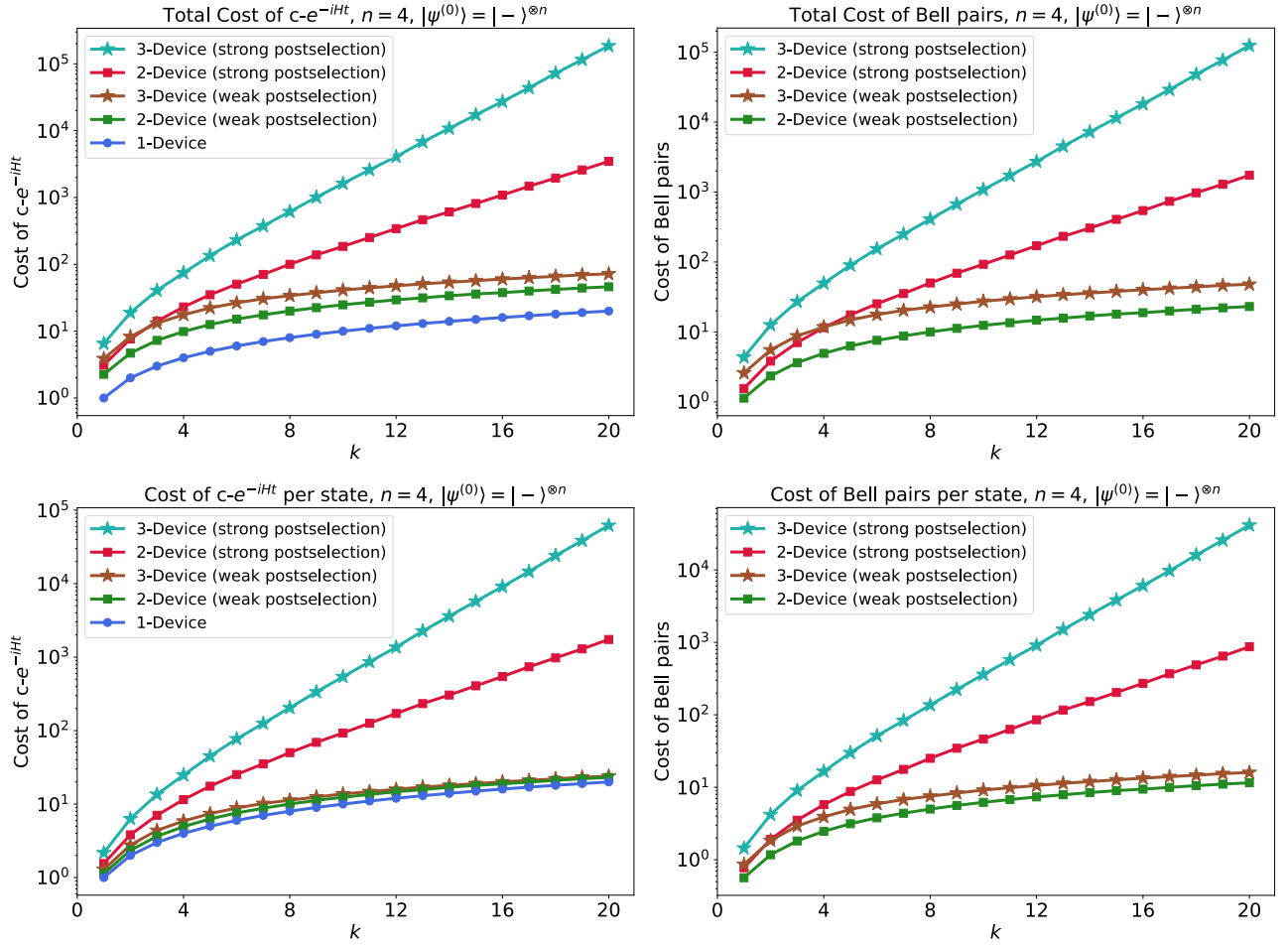


Figure 8: Empirical average number of control time evolutions (control- e^{-iHt}) and consumed Bell pairs for the cases in Fig. 5. The upper panels depict the total cost, while the lower panels present the cost normalized by the number of output states.

participating parties. Furthermore, due to the presence of postselection, the protocol may need to restart multiple times, leading to additional resource costs.

