

Optimal Coherent Quantum Phase Estimation Via Tapering

Dhruvil Patel^{1,4,*}, Shi Jie Samuel Tan^{1,2,3,4,*†}, Yiğit Subaşı⁴, and Andrew T. Sornborger⁴

¹*Department of Computer Science, Cornell University, Ithaca, New York 14850, USA*

²*Joint Center for Quantum Information and Computer Science, NIST/University of Maryland, College Park, Maryland, USA*

³*Department of Computer Science, University of Maryland, College Park, Maryland, USA*

⁴*Information Sciences, Los Alamos National Laboratory, Los Alamos, New Mexico, USA*



(Received 13 June 2025; accepted 13 February 2026; published 2 April 2026)

Quantum phase estimation is one of the fundamental primitives that underpins many quantum algorithms, including Shor’s algorithm for efficiently factoring large numbers. Due to its significance as a subroutine, in this work, we consider the coherent version of the phase estimation problem, where given an arbitrary input state and black-box access to unitaries U and controlled- U , the goal is to estimate the phases of U in superposition. Most existing phase estimation algorithms involve intermediary measurements that disrupt coherence. Only a couple of algorithms, including the standard quantum phase estimation algorithm, consider this coherent setting. However, the standard algorithm only succeeds with a constant probability. To boost this success probability, one can employ the coherent median technique, resulting in an algorithm with asymptotically optimal query complexity (the total number of calls to U and controlled- U). However, this coherent median technique requires a large number of ancilla qubits and a computationally expensive quantum sorting network. To address this, in this work, we propose an improved version of the standard algorithm called the tapered quantum phase estimation algorithm, which leverages tapering (or window) functions commonly used in classical signal processing. Our algorithm achieves the asymptotically optimal query complexity without requiring the expensive coherent median technique to boost success probability. Moreover, we find the absolutely optimal taper—not only in the asymptotic scaling but in terms of exact performance. We provide an efficiently preparable ancilla state based on an approximation of the optimal taper, which incurs at most a factor-of-two increase in the probability of error, thereby maintaining near-optimal performance in practice. In the Appendixes, we give an explicit construction of the taper state preparation circuit. Finally, we derive an error bound for coherent quantum phase estimation when the phase estimate is used as a control and subsequently uncomputed.

DOI: [10.1103/15y6-6zxv](https://doi.org/10.1103/15y6-6zxv)

I. INTRODUCTION

Quantum phase estimation (QPE) has been central to the field of quantum computing since its introduction [1,2]. It has been used in Shor’s algorithm for efficiently factoring large numbers [2], in the Harrow-Hassidim-Lloyd (HHL) algorithm to solve a system of linear equations [3], for quantum amplitude estimation [4], for quantum principal component analysis [5], for fast quantum Merlin-Arthur

(QMA) amplification [6], and as a subroutine in many other applications [7–13].

At its core, the goal of QPE is to estimate the phase of an eigenvalue of a given unitary. Let U be a unitary acting on a d -dimensional Hilbert space, $\mathcal{H}$, and suppose that $e^{2\pi i\theta}$ is one of the eigenvalues of U , where θ lies in the range $[0, 1)$. Then, the QPE problem is to output an estimate $\tilde{\theta}$ that is δ -close to θ with a probability at least $1 - \epsilon$ for some $\delta, \epsilon > 0$.

To solve this problem, it is common to assume that we are given black-box access to U and its controlled version (controlled- U), as well as sample access to the eigenvector $|\psi_\theta\rangle$ of U corresponding to the eigenvalue $e^{2\pi i\theta}$. For our purposes, we define the cost of an algorithm in terms of its query complexity, which is the number of times U and controlled- U are applied. An upper bound on the gate complexity can be obtained from the query

*The first two authors contributed equally to this work.

†Contact author: stan97@umd.edu

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

complexity for a given circuit implementation of U by simple multiplication.

Probably the simplest approach to the QPE problem is the Hadamard test. It estimates the real and imaginary parts of the overlap $\langle \psi_\theta | U | \psi_\theta \rangle$, giving $e^{2\pi i \theta}$. This can then be used to estimate θ . However, this approach requires $O(\delta^{-2} \log(1/\epsilon))$ queries to achieve a desired precision, δ . A quadratic improvement in precision, i.e., $O(\delta^{-1} \log(1/\epsilon))$, can be achieved using the well-known algorithm proposed by Kitaev in 1995 [1]. A crucial insight for this quadratic speedup was to extract the phase value bit-by-bit using controlled- U^{2^j} , rather than simply using controlled- U in each iteration like in the Hadamard test. Interestingly, the authors of Ref. [14] recently showed that any algorithm solving the QPE problem requires $\Omega(\delta^{-1} \log(1/\epsilon))$ queries, making Kitaev's algorithm optimal in that sense.

In many practical applications of QPE, it is not realistic to assume that we have access to the exact eigenvector, $|\psi_\theta\rangle$. Instead, we have access to an arbitrary quantum state, $|\psi\rangle$, that can be written as a superposition of eigenvectors $\{|\psi_{\theta_r}\rangle\}_{r=0}^{d-1}$ of U ,

$$|\psi\rangle = \sum_r c_r |\psi_{\theta_r}\rangle. \quad (1)$$

Suppose that the eigenvalues corresponding to these eigenvectors are $\{e^{2\pi i \theta_r}\}_{r=0}^{d-1}$. A more practical goal is then to estimate phases coherently in superposition, i.e., prepare a state

$$\sum_r c_r (|\omega_r^\delta\rangle + |\omega_r^{\delta\perp}\rangle) |\psi_{\theta_r}\rangle, \quad (2)$$

where

$$|\omega_r^\delta\rangle = \sum_{s: |\theta_r - \tilde{\theta}_s| \leq \delta} d_{r,s} |\tilde{\theta}_s\rangle \quad (3)$$

is the part of the ancilla state that encodes a δ -approximation to θ_r in the computational basis. We want $\| |\omega_r^\delta\rangle \| \geq \sqrt{1-\epsilon}$ which implies that a δ -approximation can be read out with probability $\geq 1-\epsilon$. In practice, the error probability will depend on the true phase: $\| |\omega_r^\delta\rangle \| = \sqrt{1-\epsilon(\theta)}$, and usually we demand $\max_\theta \epsilon(\theta) \leq \epsilon$. This coherent setting is more useful because most algorithms employ QPE as a subroutine, rather than as a standalone algorithm.

Both the Hadamard test and Kitaev's algorithm fall under the category of "iterative" QPE algorithms. An iterative algorithm consists of multiple successive computations, each consisting of simple quantum circuits. Each iteration, however, consists of measurements at the end of the circuit evaluation followed by some classical postprocessing. Applying these algorithms to an arbitrary

state aside from an eigenvector of U causes decoherence due to these inherent measurements in each iteration. Ultimately, this leads to the preparation of an incoherent state as opposed to the pure coherent state, given by Eq. (2).

To the best of our knowledge, only two algorithms exist in the literature that can perform phase estimation coherently [15,16], with the standard QPE algorithm (also known as the textbook QPE algorithm) being the most well known [15]. This algorithm incorporates the (inverse) quantum Fourier transform (QFT) as a subroutine and uses $O(\delta^{-1})$ queries to prepare the desired coherent state with probability at least $8/\pi^2$ in a single run. However, this constant success probability may not be sufficient for quantum algorithms that use QPE as a subroutine. Therefore, boosting this success probability to the desired $1-\epsilon$ becomes extremely crucial. To achieve this, the standard QPE algorithm is executed $O(\log(1/\epsilon))$ times in parallel, and the median of the outputs is computed coherently [6]. Therefore, the overall algorithm, i.e., the standard QPE combined with the coherent median computation for boosting success probability, has query complexity of $O(\delta^{-1} \log(1/\epsilon))$. This query complexity is optimal, matching the corresponding lower bound $\Omega(\delta^{-1} \log(1/\epsilon))$. However, the coherent median step involves using a large number of ancilla qubits and a quantum sorting network [17–19], which is computationally expensive.

An alternative approach to boost the success probability, without using a sorting network, involves using $m = O(\log(1/\epsilon))$ additional ancilla qubits all prepared in the uniform superposition state, but this increases the overall query complexity to $O(\delta^{-1} \epsilon^{-1})$ [15]. This is because the query complexity of the standard QPE algorithm grows exponentially with the number of total ancilla qubits $p = \ell + m$, i.e., $O(\delta^{-1} 2^m)$. Here $\ell = O(\log(1/\delta))$ is the minimum number of ancilla qubits required to achieve precision δ . Substituting $m = O(\log(1/\epsilon))$ into this yields the stated complexity. It is worth noting that this complexity is exponentially worse in ϵ compared to using the coherent median approach.

Therefore, in this work, we investigate the following question and answer it positively: can we improve the standard QPE algorithm such that it maintains the optimal query complexity of $O(\delta^{-1} \log(1/\epsilon))$ without employing the median approach and the associated computationally expensive quantum sorting network?

A. Contributions

We propose an improved version of the standard QPE algorithm, which we call the *tapered QPE (tQPE) algorithm*. The rationale behind this name will become apparent later in the discussion. We show that the tQPE algorithm uses exponentially fewer additional qubits, m , to

achieve a success probability arbitrarily close to one. This improvement directly leads to an exponentially smaller query complexity in terms of ϵ .

To be more precise, we demonstrate that our tQPE algorithm requires only $m = \lceil \log_2 \log(1/\epsilon) \rceil$ additional qubits to achieve a success probability at least $1 - \epsilon$ (see Sec. IV C). We obtain this improvement by framing the problem as an optimization over the choice of the ancilla qubit state that maximizes the success probability. This results in a more effective choice for the initial state of the ancilla register. In the case of the standard QPE algorithm, the ancilla register is initialized to the uniform superposition state. However, we propose to replace this state with the state corresponding to the most-frequency-concentrated discrete prolate spheroidal sequence (DPSS), which is a widely used *window/tapering* function in the field of classical signal processing. This motivates the name of our algorithm. We will provide a brief note on DPSS and on window/tapering functions in general later in Sec. II.

The key idea behind this substitution is that DPSS maximizes signal concentration within a given spectral band. This is particularly important for QPE, as it helps in maximizing the probability of obtaining phase estimates that are δ -close to the true phase θ . To this end, the query complexity of our algorithm is then $O(\delta^{-1} \log(1/\epsilon))$, which scales exponentially better in terms of ϵ than that of the standard QPE algorithm, which, as mentioned before, is $O(\delta^{-1} \epsilon^{-1})$. Furthermore, the query complexity of our algorithm saturates the lower bound $\Omega(\delta^{-1} \log(1/\epsilon))$, which means that it is optimal. Having said that, we would also like to emphasize one of the implications of our result: tQPE can be used directly for fast QMA amplification instead of running the standard QPE algorithm $O(\log(1/\epsilon))$ times for computing the median. This suggests an alternative approach for fast QMA amplification that employs tapering functions.

A natural question following the aforementioned result is whether one can initialize the ancilla register to this DPSS state efficiently. We answer this question by providing explicit algorithms for doing so, which we describe in detail in Sec. VII and Appendix A. We subsequently show that the gate complexity of this algorithm is comparable to that of the standard QPE algorithm for initializing the uniform superposition state, up to $\log \log$ factors.

The tQPE algorithm, like the standard QPE algorithm, involves applying the inverse QFT to the ancilla qubits. However, since tQPE uses exponentially fewer additional qubits, the inverse-QFT circuit is also smaller compared to that used in the standard algorithm. This can be seen as follows. In Ref. [20], the authors showed that the gate complexity of the inverse QFT acting on p qubits is $O(p \log p)$. Therefore, the gate complexity of the inverse QFT in the

case of the standard algorithm is

$$O\left(\left(\log \frac{1}{\delta} + \log \frac{1}{\epsilon}\right) \log \left(\log \frac{1}{\delta} + \log \frac{1}{\epsilon}\right)\right)$$

because the number of ancilla qubits required for this algorithm is $p = O(\log(1/\delta) + \log(1/\epsilon))$. On the other hand, the gate complexity of this transform in the case of tQPE is

$$O\left(\left(\log \frac{1}{\delta} + \log \log \frac{1}{\epsilon}\right) \log \left(\log \frac{1}{\delta} + \log \log \frac{1}{\epsilon}\right)\right)$$

because the ancilla qubits required for this algorithm is $p = O(\log(1/\delta) + \log \log(1/\epsilon))$, a significant improvement.