In addition to time evolutions, a shared Bell state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ is required in each iteration to teleport Bob's auxiliary qubit to Alice via local operations and classical communication. Once the teleportation is complete, Alice can apply the control-SWAP gate locally and determine whether the measurement outcomes satisfy the postselection criterion. If the protocol proceeds to the next iteration, the local auxiliary qubit registers can be reset and reused. The detailed circuit implementation is shown in Fig. 7. This teleportation-based approach generalizes straightforwardly to the s -device filter, which requires a total of $s - 1$ Bell pairs per iteration to teleport all auxiliary qubits to Alice, allowing her to perform the control permutations locally.

These resource requirements are closely tied to the cumulative success rate of the protocol. Since each iteration carries a probability of failure and possible restart, resource wastage can accumulate. Using Monte Carlo method, we compute the corresponding resource costs, as illustrated in Fig. 8. In the *weak* postselection regime, the cumulative success rate plateaus after a few iterations, implying that the protocol will almost always succeed beyond a certain k . As a result, the total resource cost ultimately scales linearly with k . However, in the *strong* postselection regime, the cumulative success rate decays exponentially, meaning that the protocol continues to have a non-negligible failure probability at each iteration. Consequently, the total resource cost grows exponentially with k . Moreover, we illustrate the cost normalized by the number of output states as shown in the lower panels of Fig. 8), since the distributed protocols can yield multiple identical eigenstates. They reveal the per-state resource requirements, enabling a fairer comparison among different settings.

We also note that when the controlled-SWAP operation is affected by noise, such as from an imperfect shared Bell pair or a noisy teleportation process, the protocol still functions, though with reduced performance depending on the noise level (cf. Fig. 12 in [16]).

E Additional numerical experiments

In the final subsection of this Appendix, we present additional numerical experiments on several product state instances of the form $|p(\theta)\rangle = (\cos\theta|0\rangle + \sin\theta|1\rangle)^{\otimes n}$, where we choose $\theta \in \{-\pi/4, \pi/4\}$ and $n \in \{4, 5, 6\}$. We show one standard deviation ($\pm\sigma_{\text{data}}$) of the data around the mean value in Fig. 9, 10, 11 and 12. Different from the numerics in the main text, where the average values are computed as accurately as possible, here we fix the number of repetition times to 10^6 for Fig. 9 and 10 and 10^5 for Fig. 11 and 12. For each iteration k , we compute the average value from the data that meet the postselection criteria, and discard the others. Then, especially for *strong* postselection, the data is computed from smaller ensembles, leading to noticeable statistical fluctuations at large k . Also, we note that due to the randomness inherent in the filtering process, the error $|\sigma_{\text{data}}|$ can occasionally be large. To clearly illustrate the decrease in average variance (not the eigenvalue spreading) across different iterations, the y-axis is plotted on a logarithmic scale. As a result, the estimates for the lower error bars can extend below zero, causing the shaded error regions to cover all the areas below the data points on the plots.