We also study a special scenario where one is not allowed to use any additional qubits, i.e., $m = 0$ (see Sec. IV B 1) and one wishes to maximize the success probability of outputting one of the two nearest estimates when θ lies exactly in between two phase estimates. The optimal input states for the ancilla qubits for this scenario define a two-dimensional subspace which includes a particular kind of sinusoidal sequence, defined in Eq. (24). Additionally, we perform numerics, plotting the success probability as a function of the distance between the true value and the closest phase estimate (see Sec. IV E). We carry out these numerics for three different input states of the ancilla register: (1) the uniform superposition state, (2) the DPSS state, and (3) the above sinusoidal state. This analysis demonstrates that DPSS performs well over the entire range as compared to the uniform superposition state.

In Sec. V we address the issue of coherent uncomputation of the phase estimate. In many applications such as HHL algorithm [3] and quantum Metropolis sampling [21], QPE is used to implement a unitary that is controlled by the phase estimate which is subsequently uncomputed. In Theorem 4 (proven in Appendix G) we give an error bound on such uses of QPE that applies to any algorithm for coherent QPE. We anticipate that this bound will be useful for analyzing such algorithms as it accounts for multiple errors that in previous analyses had to be treated individually.

In Sec. VI we show that by randomizing over the input phases as in Ref. [22], we can reduce any instance of the QPE problem to an average instance for which DPSS taper is optimal. This proves that the DPSS taper achieves the optimal performance among all tapers not only asymptotically but in an absolute sense. In Sec. VII and Appendix A, we give an explicit description of how approximate taper states with near-optimal performance can be prepared efficiently on a quantum computer. We find that for most applications, $m \leq 4$ additional ancilla qubits might be sufficient.

B. Overview

Let p be the number of ancilla qubits in the QPE circuit, initialized to the all-zeros state. For simplicity, we assume that we have sample access to one of the eigenvectors, $|\psi_\theta\rangle$, of U . In tQPE, as shown in Fig. 1, there are three main steps:

- (1) Initialize the ancilla qubits to a state $|\phi\rangle$. For now, this state serves as a placeholder and will be replaced later with a specific state depending on the problem at hand.
- (2) Apply unitaries controlled- U^{2^j} to the joint state, $|\phi\rangle|\psi_\theta\rangle$.
- (3) Apply inverse QFT to the ancilla qubits.

Notably, the standard QPE algorithm is a special case of tQPE, where the ancilla qubits are initialized to the uniform superposition state, i.e.,

$$|\phi\rangle = \frac{1}{2^{p/2}} \sum_i |i\rangle. \quad (4)$$

The application of the above three steps leads to the following transformation:

$$|0\rangle^{\otimes p} |\psi_\theta\rangle \xrightarrow{\text{tQPE}} \sum_{k=0}^{2^p-1} \hat{\phi}\left(\theta - \frac{k}{2^p}\right) |k\rangle |\psi_\theta\rangle, \quad (5)$$

where $\hat{\phi}(\cdot)$ is the discrete-time Fourier transform, defined explicitly in Eq. (12), of $|\phi\rangle$. Figure 2 shows how the probabilities with which phase estimates are output depend on the true phase θ .

As we can see from the above transformation, the final state of the ancilla register is a superposition of all possible values of phase estimates, $k/2^p$. The success probability of the algorithm is given by

$$\sum_{k: |\theta - \frac{k}{2^p}| \leq \delta} \left| \hat{\phi}\left(\theta - \frac{k}{2^p}\right) \right|^2 = 1 - \epsilon(\theta), \quad (6)$$

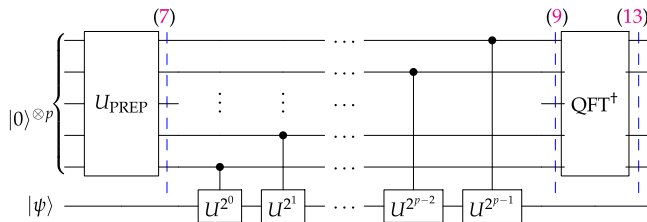


FIG. 1. Tapered quantum phase estimation quantum circuit. The system is initialized in the state $|0\rangle^{\otimes p} |\psi\rangle$, where $|\psi\rangle$ is an arbitrary input state. U_{PREP} is the unitary that prepares the taper state $|\phi\rangle$ (see Appendix A for U_{PREP} that prepares optimal tapers approximately). Dashed blue lines labeled by equation numbers denote the corresponding state at each step.

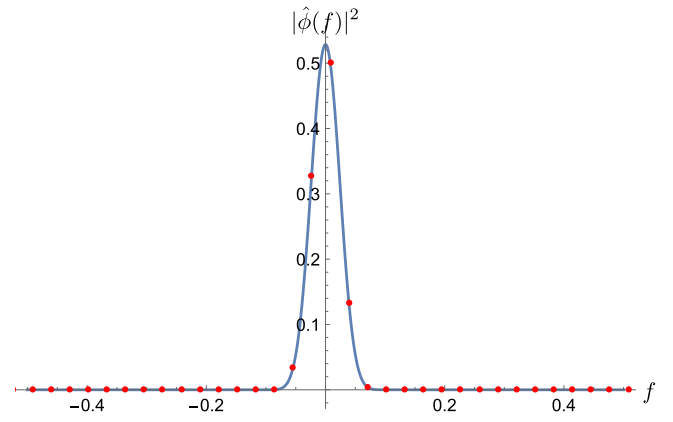


FIG. 2. For the DPSS taper with $p = 5$, we plot the absolute value squared of the taper in the frequency domain (blue curve). The function evaluated at the discrete values $\theta - k/2^p$ (red dots) corresponds to the squared amplitudes of the final ancilla state in Eq. (5) for some choice of θ .

which in general depends on θ as we will see later. Intuitively, we want to choose a state, $|\phi\rangle$, of the ancilla register whose support in the frequency domain is concentrated within the frequency band $[-\delta, \delta]$ and is at least $1 - \epsilon$, where, as mentioned before, ϵ is a uniform upper bound on the error probability across all θ , i.e., $\max_\theta \epsilon(\theta) \leq \epsilon$. Figure 2 shows an example of a taper whose support is concentrated around zero in its frequency domain.

We study error in two different settings: average- and worst-case. In the average-case error setting, we formulate an optimization problem that aims to find ancilla states, $|\phi\rangle$, that minimize the average error of the tQPE algorithm when the phases are sampled uniformly at random. We show that solving these optimization problems is equivalent to solving frequency concentration problems arising in classical signal processing, and the most-frequency-concentrated DPSS naturally emerges as the optimal solution in the average-case setting. We then show how worst-case instances of the QPE problem can be reduced to average-case instances by randomization of the phases, thus proving DPSS taper is optimal in the worst case as well.

Before we discuss related works, we supply a summary table of parameters and their roles to help guide readers who are less familiar with QPE to navigate the dense notation. Readers can find an overview of our notation in Table I.

C. Related work

QPE is a well-researched problem and hitherto many quantum algorithms have been proposed to solve it under different settings. However, most of the quantum algorithms in the literature [1, 14, 23–30] are not suitable for the coherent setting that we are considering in this paper. This point is strongly emphasized in Ref. [16]. As such,

TABLE I. Summary of notation.

Symbol	Domain	Interpretation
$\mathcal{H}$	...	d -dimensional Hilbert space
U	$\mathbf{U}(\mathcal{H})$	Black-box unitary
θ	$[0, 1)$	Eigenphase of U
$ \psi_\theta\rangle$	$\mathcal{H}$	Eigenstate of U with phase θ
$ \psi\rangle$	$\mathcal{H}$	Arbitrary input $\sum_r c_r \psi_{\theta_r}\rangle$
$\hat{\theta}$	$[0, 1)$	Phase estimate
δ	$(0, 1/2]$	Target precision: success if $ \hat{\theta} - \theta \leq \delta$
$\epsilon(\theta)$	$[0, 1]$	Error probability for true phase θ
ϵ	$(0, 1]$	Uniform upper bound on the error probability
$\bar{\epsilon}^{\text{opt}}$	$(0, 1]$	Optimal average error probability for the tQPE algorithm, attained by the DPSS taper
$\bar{\epsilon}^\phi$	$(0, 1]$	Average error probability for tQPE algorithm when $ \phi\rangle$ is used as a taper
$\bar{\epsilon}$	$(0, 1]$	Upper bound on the average error probability
ℓ	$\mathbb{N}$	Minimum number of ancilla qubits required to achieve precision δ
m	$\mathbb{N}$	Number of additional ancilla qubits used to boost success probability
p	$\mathbb{N}$	Total number of ancilla qubits: $\ell + m$
N	$\mathbb{N}$	Number of phase grid points: $2^{\ell+m}$
$\mathcal{H}_t$	...	N -dimensional Hilbert space corresponding to the ancilla/taper register
Δ	$(-1/2N, 1/2N]$	Offset between the true phase and the nearest phase estimate on the grid
K	$\{0, \dots, (N/2) - 1\}$	Used to define number of acceptable phase estimates: $2K + 1$
W	$(0, 0.5)$	Half-bandwidth of the DPSS taper in the frequency domain: $W = \delta - 1/(2N)$
λ_{max}	$[0, 1]$	Maximum eigenvalue of DPSS kernel

only two algorithms, such as those presented in Refs. [15, 16], provide coherent phase estimation algorithms and are thus suitable for use as subroutines in larger algorithms. The approach in Ref. [16] is complementary to ours and achieves similar query complexity. Their algorithm employs block-encoding techniques to obtain the phase estimate bit-by-bit and employs techniques from the quantum singular value transform (QSVT) [31,32] to boost the success probability.

Tapers are not new in the study of quantum algorithms. In 1999, Bužek *et al.* used a sinusoidal taper for the construction of optimal quantum clocks [33]. The authors

of Ref. [34] found that a particular kind of sinusoidal taper is optimal as an N -photon two-mode input state for obtaining an estimate of the phase difference between two arms of an interferometer. The effect of a cosine taper on the quantum phase estimation has been studied in Ref. [35], and the authors showed a cubic improvement of m in terms of error probability ϵ , i.e., $m = O(\log(1/\epsilon^{1/3}))$ as opposed to $m = O(\log(1/\epsilon))$ for the standard QPE algorithm. Additionally, the cosine taper has been employed in QPE-based algorithms, including the HHL algorithm for solving systems of linear equations [3]. Bump functions have found application in spectral estimates based on time series analysis [36]. Tapers have also been explored within quantum spectral filtering methods aimed at efficient state initialization [37]. Furthermore, approximate Gaussian tapers have been utilized in quantum algorithms for spectral density estimation [38], thermal state preparation [39–41], and ground-state energy estimation [42]. The Kaiser taper [43–46] has been considered in QPE [47] and achieves nearly asymptotically optimal scaling. We will further comment on Kaiser tapers and their relation to the DPSS taper in Sec. VIII. Note that none of the tapers discussed above are optimally frequency-concentrated, resulting in side-lobes that warrant further reduction.

Previous works have used randomness to improve the performance of QPE. In 2023, Lu and Lin applied phase randomization to achieve unbiased phase estimation [22]. By applying random single-qubit rotations to the ancilla register, a random offset to θ is applied which removes the bias that plagues QPE. We note that we essentially utilize their technique in our work for our worst-case to average-case reduction to argue for the optimality of our QPE approach. Linden and De Wolf also proved that randomization of the inputs to QFT can help achieve good worst-case performance for phase estimation [48]. A different work by Wan *et al.* showed that randomizing how the unitary powers are sampled can help convert algorithmic errors into stochastic errors, leading to a doubly random approach which once again facilitates the reduction of the worst-case instances to the average-case instance [30].

II. A BRIEF NOTE ON DISCRETE PROLATE SPHEROIDAL SEQUENCES

In the field of classical signal processing and statistics, DPSS, sinusoidal sequences, and other such sequences are referred to as window functions, tapering functions, or simply tapers. These functions are widely used to analyze and modify the frequency spectrum of a given signal. It is important to note that the uniform superposition state used in the standard QPE algorithm is also a type of window function known as a rectangular window. This function is also sometimes referred to as the *tophat* taper, and we will use this name throughout our paper to refer to this taper.

In a series of seminal papers in the field of signal analysis [49–53], Slepian, Pollack, and Landau studied the extent to which a time-limited signal can be band-limited. In other words, they investigated how much the Fourier transform of a signal can be concentrated in a small interval in the frequency domain (i.e., band-limited), given that the signal is only nonzero in a finite interval in the time domain (i.e., time-limited). The discrete-time case was studied in the fifth paper of this series of papers [52]. In this paper, Slepian introduced DPSSs (also called Slepian sequences) as eigenvectors of a kernel arising from a particular frequency concentration algorithm and demonstrated that there exists a DPSS that is both time-limited and maximally band-limited.

For completeness, we provide a detailed derivation of this in Appendix E. In brief, Slepian constructed and optimized a cost function to find a discrete sequence, whose Fourier transform is maximally concentrated in a given bandwidth. The optimization results in a linear kernel, whose leading eigenvector solves the problem. Furthermore, the authors of Ref. [54] showed how to efficiently compute the entries of this eigenvector numerically.

Due to the ability of DPSSs to concentrate the support of the Fourier-transformed function in a small interval in the frequency domain, while also maintaining finite support in the time domain, it is well-suited for a wide variety of applications in signal processing, including signal filtering, and high-resolution spectral and harmonic analysis methods.

Classical signal analysis methods based on Slepian *et al.*'s analysis have resulted in significant advances in spectral and harmonic analysis [52,55–59]. Maximizing signal concentration within a given spectral band is not only desirable for classical signal analysis but is also of special significance for QPE as we will see in this paper. Intuitively, the band-limiting property of a taper is important especially for QPE in the sense that it helps in increasing the probability of outputting phase estimates that are δ -close to the true phase θ .

III. TAPERED QUANTUM PHASE ESTIMATION

For simplicity, we focus on a particular eigenvector $|\psi_\theta\rangle$ of U with eigenvalue $e^{2\pi i\theta}$ and lay out the details of our tQPE algorithm. This approach can be readily extended to a superposition of eigenvectors as well. To begin with, as shown in Fig. 1, we prepare an ancilla register in a state $|\phi\rangle$ defined as

$$|\phi\rangle := \sum_{n=0}^{N-1} \phi[n] |n\rangle, \quad (7)$$

where $\phi[n] \in \mathbb{C}$ for all $n \in \{0, \dots, N-1\}$. Also, $N := 2^p$, where p is the total number of qubits in the ancilla register. Throughout this paper, we refer to the state of the ancilla

register, $|\phi\rangle$, as a taper, and we use the terms “taper register” and “ancilla register” interchangeably. For the time being, please note that $|\phi\rangle$ is an arbitrary state, with a particular choice specified later. We then apply the conditional unitary

$$\sum_{n=0}^{N-1} |n\rangle\langle n| \otimes U^n \quad (8)$$

to $|\phi\rangle |\psi_\theta\rangle$, resulting in the state

$$\sum_{n=0}^{N-1} \phi[n] e^{2\pi i\theta n} |n\rangle |\psi_\theta\rangle. \quad (9)$$

Subsequently, we apply the inverse QFT ($\text{QFT}^{-1} = \text{QFT}^\dagger$) on the ancilla register. This transforms the basis $\{|n\rangle\}_{n=0}^{N-1}$ as

$$|n\rangle \rightarrow \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} e^{-2\pi i n k / N} |k\rangle, \quad (10)$$

giving us the following final state:

$$\sum_{k=0}^{N-1} \left(\frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \phi[n] e^{2\pi i n (\theta - k/N)} \right) |k\rangle |\psi_\theta\rangle. \quad (11)$$

Note that the expression above inside the parentheses is the discrete-time Fourier transform, defined as

$$\hat{\phi}(f) := \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \phi[n] e^{2\pi i n f}, \quad (12)$$

of the time-limited signal, $\phi[n]$, evaluated at the frequency, $\theta - k/N$. Using this, the final state of the algorithm, given by Eq. (11), can be expressed more concisely as

$$\sum_{k=0}^{N-1} \hat{\phi}(\theta - k/N) |k\rangle |\psi_\theta\rangle, \quad (13)$$

where we can think of $\{\hat{\phi}(\theta - k/N)\}_k$ as the probability amplitudes of the discrete frequencies $\{k/N\}_k$ indexed by k . In other words, if the first register corresponding to the taper is measured, the tQPE algorithm outputs a phase estimate, k/N , with probability $|\hat{\phi}(\theta - k/N)|^2$ (see Fig. 2).

In Appendix F, we show that the coefficient in Eq. (13) may be understood as the convolution

$$\hat{\phi}(\theta - f) = (\hat{\phi} * \delta)(\theta - f) \quad (14)$$

evaluated at $f = k/N$, where $*$ denotes convolution and $\delta(\cdot)$ is the Dirac delta function (not to be confused with

the precision parameter δ used elsewhere). From the above equality, observe that the taper is centered about, and independent of, the frequency θ . Conceptually, and important for our discussion below, this distinguishes the taper as a variable function that can take various forms depending on the particular optimization problem of interest.

With that in place, recall that the goal of tQPE is to prepare the final state, given by Eq. (13), such that the following holds:

$$\epsilon(\theta) := \sum_{k: |\theta - k/N| > \delta} \left| \hat{\phi}(\theta - k/N) \right|^2 \leq \epsilon. \quad (15)$$

To simplify the analysis, we choose $\delta = 2^{-\ell-1}$ for some positive integer ℓ without loss of generality. If θ can be expressed exactly on ℓ bits, then there exists a $k \in \{0, 1, \dots, N-1\}$ such that $\theta = k/N$. Otherwise, there is always a phase estimate k'/N such that $|\theta - k'/N| \leq 2^{-p-1}$, and a θ that saturates this bound (i.e., it lies exactly between two possible phase estimates). This constraint naturally arises because the resolution of the ancilla register is 2^{-p} . Therefore, we only consider the case where $\delta \geq 2^{-p-1}$ which implies that $p \geq \ell$. Thus, we set $p = \ell + m$ for some $m \geq 0$ where m can be interpreted as the number of additional qubits used to boost the success probability of estimating phases that are δ -close to θ .

As we can see, the standard QPE algorithm is a special case of the tQPE algorithm, where the taper being used is the tophat taper ($\phi[n] = 1/\sqrt{N}; \forall n$). For $m = 0$ (i.e., no additional qubits used to boost the success probability), the tQPE algorithm using the tophat taper outputs a δ -close phase estimate, i.e., one of the two grid points closest to the true phase, with probability at least $8/\pi^2 \approx 0.81$. As previously stated in the Introduction section, in order to increase the success probability further, a natural approach is to increase m . By doing so, the number of δ -close phase estimates increases, which in turn increases the total probability of outputting such estimates. Specifically, the authors of Ref. [15] showed that it is sufficient to choose $m = \lceil \log_2(1/(2\epsilon) + 1/2) \rceil$ to boost the success probability to at least $1 - \epsilon$ when considering the tophat taper.

Although increasing m can boost the success probability, it also increases the computational cost of the algorithm, since the number of queries to U is proportional to $2^p = 2^{\ell+m}$ (see Fig. 1). It also increases the size of the inverse-QFT circuit because, as mentioned before, the size of this circuit is of the order $O(p \log p)$. Thus, it is of utmost importance to choose the taper, $|\phi\rangle$, optimally so as to minimize m for a given ϵ .

IV. OPTIMAL TAPERS

In what follows, we derive optimal tapers for two settings: first, when the relative position of the true phase with respect to the grid of phase estimates is known; and second,

for the average-case scenario, which we define later. We then analyze the worst-case error behavior of these tapers.

A. The optimization problem

We now state the optimization problem more formally:

$$\begin{aligned} \max_{|\phi\rangle \in \mathcal{H}_t} \quad & \sum_{k: |\theta - k/N| \leq \delta} \left| \hat{\phi}\left(\theta - \frac{k}{N}\right) \right|^2 \\ \text{subject to} \quad & \sum_{n=0}^{N-1} |\phi[n]|^2 = 1. \end{aligned} \quad (16)$$

Here, $\mathcal{H}_t$ denotes the N -dimensional Hilbert space corresponding to the taper register. The objective function above represents the success probability of the tQPE algorithm, i.e., probability of obtaining a phase estimate that is δ -close to the exact value θ , given $|\phi\rangle$ as the taper state. The constraint arises from the fact that $|\phi\rangle$ is a quantum state, and therefore, it must be normalized.

Due to our choice of $\delta = 2^{-\ell-1}$ and $p = \ell + m$, there are generically 2^m estimates that are δ -close to θ . In the special cases when the phase is on a grid point or exactly between two grid points, there are $2^m + 1$ and $2^m + \delta_{m0}$ estimates that are δ -close to θ , respectively. Here, δ_{ij} denotes the Kronecker delta, which equals 1 when $i = j$ and 0 otherwise, and, again, should not be confused with the precision parameter δ used elsewhere.

Depending on the exact value of θ , the index k in the objective function of Eq. (16) can range over different values. This significantly complicates the optimization problem. In order to simplify the above expression further, we define the following:

$$\Delta := \theta - \frac{k^*}{N}, \quad (17)$$

where k^*/N is the phase estimate that is closest to θ on the grid. In other words, Δ is the difference between the best possible phase estimate afforded by the p -qubit taper and the true value, θ . We can now rewrite $\hat{\phi}(\theta - k/N)$ as $\hat{\phi}(\Delta + j/N)$ since the sum is over a dummy index.

We observe that phase estimates corresponding to $|k| \leq K$ with $K = 2^{m-1} - 1$ are always δ -close for $m \geq 2$ and $K = 0$ for $m = 0$. This generically leaves out one phase estimate that is furthest away from the true value. In the special cases when the phase is on a grid point or exactly between two grid points, the miscount is $2 - 2\delta_{m0}$ and 1, respectively. Finally, all tapers of practical interest that we know of have Fourier transforms that are symmetric, have a peak at 0, and decay rapidly, meaning their value at $\Delta \pm (K+1)/N$ will be extremely small for large K or equivalently small ϵ . In light of this, we ignore the contribution of those points to the success probability in the optimization problem.

With these considerations, we restate the optimization problem Eq. (16) as

$$\begin{aligned} & \max_{|\phi\rangle \in \mathcal{H}_t} \sum_{j=-K}^K \left| \hat{\phi} \left(\Delta + \frac{j}{N} \right) \right|^2 \\ & \text{subject to} \quad \sum_{n=0}^{N-1} |\phi[n]|^2 = 1. \end{aligned} \quad (18)$$

To convert the above constrained optimization problem to an unconstrained one, we use the Lagrangian formulation:

$$\begin{aligned} \mathcal{L}(|\phi\rangle, \lambda) = & \sum_{j=-K}^K \left| \hat{\phi} \left(\Delta + \frac{j}{N} \right) \right|^2 \\ & - \lambda \left(\sum_{n=0}^{N-1} |\phi[n]|^2 - 1 \right). \end{aligned} \quad (19)$$

where $\mathcal{L}$ is the Lagrangian and $\lambda \in \mathbb{R}$ is a Lagrange multiplier.