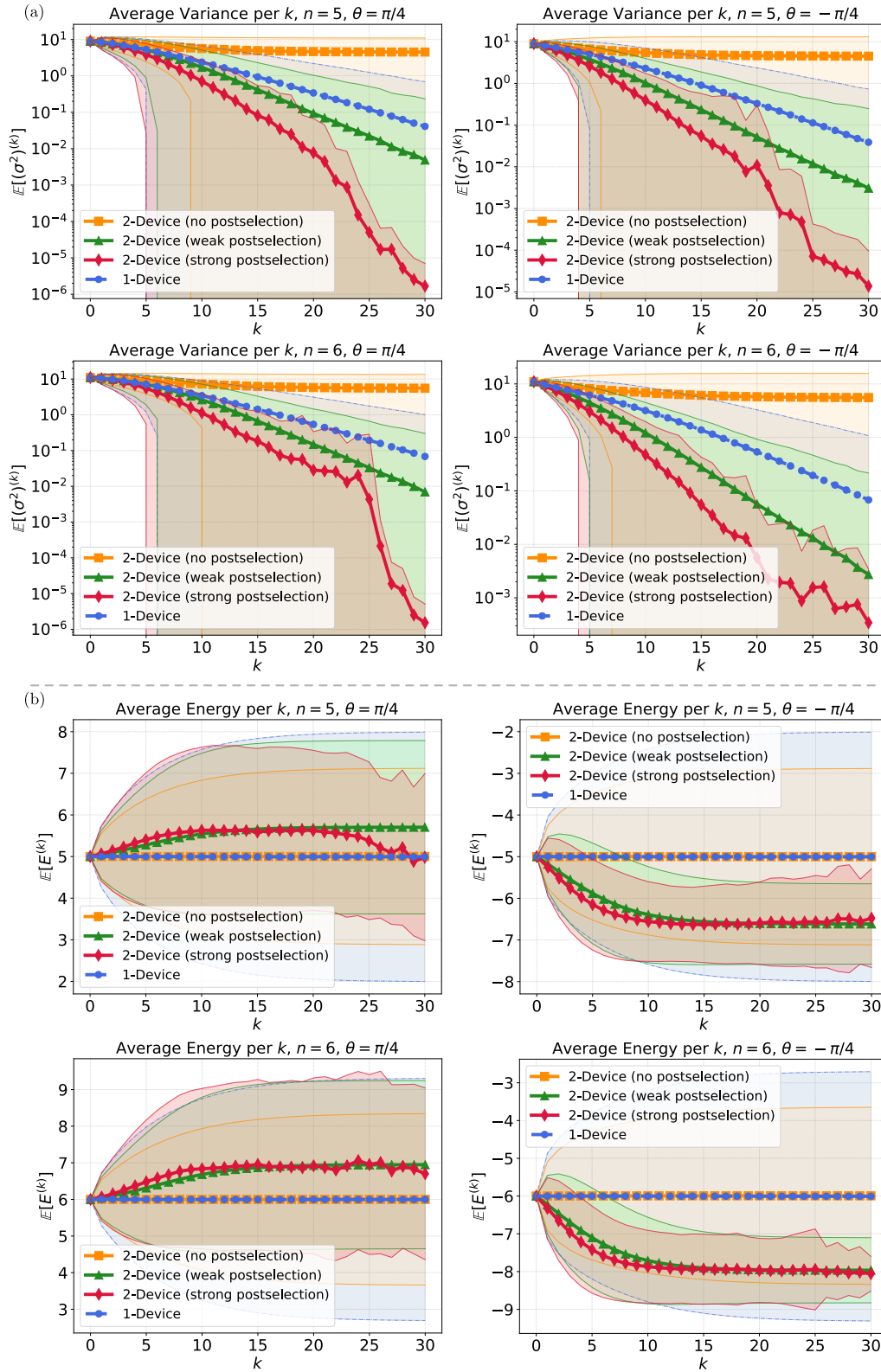


Figure 9: Average variance and energy of the pure state ensemble per iteration k . Four cases are shown including the instances in the main text: $n = 5$ or 6 ; $\theta = \pm\pi/4$. The algorithm is repeated 10^6 times, with each repetition lasting up to 30 iterations. The fluctuations in the average variance for the *strong* postselection case at large k are attributed to the limited data available from the exponentially decreasing number of surviving rounds.

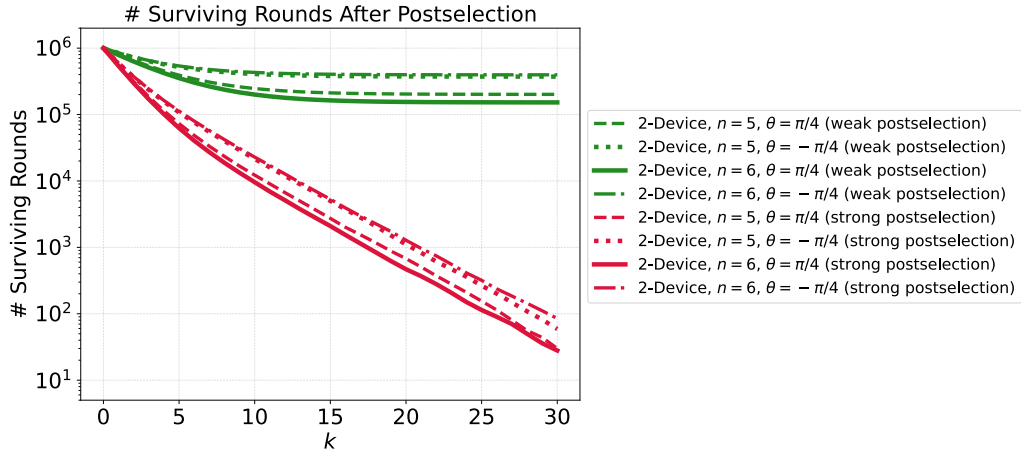


Figure 10: The surviving rounds of repetitions after postselection at each iteration, for all four numerical instances in Fig. 9.