B. Ideal case optimal tapers

Finding the optimal taper is equivalent to finding the stationary point of $\mathcal{L}$ that maximizes the objective function of Eq. (18). The stationary points of $\mathcal{L}$ can be found by setting all the partial derivatives of $\mathcal{L}$ to zero. Doing so, we get the following two conditions (see Appendix B for a detailed derivation):

$$\begin{aligned} & \frac{1}{N} \sum_{n=0}^{N-1} e^{2\pi i \Delta(n-m)} \left(\frac{\sin(\pi(m-n)(2K+1)/N)}{\sin(\pi(m-n)/N)} \right) \phi[n] \\ & = \lambda \phi[m]; \quad \forall m \in \{0, \dots, N-1\}, \end{aligned} \quad (20)$$

$$\sum_{n=0}^{N-1} |\phi[n]|^2 = 1. \quad (21)$$

The first equation is an eigenvalue equation, while the second equation is the normalization constraint. Note that all the eigenpairs $(|\phi\rangle, \lambda)$ of the above eigenvalue equation satisfy both the aforementioned conditions, and therefore, these eigenpairs are the stationary points of $\mathcal{L}$. Now, by substituting all these stationary points $(|\phi\rangle, \lambda)$ into the objective function of Eq. (18), we obtain the following:

$$\sum_{j=-K}^K \left| \hat{\phi} \left(\Delta + \frac{j}{N} \right) \right|^2 = \lambda. \quad (22)$$

This implies that the stationary point $(|\phi\rangle, \lambda)$ that maximizes the objective function is the eigenvector $|\phi\rangle$ with

maximum eigenvalue, λ . Since the objective function is the probability of outputting one of the $2K+1$ phase estimates δ -close to θ , using the eigenvector of Eq. (20) with the maximum eigenvalue as our taper will maximize this probability.

In Appendix C, we establish an explicit connection between the eigenvectors of Eq. (20) and the periodic discrete prolate spheroidal sequences (P-DPSS) [60]. We call the former quantum periodic discrete prolate spheroidal sequences (QP-DPSS), and it is easy to see that they depend on the value of Δ . Then, combining results from Appendix C with known results for the P-DPSS in [61], we observe that the eigenvector of Eq. (20) with the largest eigenvalue has an eigenvalue of 1, regardless of the value of Δ . In other words, there exists a taper for which the tQPE algorithm outputs one of the δ -close $2K+1$ phase estimates with probability 1.

It should not be surprising that the above observation holds true. To understand this intuitively, we can break it down into two cases. In the first case, we consider $\Delta = 0$, that is θ lies exactly on one of the grid points [see Eq. (17)]. In this case, the standard QPE algorithm using the tophat taper always returns θ with probability 1 because now θ itself is one of the possible phase estimates. On the other hand, if we consider the case where $0 < \Delta \leq 1/2N$ or $-1/2N \leq \Delta < 0$, then this case can also be converted into the $\Delta = 0$ case by shifting the grid of possible phase estimates by Δ . To accomplish this, we simply apply a unitary operator parameterized by Δ to the tophat taper. This operation shifts all the 2^p grid points by Δ , such that θ now lies exactly on a grid point. After that, we output θ with probability 1 because it is now a possible phase estimate.

For example, for $K = 0$ one can achieve the success probability 1 with the optimal taper given by

$$\phi^\Delta[n] = \frac{e^{2\pi i \Delta n}}{\sqrt{N}}. \quad (23)$$

As a check, we note that for $\Delta = 0$, we recover the tophat taper, which is known to have success probability 1 when the true phase is on the grid of estimates. In Fig. 3 we plot the success probability of this taper for $\Delta = \pm 1/2N$. The above procedure is described in greater detail in Appendix C.

1. Special case: $\Delta = \pm 1/2N$

In this case, there are pairs of grid points that are equidistant to the exact solution. So, it makes sense to consider the probability of outputting an even number of estimates as opposed to an odd number. More specifically, we will focus on the case where we only consider the nearest two estimates. Equation (23) shows tapers that return the closest estimate with probability one for arbitrary Δ . In the special case $\Delta = \pm 1/2N$, there are actually two such

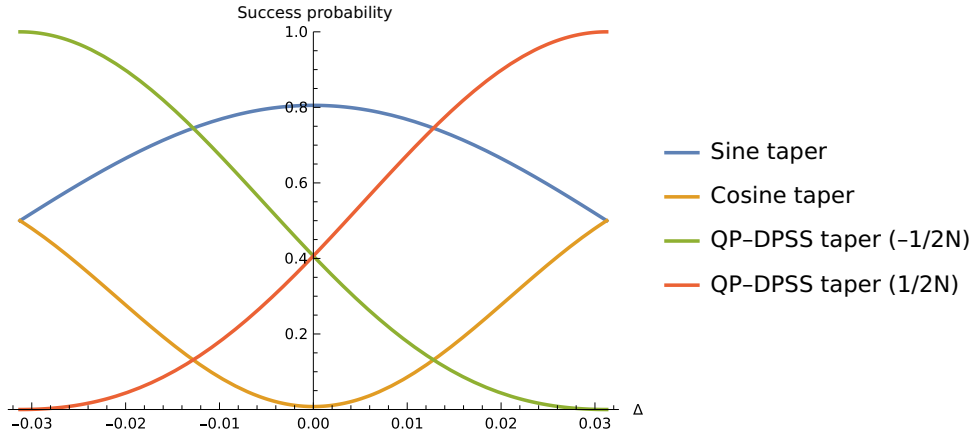


FIG. 3. We plot the probability of the $\phi^{-1/2N}$ (green), $\phi^{1/2N}$ (red), $\phi^{\sin}$ (blue), and $\phi^{\cos}$ (orange) tapers to output the closest phase estimate as a function of Δ , for $N = 2^5$. Both cosine and sine tapers achieve 0.5 at $\Delta = \pm 1/2N$. This means with unit probability one of the two closest estimates will be returned.

tapers: one shown in Eq. (23) and the other obtained by letting $\Delta \rightarrow -\Delta$. One of these tapers outputs the larger and the other the smaller of the two closest estimates with unit probability. Any linear combination of these tapers also succeeds with unit probability. One might be interested in combinations that output each closest estimate with equal probability. One such pair of tapers is

$$\phi^{\sin}[n] = \frac{\sin(\pi n/N)}{\sqrt{N/2}}, \quad (24)$$

$$\phi^{\cos}[n] = \frac{\cos(\pi n/N)}{\sqrt{N/2}}. \quad (25)$$

Although these two tapers perform the same on $\Delta = \pm 1/2N$, their performance on other Δ is very different. In Fig. 3 we show that the sine taper is superior to the cosine taper because it performs better at all other values of Δ . Note that $\phi^{\pm 1/2N}$ has similar properties to the tophat taper by construction but the sine taper behaves qualitatively differently. This is because in constructing the sine taper we used the additional degree of freedom afforded by taking a linear combination of $\phi^{\pm 1/2N}$ to improve the performance of the taper over all Δ . In fact, the sine (cosine) taper can be obtained by minimizing (maximizing) the average-case error defined in the next section among all linear combinations between the two QP-DPSS tapers above.

C. Average-case optimal tapers

From the development above, it is important to note that in general the value of Δ is not known *a priori*. Moreover, since Δ depends on the phase θ , it can differ for distinct phases, particularly when the input state is a superposition of eigenvectors of U . Since we are interested in the coherent QPE problem, the same taper must work well for all

values of Δ in order to be useful. Therefore, in this subsection, we focus on finding the optimal taper that works best on average.

The optimization problem corresponding to this is

$$1 - \bar{\epsilon}^{\text{opt}} := \max_{|\phi\rangle \in \mathcal{H}_t} \mathbb{E}_{\Delta \sim \mathcal{D}} \left[\sum_{j=-K}^K \left| \hat{\phi} \left(\Delta + \frac{j}{N} \right) \right|^2 \right] \quad (26)$$

subject to $\sum_{n=0}^{N-1} |\phi[n]|^2 = 1,$

where $\bar{\epsilon}^{\text{opt}}$ denotes the optimal average error probability. In addition, let us denote the average success probability for any given taper $|\phi\rangle$, that is, the objective function above, as follows:

$$1 - \bar{\epsilon}^{\phi} := \mathbb{E}_{\Delta \sim \mathcal{D}} \left[\sum_{j=-K}^K \left| \hat{\phi} \left(\Delta + \frac{j}{N} \right) \right|^2 \right]. \quad (27)$$

The Lagrangian, $\bar{\mathcal{L}}$, for the modified optimization problem in Eq. (26) is then

$$\bar{\mathcal{L}}(|\phi\rangle, \lambda) = \mathbb{E}_{\Delta \sim \mathcal{D}} \left[\sum_{j=-K}^K \left| \hat{\phi} \left(\Delta + \frac{j}{N} \right) \right|^2 \right] - \lambda \left(\sum_{n=0}^{N-1} |\phi[n]|^2 - 1 \right), \quad (28)$$

where $\mathcal{D}$ is the uniform distribution defined over the interval $[-1/2N, 1/2N]$. Then solving for the above Lagrangian, we get the following eigenvalue equation (see

Appendix D for a detailed derivation):

$$\frac{1}{N} \sum_{n=0}^{N-1} \frac{\sin(\pi(m-n)(2K+1)/N)}{\pi(m-n)} \phi[n] = \lambda \phi[m]; \quad \forall m \in \{0, \dots, N-1\}. \quad (29)$$

We observe that this eigenvalue equation appears commonly in the field of classical signal processing and that DPSSs are eigenvectors of this equation. Now, with a similar reasoning as we used previously for the ideal case, the eigenvector $|\phi\rangle$ that maximizes the objective function, $\bar{\mathcal{L}}$ (i.e., average success probability), is the one with maximum eigenvalue, denoted by $\lambda_{\max}$. Hence, the optimal eigenvector is the DPSS associated with $\lambda_{\max}$, and the optimal value itself equals this maximum eigenvalue,

$$1 - \bar{\epsilon}^{\text{opt}} = 1 - \bar{\epsilon}^{\text{DPSS}} = \lambda_{\max}. \quad (30)$$

Thus, using the DPSS taper achieves the optimal average success probability, equal to $\lambda_{\max}$. Also, note that when we mention DPSS from here on, we refer to the DPSS with $\lambda_{\max}$. For completeness, we provide a brief overview of DPSSs from the classical signal analysis point of view in Appendix E.

Remark 1. An important thing to note is that the derivation of DPSSs from the classical signal processing point of view (provided in Appendix E) relies heavily on the fact that the spectrum of frequencies is continuous. In contrast, we have a discrete frequency spectrum in the case of QPE. However, by taking the average of Δ over the uniform distribution, we effectively transform the problem to a continuous one. Thus, the average-case optimization problem has the same form as the classical continuous-frequency case, and as a result, the DPSS taper turns out to be the optimal taper.

We now state an important theorem from the classical signal processing literature [45] that provides (to date) the most stringent nonasymptotic bounds on the eigenvalues of the DPSS kernel in Eq. (29). Note that this kernel is also known in the literature as the DPSS kernel. The key idea is to use this theorem to derive a nonasymptotic bound on the number of additional ancilla qubits, i.e., m , needed when employing the DPSS taper to achieve the average success probability $1 - \bar{\epsilon}^{\text{DPSS}}$ of at least $1 - \bar{\epsilon}$, for some $\bar{\epsilon} > 0$. Since $1 - \bar{\epsilon}^{\text{opt}} = 1 - \bar{\epsilon}^{\text{DPSS}}$, we also directly have

$$1 - \bar{\epsilon}^{\text{opt}} \geq 1 - \bar{\epsilon}. \quad (31)$$

Theorem 1 ([Corollary 1] [45]). For any $N, K \in \mathbb{N}$ such that $K \in \{0, \dots, N/2 - 1\}$, the maximum eigenvalue $\lambda_{\max}$

of the DPSS kernel in Eq. (29) satisfies

$$\lambda_{\max} \geq 1 - \min \left\{ 8 \exp \left[-\frac{2K-1}{\frac{2}{\pi^2} \log(4N)} \right], 10 \exp \left[-\frac{2K-6}{\frac{2}{\pi^2} \log(100K+75)} \right] \right\}. \quad (32)$$

As the maximum eigenvalue of the DPSS kernel in Eq. (29) is equal to the DPSS average success probability, the bound on the former from the theorem given above naturally bounds the latter. Therefore, one way to infer the above theorem statement is as follows: when the DPSS taper is used, the tQPE algorithm will output one of the $2K+1$ phase estimates closest to θ with an average success probability lower-bounded by the quantity on the right-hand side of Eq. (32).