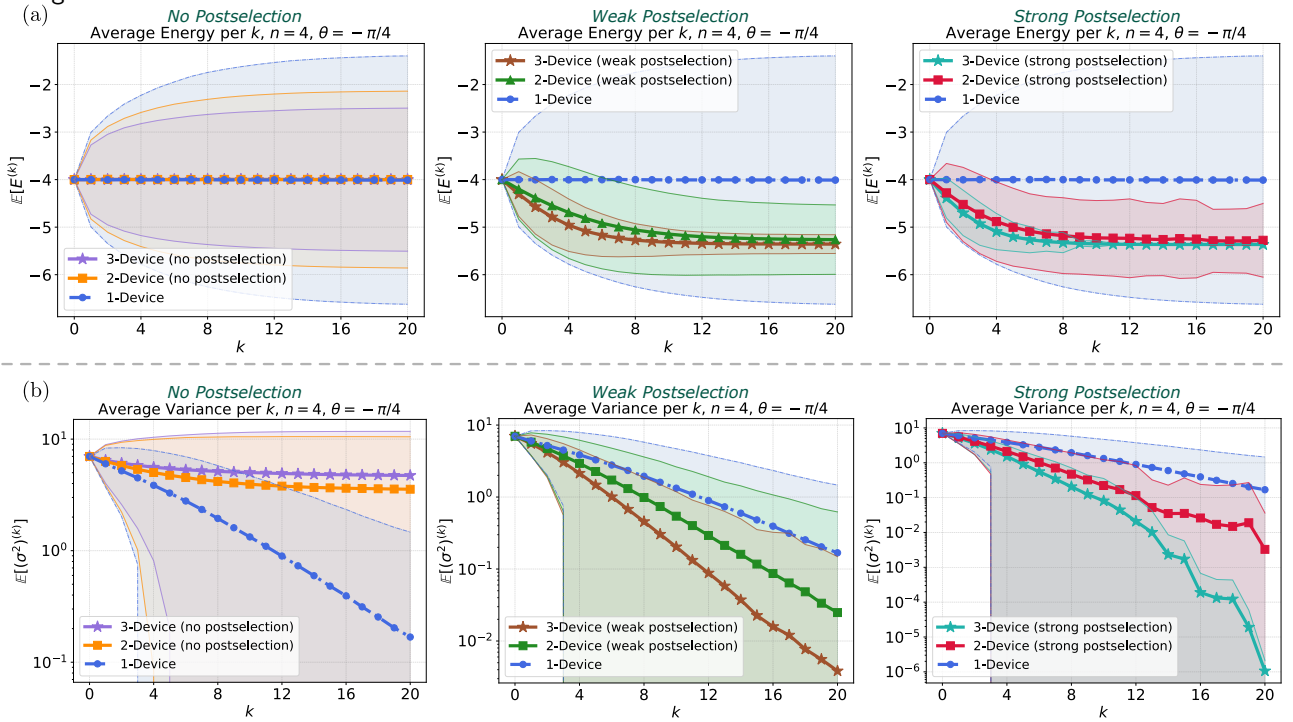


Figure 11: The numerical experiment results with the same inputs in Fig. 5 (comparing single-, two- and three-device algorithms with $n = 4$, $\theta = -\pi/4$) but with fixed repetition numbers (10^5) of the algorithm at the beginning.

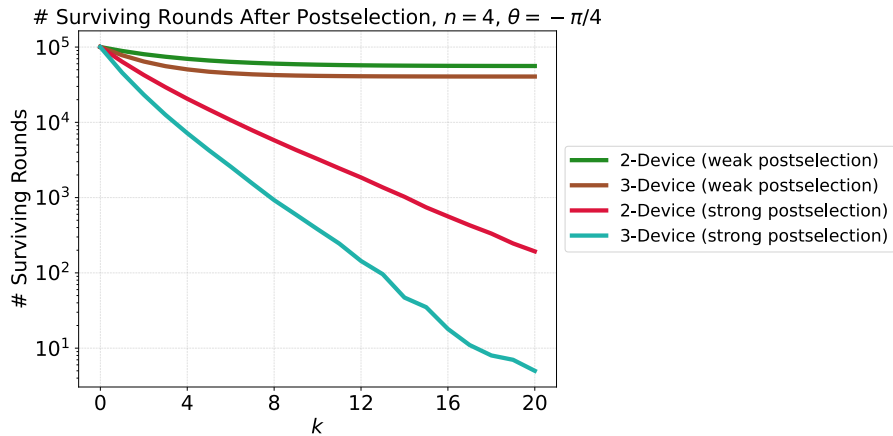


Figure 12: The surviving rounds of repetitions after postselection at each iteration for the numerical instances in Fig. 11.