Applying Theorem 1 for our case, i.e., for the tQPE algorithm that uses the DPSS taper, we obtain the following result that holds for all $N, K \in \mathbb{N}$ which we term as the nonasymptotic regime.

Theorem 2 (Nonasymptotic). Let $\delta \in (0, 1]$. To guarantee that the tQPE algorithm outputs a phase estimate that lies within a δ distance of the true phase, with an average success probability of at least $1 - \bar{\epsilon}$, it suffices to use

$$m = \lceil \log_2 (\lceil 175 (\log(10/\bar{\epsilon}) + 1)^2 \rceil + 1) \rceil + 1 \quad (33)$$

additional ancilla qubits.

Proof. We employ the tQPE algorithm that uses the DPSS taper and outputs one of the δ -close p -bit phase estimates. Here $p = \ell + m$ and ℓ is the minimum positive integer such that $\delta \geq 1/2^{\ell+1}$. This implies that there are 2^m grid points (phase estimates) that are within δ distance to the true phase.

In Theorem 1, we set $N = 2^p$ and $K = 2^{m-1} - 1$ so that we use the bound from Theorem 1 to lower-bound the DPSS average success probability for outputting one of the $2K+1$ p -bit phase estimates that are within δ distance to the true phase. Since we want

$$1 - \bar{\epsilon}^{\text{DPSS}} = \lambda_{\max} \geq 1 - \bar{\epsilon}, \quad (34)$$

we have the following directly from Eq. (32):

$$\bar{\epsilon} \geq \min \left\{ 8 \exp \left[-\frac{2K-1}{\frac{2}{\pi^2} \log(4N)} \right], 10 \exp \left[-\frac{2K-6}{\frac{2}{\pi^2} \log(100K+75)} \right] \right\}. \quad (35)$$

Without loss of generality, we choose to work with the second argument of the min function in the inequality

above, i.e.,

$$\bar{\epsilon} \geq 10 \exp \left[-\frac{2K-6}{\frac{2}{\pi^2} \log(100K+75)} \right]. \quad (36)$$

Taking logarithms and flipping the sign of both sides, we find

$$\log(10/\bar{\epsilon}) \leq \frac{2K-6}{\frac{2}{\pi^2} \log(100K+75)}. \quad (37)$$

By assuming that $K \geq 1$, we first provide a lower bound for the right-hand side of the above inequality in the following way:

$$\frac{2K-6}{\frac{2}{\pi^2} \log(100K+75)} \geq \frac{2K-6}{\frac{2}{\pi^2} \log(175K)} \quad (38)$$

$$\geq \frac{K-3}{\log(175K)} \quad (39)$$

$$= \frac{1}{175} \left(\frac{175K}{\log(175K)} \right) - \frac{3}{\log(175K)} \quad (40)$$

$$\geq \frac{1}{175} \left(\frac{175K}{\log(175K+1)} \right) - 1 \quad (41)$$

$$\geq \frac{\sqrt{175K}}{175} - 1 \quad (42)$$

$$= \sqrt{\frac{K}{175}} - 1, \quad (43)$$

where the second-to-last inequality comes from the following standard logarithmic inequality:

$$\frac{x}{\log(x+1)} \geq \sqrt{x+1} \quad \text{for } x \geq -1. \quad (44)$$

Now, by enforcing the inequality,

$$\log(10/\bar{\epsilon}) \leq \sqrt{\frac{K}{175}} - 1, \quad (45)$$

we obtain $K \geq 175 (\log(10/\bar{\epsilon}) + 1)^2$. In other words, as long as K is at least $\lceil 175 (\log(10/\bar{\epsilon}) + 1)^2 \rceil$, we are guaranteed that Eq. (36) holds, and the error is bounded by $\bar{\epsilon}$. Recall that 2^m of the p -bit phase estimates are δ -close. Setting $K = 2^{m-1} - 1$, we only need

$$m = \lceil \log_2 (\lceil 175 (\log(10/\bar{\epsilon}) + 1)^2 \rceil + 1) \rceil + 1 \quad (46)$$

additional qubits to ensure that our tQPE algorithm has average success probability at least $1 - \bar{\epsilon}$. This concludes the proof. $\blacksquare$

Remark 2. Although the analytical proof for Theorem 2 implies that K needs to be of order $\log^2(1/\bar{\epsilon})$ to achieve an average success probability of at least $1 - \bar{\epsilon}$, for all practical purposes K only needs to be of order $\log(1/\bar{\epsilon})$. For example, for all $\bar{\epsilon} \geq 10^{-81}$, we have that $K \leq 192$, and as a result, we can replace K inside the logarithm in Eq. (37) with 192 and find that $K = \lceil \log(10/\bar{\epsilon}) \rceil + 3$ is sufficient.

Since Theorem 1 holds for any $N \in \mathbb{N}$, our result described in Theorem 2 also holds for any N including the asymptotic regime where $N \rightarrow \infty$ and $\bar{\epsilon} \rightarrow 0$. Consequently, any future improvement or tightening of the bound in Theorem 1 will directly yield a correspondingly sharper and better bound in Theorem 2. If instead we focus on the asymptotic regime, stronger bounds for m can be obtained. More precisely, in this setting, we fix a target precision $\delta \in (0, 1/8]$ and study the limit where the DPSS average success probability approaches 1, i.e., $1 - \bar{\epsilon} \rightarrow 1$ (or equivalently, $\bar{\epsilon} \rightarrow 0$). The following theorem states our result in this asymptotic case.

Theorem 3 (Asymptotic). Let $\delta \in (0, 1/8]$. To guarantee that the tQPE algorithm outputs a phase estimate that lies within a δ distance of the true phase, with an average success probability of at least $1 - \bar{\epsilon}$, it suffices to use

$$m \sim \left\lceil \log_2 \log \frac{1}{\bar{\epsilon}} \right\rceil \quad (47)$$

additional ancilla qubits.

Proof. Here as well we use the tQPE algorithm with the DPSS taper, and the notations are similar to those used in the proof of Theorem 2.

For this proof, we use the result of Ref. [52] that gave asymptotic expressions for the eigenvalues of the DPSS kernel. In Ref. [52], the author denotes the half-bandwidth of the DPSS taper in the frequency domain by W and so to keep the notation consistent with this paper that corresponds to $W = \delta - 1/(2N)$. From Ref. [52], for a fixed W , we have

$$1 - \lambda_{\max} \sim C(W) N^{1/2} e^{-\gamma(W)N}, \quad (48)$$

where

$$C(W) = \pi^{1/2} 2^{9/4} \alpha(W)^{1/4} (2 - \alpha(W))^{-1/2}. \quad (49)$$

Here $\alpha(W)$ and $\gamma(W)$ are given by

$$\alpha = 1 - \cos 2\pi W, \quad (50)$$

$$\gamma = \log \left(1 + \frac{2\sqrt{\alpha}}{\sqrt{2} - \sqrt{\alpha}} \right). \quad (51)$$

By the definition of $\sim$, for every $\eta \in (0, 1)$ there exists a positive number N_0 such that for all $N \geq N_0$, we have

$$1 - \lambda_{\max} \leq (1 + \eta)C(W)N^{1/2}e^{-\gamma(W)N}. \quad (52)$$

Now since $W = (2^m - 1)/(2N)$ and $N = 2^{\ell+m}$, we have $NW = (2^m - 1)/2 =: x$. Plugging $N = x/W$ into Eq. (52), we get

$$1 - \lambda_{\max} \leq A(W, \eta)x^{1/2}e^{-\tau(W)x}, \quad (53)$$

where, again, for brevity, we introduce the following two quantities:

$$A(W, \eta) := \frac{(1 + \eta)C(W)}{\sqrt{W}}, \quad \tau(W) := \frac{\gamma(W)}{W}. \quad (54)$$

We want that $1 - \lambda_{\max} \leq \bar{\epsilon}$. This implies that we want to find an x such that the following holds:

$$A(W, \eta)x^{1/2}e^{-\tau(W)x} \leq \bar{\epsilon}. \quad (55)$$

Let us pick the following x :

$$x = \frac{1}{\tau(W)} \left(\log \left(\frac{A(W, \eta)}{\bar{\epsilon}} \right) + \frac{1}{2} \log \log \left(\frac{A(W, \eta)}{\bar{\epsilon}} \right) \right). \quad (56)$$

Now let us check if this value of x satisfies the condition Eq. (55). We begin by plugging this value in Eq. (55),

$$A(W, \eta)x^{1/2}e^{-\tau(W)x} = \bar{\epsilon} \sqrt{\frac{x}{\log \left(\frac{A(W, \eta)}{\bar{\epsilon}} \right)}}. \quad (57)$$

Now using the inequality

$$\sqrt{\frac{x}{\log \left(\frac{A(W, \eta)}{\bar{\epsilon}} \right)}} \leq \sqrt{\frac{2}{\tau(W)}}, \quad (58)$$

we have

$$A(W, \eta)x^{1/2}e^{-\tau(W)x} \leq \bar{\epsilon} \sqrt{\frac{2}{\tau(W)}}. \quad (59)$$

Furthermore, since for all $W \in (0, 1/2]$, we have $\tau(W) \geq 2\pi$, we can say that

$$A(W, \eta)x^{1/2}e^{-\tau(W)x} \leq \bar{\epsilon}. \quad (60)$$

This completes the check that the value of x that we picked satisfies the condition Eq. (55).

Since $x = (2^m - 1)/2$, any integer

$$m \geq \left\lceil \log_2 \left(\frac{1}{\tau(W)} \left[\log \left(\frac{A(W, \eta)}{\bar{\epsilon}} \right) + \frac{1}{2} \log \log \left(\frac{A(W, \eta)}{\bar{\epsilon}} \right) \right] \right) \right\rceil \quad (61)$$

guarantees $1 - \lambda_{\max} \leq \bar{\epsilon}$. Now for all $W \in (0, 1/4]$, we bound $A(W, \eta)$ from above as

$$A(W, \eta) \leq 8\sqrt{2}\pi. \quad (62)$$

With this, we finally get a fully explicit sufficient bound for m in terms of $\bar{\epsilon}$ and W ,

$$m \geq \left\lceil \log_2 \left(\frac{1}{\tau(W)} \left[\log \left(\frac{8\sqrt{2}\pi}{\bar{\epsilon}} \right) + \frac{1}{2} \log \log \left(\frac{8\sqrt{2}\pi}{\bar{\epsilon}} \right) \right] \right) \right\rceil. \quad (63)$$

Since $\tau(W) \geq 2\pi$ for all $W \in (0, 1/2]$, asymptotically as $\epsilon \rightarrow 0$, we finally have

$$m \sim \left\lceil \log_2 \log \frac{1}{\bar{\epsilon}} \right\rceil \quad (64)$$

in the finite- W or finite- δ (precision) regime. This concludes the proof. $\blacksquare$

Above, we have given both asymptotic and nonasymptotic bounds on the DPSS average error probability $\bar{\epsilon}$. But the true error probabilities can be much better than suggested by these bounds. In Fig. 4 we show the average error

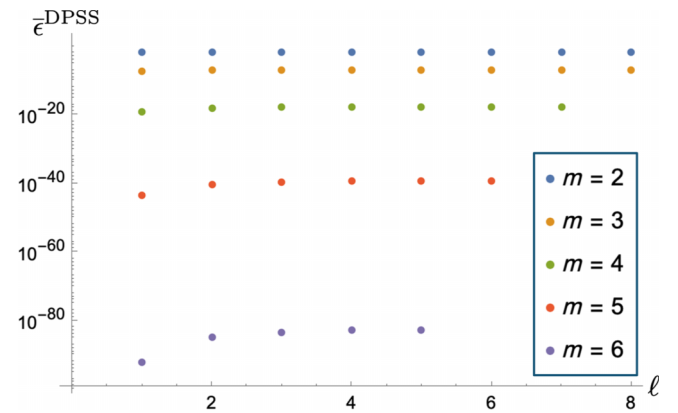


FIG. 4. DPSS average error probability as a function of ℓ for values of $m \leq 6$. We observe that the error probabilities quickly converge with increasing ℓ . Calculations were performed with Mathematica by setting `WorkingPrecision=250`.

probabilities for some values of (ℓ, m) . We observe that already with $m = 6$, the average error probabilities below 10^{-80} are achieved. With $m = 4$, the error probability is below 10^{-17} .

D. Analysis of worst-case error

In the discussion above, we optimized with respect to the average over the unknown offset Δ , which led us naturally to the DPSS taper: it is the taper that does best on average when the true phase can sit anywhere between two grid points. But an average guarantee does not tell us what happens at the worst-case (most unfavorable) Δ , which we denote it as Δ^* . Ideally, we would like to find a taper that has the smallest error probability for its worst-case Δ^* . For most tapers used in classical signal processing (including the tophat and DPSS tapers), $\Delta^* = \pm 1/2N$ corresponds to their worst-case offset, where the corresponding error probability reaches its maximum. However, in general, different tapers have different worst-case Δ^* , which makes the corresponding optimization problem too hard to solve. In this section, we analyze the performance of the DPSS taper in its worst case, which, as mentioned above, is at $\Delta^* = \pm 1/2N$. We find that although not optimized for the worst-case scenario, the DPSS taper performs well and is optimal asymptotically.

To perform such a comparison, we first restate the kernel of the DPSS taper which coincides with the eigenvalue equation shown in Eq. (29),

$$\frac{1}{N} \sum_{n=0}^{N-1} \frac{\sin(\pi(m-n)(2K+1)/N)}{\pi(m-n)} \phi[n] = \lambda \phi[m]; \quad \forall m \in \{0, \dots, N-1\}. \quad (65)$$

Recall that the DPSS taper is the eigenvector of the DPSS kernel with the largest eigenvalue. When considering the success probability of the DPSS taper in the worst case, where $\Delta = \pm 1/2N$, we let $|\phi\rangle$ denote the DPSS taper and adapt the calculations in Appendix D to include an additional complex rotation $e^{2\pi i(n'-n)\Delta}$ that translates the phase estimates in the frequency domain by $\Delta = \pm 1/2N$ to obtain the following expression:

$$\begin{aligned} & \sum_{j=-K}^K \left| \hat{\phi} \left(\Delta + \frac{j}{N} \right) \right|^2 \\ &= \frac{1}{N} \sum_{n,m=0}^{N-1} e^{2\pi i(m-n)\Delta} \\ & \times \frac{\sin(\pi(m-n)(2K+1)/N)}{\sin(\pi(m-n)/N)} \phi^*[n] \phi[m] \quad (66) \end{aligned}$$

$$\begin{aligned} &= \frac{1}{N} \sum_{n,m=0}^{N-1} \cos(\pi(m-n)/N) \\ & \times \frac{\sin(\pi(m-n)(2K+1)/N)}{\sin(\pi(m-n)/N)} \phi^*[n] \phi[m], \quad (67) \end{aligned}$$

where the last equality comes from substituting $\Delta = 1/2N$ and observing how the complex terms from the complex exponential annihilate each other when we run over all summands indexed by m and n . In other words, the success probability for the DPSS taper in the worst case would then be the quantity shown in Eq. (67).

From Theorem 1, we make the observation that the maximum eigenvalue, or the success probability, of the average-case optimal DPSS taper tends to 1 as $N \rightarrow \infty$. This is well-aligned with our intuition because we would expect our DPSS taper to have greater spectral concentration in the central lobe in the frequency domain. At the same time, we note that the worst-case success probability of the DPSS taper also converges to 1 when $N \rightarrow \infty$. This is a result of the complex exponential term in Eq. (66) approaching 1 as $N \rightarrow \infty$, resulting in the worst-case DPSS success probability expression approaching the average-case success probability in the same limit. Numerical evidence in Fig. 5 suggests that the success probability of DPSS taper in the worst case is still $O(\epsilon)$.

E. Comparison of tapers

In this section, we provide numerics for three of the tapers we have discussed so far: the DPSS taper, the sine taper, and the tophat taper. In Figs. 6(a)–6(d), we plot the success probability of each taper as a function of the offset Δ , i.e., the distance between the true phase and the grid point closest to it, for various values of m . The DPSS taper is optimal for the average case. The tophat taper is designed to output the true phase with unit probability when the true phase coincides with a grid point and the sine taper has the property that it outputs one of the two nearest estimates with unit probability when the true phase is exactly between two grid points. Both of these properties are confirmed by the plots. Although the DPSS kernel does not succeed with unit probability for any Δ it does perform well for the entire range, which is consistent with the fact that it is optimal for the average case.

In Fig. 5, we plot the failure probability of the DPSS taper as a function of Δ for various values of l and m (solid blue lines). In contrast to Figs. 6(a)–6(d), we do not plot the sine and tophat tapers. We see that the error profile of the DPSS tapers has more structure compared to the tophat and sine tapers. One trend is that for increasing m the error probability concentrates around its average value for most Δ except near $\Delta = 1/(2N)$, where it shoots up. But even in that region, the error appears to be $O(\epsilon)$. In a sense, for large m , i.e., high success probability, the DPSS taper

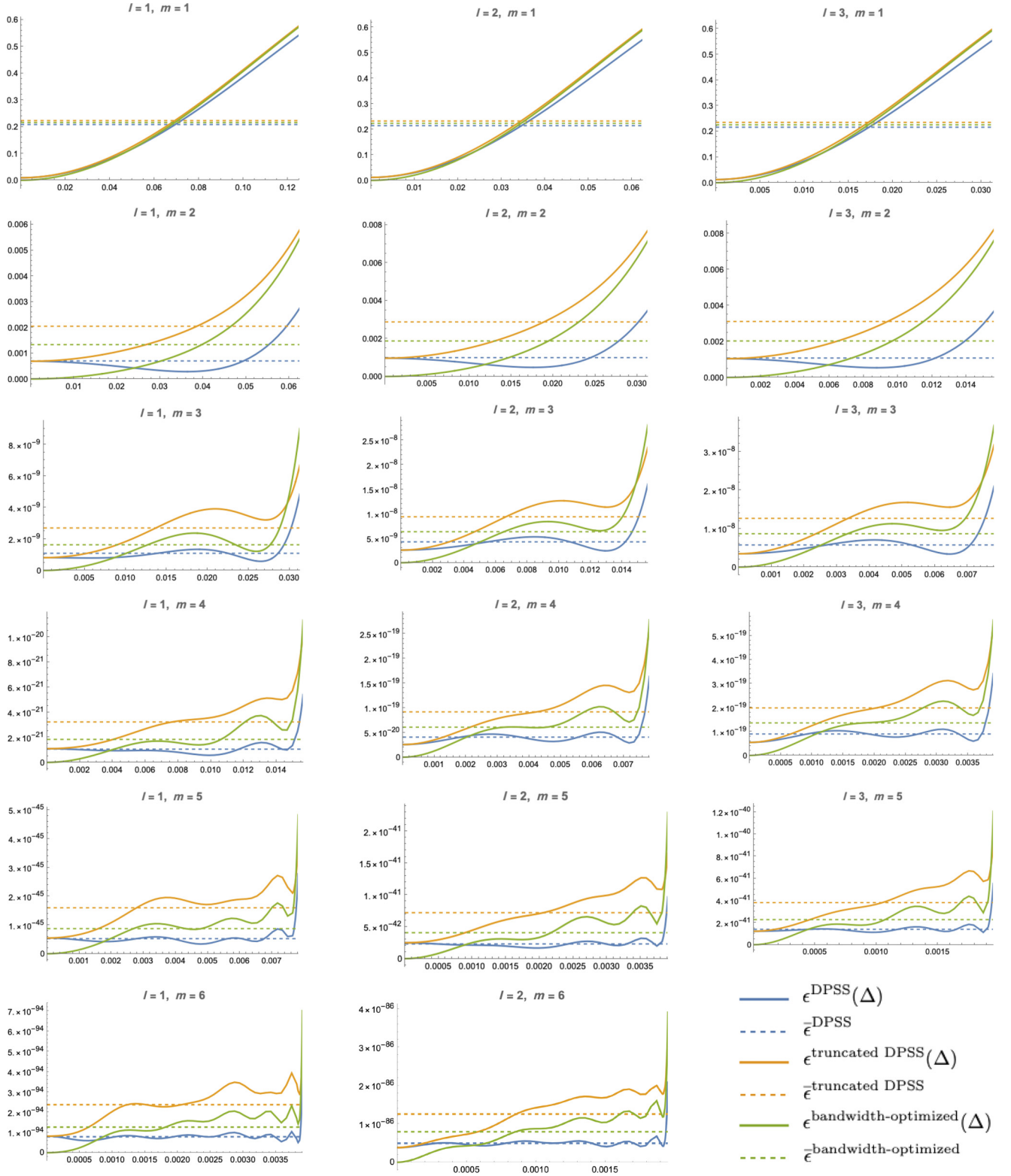


FIG. 5. Error probability of the various tapers for $\ell = 1, \dots, 3$ and $m = 1, \dots, 6$. The solid lines denote error probability as a function of Δ from 0 to $1/(2N)$, whereas the dashed lines of the same color are the corresponding averages in this interval. DPSS tapers are described in Sec. IV C, truncated DPSS tapers in Sec. VII A, and bandwidth-optimized tapers in Sec. VII B. We find that the average error of the bandwidth-optimized taper is always less than twice the error of the DPSS taper. Calculations were performed with Mathematica by setting `WorkingPrecision=250`.

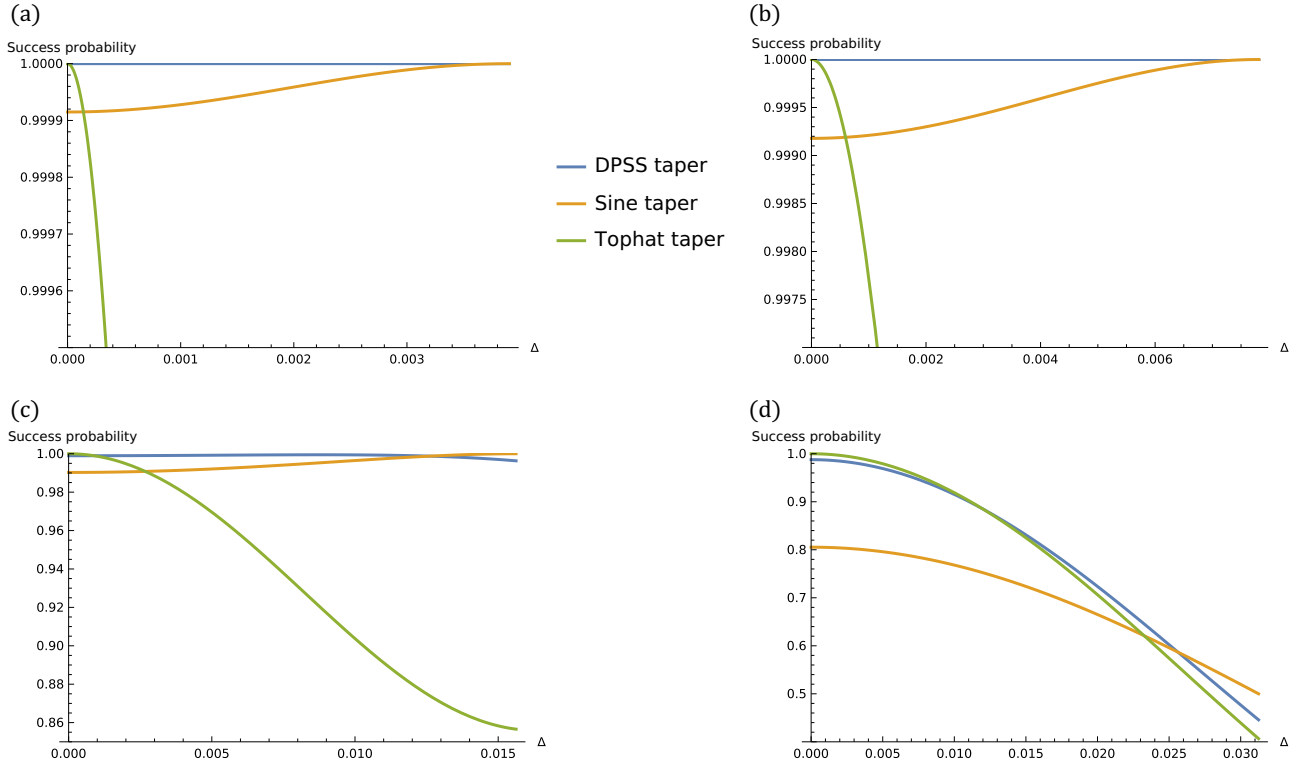


FIG. 6. Comparison between the DPSS taper, the sine taper, and the tophat taper. (a) $m = 4, K = 7$. (b) $m = 3, K = 3$. (c) $m = 2, K = 1$. (d) $m = 1, K = 0$.

does very well almost uniformly (besides being optimal on average).

V. ERROR ANALYSIS FOR QPE WITH UNCOMPUTATION

In this section, we bound the error in applications of QPE that require uncomputation of the estimated phase. The analysis below applies to all coherent QPE algorithms. In applications that require coherent QPE, like HHL and quantum Metropolis sampling, QPE is used in order to approximately implement a unitary of the form

$$cT_{CA} := \sum_j T(\theta_j)_C \otimes |\psi_j\rangle\langle\psi_j|_A, \quad (68)$$

where A and C are two registers (see the left-hand side of Fig. 7) and $T(\theta)$ is a unitary operator parametrized by $\theta \in [0, 2\pi)$. Furthermore, $\{|\psi_j\rangle\}_j$ are the eigenvectors of some unitary U with the corresponding eigenvalues $\{e^{2\pi i\theta_j}\}_j$. We will assume that T is L_T -Lipschitz continuous for some $L_T > 0$,

$$\|T(\theta) - T(\theta')\| \leq L_T \min(|\theta - \theta'|, 2\pi - |\theta - \theta'|). \quad (69)$$

We approximately implement the unitary cT using the following algorithm (see the right-hand side of Fig. 7):

- (1) Add an ancillary register B and initialize it in the all-zeros state. That is, $|0\rangle^{\otimes p}$ for some $p \in \mathbb{N}$.
- (2) Apply a unitary Q (this corresponds to any QPE algorithm, including the tQPE algorithm and the median-based QPE algorithms) on the registers A and B where A holds the system state and B encodes information about the phase. The exact form of this encoding, and how it is later decoded, depends on the specific QPE procedure employed. The general action of Q can be written as

$$Q_{BA} := \sum_{j,x,y} f_{jxy} |x\rangle\langle y|_B \otimes |\psi_j\rangle\langle\psi_j|_A, \quad (70)$$

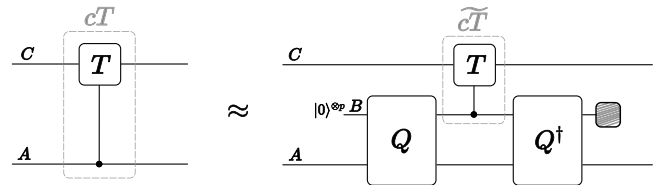


FIG. 7. The unitary cT (left) is approximated by the quantum channel (right). Note that Q can be any quantum phase estimation algorithm. For tapered quantum phase estimation, it would include the state preparation unitary for the taper on the ancilla register.

where f_{jxy} are the complex numbers determined by the particular QPE algorithm. These amplitudes satisfy, for some $\delta, \epsilon > 0$,

$$\sum_{x: |\theta_j - g(x)| > \delta} |f_{jx0}|^2 \leq \epsilon, \quad (71)$$

where $g(x) : \{0, 1\}^p \rightarrow [0, 1)$ can be seen as a phase-decoding function associated with the chosen QPE scheme. This function maps computational-basis outcomes x to a numerical phase estimate. The formulation above is general enough to capture all coherent QPE algorithms. For example:

- (a) Standard QPE or tQPE: The register B is initialized in $|0\rangle^{\otimes p}$, and the estimate is simply the binary fraction encoded in x , i.e.,

$$g(x) = \frac{x}{2^p}. \quad (72)$$

Here and in what follows, x refers both to the binary string in the computational basis as well as the associated integer it represents in binary. The meaning should be clear from context.

- (b) Median-based QPE: If x represents a tuple of outcomes $(x^{(1)}, \dots, x^{(m)})$ from m repeated QPE runs each with l bits, then the estimated phase is given by

$$g(x) = \text{median} \left(\frac{x^{(1)}}{2^l}, \dots, \frac{x^{(m)}}{2^l} \right). \quad (73)$$

Here, the median is taken coherently to suppress outliers among repeated noisy estimates, providing a more robust phase estimate.

- (3) Apply the following controlled unitary on registers B and C with the control on register B :

$$\tilde{c}T_{CB} := \sum_x T(g(x))_C \otimes |x\rangle\langle x|_B. \quad (74)$$

- (4) Apply $Q^\dagger$ on the registers A and B .
(5) Trace out register B .

Note that although cT and $\tilde{c}T$ act in a similar way on the C register, the former controls that action on the A register and the latter on the B register. Since the control on the B register is with respect to the computational basis, it is straightforward to implement.

Now, before proceeding with the error analysis of the above algorithm (Theorem 4), we define the quantum

channel versions of the unitaries cT , Q , and $\tilde{c}T$,

$$\mathcal{T}(\cdot) := cT(\cdot)cT^\dagger, \quad (75)$$

$$\mathcal{Q}(\cdot) := Q(\cdot)Q^\dagger, \quad (76)$$

$$\tilde{\mathcal{T}}(\cdot) := \tilde{c}T(\cdot)\tilde{c}T^\dagger. \quad (77)$$

We also define an appending channel $\mathcal{A}$, which essentially appends the all-zeros state to the input state, to represent Step 1 of the above algorithm,

$$\mathcal{A}_D(\rho) := \rho_D \otimes |0\rangle\langle 0|_B. \quad (78)$$

Note that in the above equation and in what follows, we use a shorthand $|0\rangle\langle 0|_B \equiv |0\rangle\langle 0|_B^{\otimes p}$ for simplicity.

With the above notions in place, we now present our QPE uncomputation error analysis in the following theorem.

Theorem 4. The following holds:

$$\frac{1}{2} \left\| \mathcal{T}_{CA} - \text{Tr}_B \circ \mathcal{Q}_{BA}^\dagger \circ \tilde{\mathcal{T}}_{CB} \circ \mathcal{Q}_{BA} \circ \mathcal{A}_{CA} \right\|_\diamond \leq 2L_T\delta + 4\epsilon. \quad (79)$$

Here, Tr_B denotes the partial-trace channel, where the register B is traced out.

The proof of the theorem can be found in Appendix G.

Equation (79) quantifies the error resulting from the uncomputation of QPE combined with the error of approximating the phase in the first place. Let us compare this error to the error introduced solely due the approximation of the phase. For this purpose, we consider the unitary

$$\overline{c}T := \sum_j T(\bar{\theta}_j)_C \otimes |\psi_j\rangle\langle \psi_j|_A, \quad (80)$$

where $\bar{\theta}_j$ is a single estimate with $|\bar{\theta}_j - \theta_j| \leq \delta$. Let us denote the associated channel by $\overline{\mathcal{T}}$. To implement this unitary one would need access to a QPE algorithm that always outputs a single estimate, which is known not to exist (Here we are ignoring the failure probability, ϵ , which we assume is negligible compared to $L_T\delta$). Nevertheless, in such a case uncomputation would be exact and the only error would be due to the approximation of the eigenphases. In

this case

$$\begin{aligned}
\frac{1}{2} \|T - \bar{T}\|_{\diamond} &\leq \|cT - c\bar{T}\| \\
&= \left\| \sum_j T(\theta_j)_C \otimes |\psi_j\rangle\langle\psi_j|_A - \sum_j T(\bar{\theta}_j)_C \otimes |\psi_j\rangle\langle\psi_j|_A \right\| \\
&= \left\| \sum_j (T(\theta_j)_C - T(\bar{\theta}_j)_C) \otimes |\psi_j\rangle\langle\psi_j|_A \right\| \\
&= \max_j \|T(\theta_j)_C - T(\bar{\theta}_j)_C\| \\
&\leq L_T \delta.
\end{aligned} \tag{81}$$

The first inequality follows from Lemma H2, the third equality follows from Lemma H1, and the last inequality follows from the L_T -Lipschitz property of the operator T . All in all, from Eq. (79) and Eq. (81), we can say that the additional error of QPE due to the uncomputing is the same order as the error due to the approximation of the phase.

VI. RANDOMIZED ALGORITHM

In Sec. IV E, we have seen that the success probability of the tQPE algorithm depends on the true unknown phase. To eliminate this dependence and achieve minimum error, we utilize a technique presented in Ref. [22], the details of which are reproduced below in the next paragraph for completeness. Although the authors of Ref. [22] used randomness to achieve unbiased phase estimation, we utilize their technique to reduce all instances Δ , including the worst-case instance, to the average-case instance, guaranteeing the optimality of our approach. We also note that randomization has been used in other works related to QFT and QPE [30,48].

In the randomized approach, we introduce a random global phase shift u before running QPE by replacing the target unitary U with $e^{2\pi i u} U$. QPE is then performed on this modified unitary to obtain an estimate $\tilde{\theta}$, after which the algorithm outputs $\tilde{\theta} - u$. This simple adjustment effectively randomizes the relative alignment between the true phase and the QPE sampling grid, turning any fixed instance into an average-case one while keeping the underlying phase unchanged.

Let Q be a QPE algorithm such that when acting on a phase θ , its error probability is $\epsilon(\theta, Q)$. In the randomized algorithm, we sample u uniformly at random from the range $[-1/(2N), 1/(2N))$ and pass it to the algorithm Q . This algorithm then returns as estimate $\tilde{\theta}$ satisfying $|\tilde{\theta} - (\theta + u)| \leq \delta$ with probability at least $1 - \epsilon(\theta + u, Q)$, by assumption. The randomized algorithm outputs $\tilde{\theta}_{\text{rand}} :=$

$\tilde{\theta} - u$, which satisfies

$$|\tilde{\theta}_{\text{rand}} - \theta| = |\tilde{\theta} - (\theta + u)| \tag{82}$$

$$\leq \delta, \tag{83}$$

with success probability $1 - \epsilon(\theta + u, Q)$. Therefore, the probability that the estimate $\tilde{\theta}_{\text{rand}}$ is δ -close to θ is then given by

$$\begin{aligned}
\text{Prob}(|\tilde{\theta}_{\text{rand}} - \theta| \leq \delta) &= \mathbb{E}_{u \sim \mathcal{D}} [(1 - \epsilon(\theta + u, Q))] \\
&= 1 - \mathbb{E}_{u \sim \mathcal{D}} [\epsilon(\theta + u, Q)]
\end{aligned} \tag{84}$$

$$= 1 - \bar{\epsilon}(Q) \tag{85}$$

$$= 1 - \bar{\epsilon}(Q), \tag{86}$$

where $\bar{\epsilon}(Q)$ is the average error probability of the QPE algorithm Q .

This procedure acts as a random self-reduction, effectively reducing any specific instance to an average instance. As a result, the performance of any randomized QPE algorithm is independent of the true phase θ , and the error probability can be upper-bounded by the average error probability of the corresponding pre-randomized QPE algorithm. Since we have already argued that the DPSS taper has the optimal average error probability, the randomized tQPE algorithm with the DPSS taper is provably optimal in the worst case as well.

The random phase shift can be implemented in tQPE with minimal additional gate cost and no increase in the number of qubits. We only need to modify the middle part of the tQPE algorithm, given in Eq. (8), as

$$\sum_{n=0}^{N-1} e^{2\pi i n u} |n\rangle\langle n| \otimes U^n. \tag{87}$$

In practice, this effect can be achieved by adding a sequence of single-qubit rotations acting on the ancilla qubits since

$$\bigotimes_{s=0}^{p-1} \left(e^{2\pi i 2^s u Z} \right)_{a_s} = e^{-2\pi i (N/2) u} \sum_{n=0}^{N-1} e^{2\pi i n u} |n\rangle\langle n|. \tag{88}$$

Here we adopt the convention that the ancilla qubits are labeled as $0, 1, \dots, p-1$, starting from the bottom ancilla qubit in Fig. 1. Recall that the s th ancilla qubit controls the application of U^{2^s} in the deterministic algorithm. The subscript a_s indicates the ancilla register the operator is acting on. The global phase outside the sum on the right-hand side is physically irrelevant.

VII. TAPER PREPARATION AND TOTAL ERROR OF THE ALGORITHM

The important issue of how to efficiently prepare a good taper state is addressed in detail in Appendix A and an

example worked out in Appendix A 1. Here we explain the ideas behind that method and use the results to establish the total error of the tQPE algorithm.

A. Truncated DPSS taper

In Sec. III, we have seen that the tQPE algorithm implements the following transformation on the taper states:

$$|\phi\rangle |\psi_\theta\rangle \xrightarrow{\text{tQPE}} |\hat{\phi}_\theta\rangle |\psi_\theta\rangle, \quad (89)$$

where

$$|\hat{\phi}_\theta\rangle := \sum_{k=0}^{N-1} \hat{\phi}(\theta - k/N) |k\rangle. \quad (90)$$

In the special case when $\theta = 0$, the outcome state

$$|\hat{\phi}\rangle := \sum_{k=0}^{N-1} \hat{\phi}(-k/N) |k\rangle \quad (91)$$

is the discrete-time Fourier transform of the taper. Since the DPSS taper has only $2K + 1$ significant frequency components and the rest are extremely suppressed, it can be safely truncated to these few dominant basis states while introducing a negligible error.

To see this, let P_θ be the projector into the subspace spanned by the $2K + 1$ computational-basis states closest to θ , which for $\theta = 0$ corresponds to the central lobe of the DPSS taper. With this, we define a truncation of the DPSS taper as

$$|\hat{\phi}^*\rangle = \frac{P_0 |\hat{\phi}\rangle}{\|P_0 |\hat{\phi}\rangle\|_2}. \quad (92)$$

By the properties of the DPSS taper we stated above, it holds that $|\langle \hat{\phi} | \hat{\phi}^* \rangle|^2 = 1 - \epsilon(0)$, which implies

$$|\hat{\phi}^*\rangle = \sqrt{1 - \epsilon(0)} |\hat{\phi}\rangle + \sqrt{\epsilon(0)} |\hat{\phi}^\perp\rangle, \quad (93)$$

where $|\hat{\phi}^\perp\rangle$ is orthogonal to $|\hat{\phi}\rangle$. Note that $\epsilon(\theta)$ had been defined in Eq. (15). Since QFT is unitary, the same holds for the time-domain tapers, i.e.,

$$|\phi^*\rangle = \sqrt{1 - \epsilon(0)} |\phi\rangle + \sqrt{\epsilon(0)} |\phi^\perp\rangle. \quad (94)$$

The idea is then to first prepare $|\hat{\phi}^*\rangle$ using brute-force quantum-state-synthesis approaches such as the one presented in Ref. [62] and then apply QFT to recover the approximate taper in the time domain. In Appendix A, we

show that the gate complexity for this state preparation algorithm is

$$\mathcal{C}_\phi = O(\log^2(1/\bar{\epsilon}) + \log^2(1/\delta)). \quad (95)$$

When using the approximate taper $|\phi^*\rangle$, the total error probability of the tQPE algorithm gets contributions from two sources: (i) the error probability when the exact DPSS taper is used and (ii) the additional error incurred due to the deviation of $|\phi^*\rangle$ from the exact DPSS taper. The corresponding ancilla state at the end of the tQPE algorithm

$$|\phi^*\rangle |\psi_\theta\rangle \xrightarrow{\text{tQPE}} |\hat{\phi}_\theta^*\rangle |\psi_\theta\rangle \quad (96)$$

is

$$|\hat{\phi}_\theta^*\rangle = \sqrt{1 - \epsilon(0)} |\hat{\phi}_\theta\rangle + \sqrt{\epsilon(0)} |\hat{\phi}_\theta^\perp\rangle. \quad (97)$$

The success probability when using the approximate DPSS taper equation (92) is then lower-bounded as

$$\begin{aligned} & \langle \hat{\phi}_\theta^* | P_\theta | \hat{\phi}_\theta^* \rangle \\ &= (1 - \epsilon(0)) \langle \hat{\phi}_\theta | P_\theta | \hat{\phi}_\theta \rangle + \epsilon(0) \langle \hat{\phi}_\theta^\perp | P_\theta | \hat{\phi}_\theta^\perp \rangle \\ & \quad + 2\sqrt{\epsilon(0)(1 - \epsilon(0))} \Re \langle \hat{\phi}_\theta^\perp | P_\theta | \hat{\phi}_\theta \rangle \end{aligned} \quad (98)$$

$$\begin{aligned} & \geq (1 - \epsilon(0))(1 - \epsilon(\theta)) \\ & \quad - 2\sqrt{\epsilon(0)(1 - \epsilon(0))} |\langle \hat{\phi}_\theta^\perp | P_\theta | \hat{\phi}_\theta \rangle|, \end{aligned} \quad (99)$$

where to get the inequality, we used the properties of the DPSS taper for the first term, set the second term to 0, and replaced the last term with the negative of its absolute value.

Next, we will show that $|\langle \hat{\phi}_\theta^\perp | P_\theta | \hat{\phi}_\theta \rangle|$ is small. Let $P_\theta^\perp = I - P_\theta$. Since $0 = \langle \hat{\phi}_\theta^\perp | \hat{\phi}_\theta \rangle = \langle \hat{\phi}_\theta^\perp | (P_\theta + P_\theta^\perp) | \hat{\phi}_\theta \rangle$ it follows that

$$|\langle \hat{\phi}_\theta^\perp | P_\theta | \hat{\phi}_\theta \rangle| = |\langle \hat{\phi}_\theta^\perp | P_\theta^\perp | \hat{\phi}_\theta \rangle|. \quad (100)$$

Using Cauchy-Schwarz inequality, the left-hand side is bounded from above by

$$\|P_\theta | \hat{\phi}_\theta \rangle\| \|P_\theta | \hat{\phi}_\theta^\perp \rangle\| = \sqrt{1 - \epsilon(\theta)} \sqrt{a}, \quad (101)$$

whereas the right-hand side is bounded from above by

$$\|P_\theta^\perp | \hat{\phi}_\theta \rangle\| \|P_\theta^\perp | \hat{\phi}_\theta^\perp \rangle\| = \sqrt{\epsilon(\theta)} \sqrt{1 - a}, \quad (102)$$

where

$$a := \|P_\theta | \hat{\phi}_\theta^\perp \rangle\|^2 \in [0, 1]. \quad (103)$$

Thus,

$$|\langle \hat{\phi}_\theta^\perp | P_\theta | \hat{\phi}_\theta \rangle| \leq \min\left(\sqrt{a(1-\epsilon(\theta))}, \sqrt{\epsilon(\theta)(1-a)}\right) \quad (104)$$

$$\leq \sqrt{\epsilon(\theta)(1-\epsilon(\theta))}, \quad (105)$$

where for the last inequality, we set $a = \epsilon(\theta)$, which maximizes the minimum for $a \in [0, 1]$. Substituting the above bound back in Eq. (99), we obtain

$$\begin{aligned} \langle \hat{\phi}_\theta^* | P_\theta | \hat{\phi}_\theta^* \rangle &\geq (1 - \epsilon(0))(1 - \epsilon(\theta)) \\ &\quad - 2\sqrt{\epsilon(0)(1 - \epsilon(0))}\sqrt{\epsilon(\theta)(1 - \epsilon(\theta))} \quad (106) \end{aligned}$$

$$\geq 1 - \left(\epsilon(0) + \epsilon(\theta) + 2\sqrt{\epsilon(0)\epsilon(\theta)}\right). \quad (107)$$

Finally, we consider the average success probability, which is

$$\mathbb{E}_{\theta \sim \mathcal{D}} \left[\langle \hat{\phi}_\theta^* | P_\theta | \hat{\phi}_\theta^* \rangle \right]. \quad (108)$$

Taking expected value on both sides of Eq. (107) and using Jensen's inequality $\mathbb{E}[\sqrt{x}] \leq \sqrt{\mathbb{E}[x]}$, we get

$$\begin{aligned} \mathbb{E}_{\theta \sim \mathcal{D}} \left[\langle \hat{\phi}_\theta^* | P_\theta | \hat{\phi}_\theta^* \rangle \right] &\geq 1 - \left(\epsilon(0) + \bar{\epsilon}^{\text{DPSS}} + 2\sqrt{\epsilon(0)\bar{\epsilon}^{\text{DPSS}}} \right) \quad (109) \\ &\geq 1 - 4 \max(\epsilon(0), \bar{\epsilon}^{\text{DPSS}}), \quad (110) \end{aligned}$$

where, as defined before $\bar{\epsilon}^{\text{DPSS}}$ is the DPSS average error probability. In most cases $\epsilon(0) \approx \bar{\epsilon}^{\text{DPSS}}$. Consequently, if we want to lower-bound the success probability by $1 - \epsilon$, choosing $\bar{\epsilon} = \bar{\epsilon}^{\text{DPSS}} = \epsilon/4$ in the theorems and lemmas in Sec. IV C is sufficient.

B. Optimized bandwidth-limited taper

The approximate taper above was obtained by first computing the optimal taper and then chopping its tails off in the frequency domain. This was necessary for the state preparation to be efficient. We can achieve a factor of two improvement by directly optimizing over the band-limited tapers that can be prepared efficiently. This is a strict improvement over the truncated DPSS taper, as the latter is also band-limited in the frequency domain. We start with the optimization problem in the average-case

error setting. The cost function is

$$p_{\text{success}} = \sum_{n,n'=0}^{N-1} \phi[n] C(n, n') \phi[n']^*, \quad (111)$$

$$C(n, n') = \frac{\sin(\pi(n - n')(2K + 1)/N)}{\pi(n - n')}. \quad (112)$$

Let us write the success probability in terms of the Fourier coefficients of the taper,

$$\hat{\phi}[k] := \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \phi[n] e^{-2\pi i n k / N}. \quad (113)$$

Note that here we are not referring to the discrete-time Fourier transform of Eq. (12). We use angular brackets to indicate that the argument is an integer index. Then

$$p_{\text{success}} = \sum_{k,k'=0}^{N-1} \hat{\phi}[k] D(k, k') \hat{\phi}[k']^*, \quad (114)$$

$$D[k, k'] = \frac{1}{N} \sum_{n,n'=0}^{N-1} C[n, n'] e^{2\pi i (nk - n'k')/N}. \quad (115)$$

In the previous approach we chopped off the tails of the taper in the frequency space, hence setting $\hat{\phi}[k] = 0$ for $K < |k| \leq N$. Thus the success probability is

$$p_{\text{success}}^{\text{truncated}} = \sum_{k,k'=-K}^K \hat{\phi}[k] D[k, k'] \hat{\phi}[k']^* e^{2\pi i (nk - n'k')/N}. \quad (116)$$

We can directly maximize this expression. There are some benefits to this approach. First, this success probability is larger than that of the truncated DPSS since it is a maximum over a set that does include it. Second, the size of the eigenvalue problem is of dimension $2K + 1 \approx 2^m$ as opposed to $N = 2^{\ell+m}$. The first downside is that the matrix elements of the new kernel $D[n, n']$ require a double sum over N items leading to N^2 complexity if done by brute force. Second, the new kernel $D[k, k']$ might not be as well studied as DPSS and it might be less efficient to compute classically.

In the numerics for $l + m \leq 10$ we found that the average success probability is larger than $1 - 2\bar{\epsilon}$ which can also be seen in Fig. 5. Consequently, if we want to lower-bound the success probability by $1 - \epsilon$ choosing $\bar{\epsilon} = \epsilon/2$ in the theorems and lemmas in Sec. IV C is likely sufficient.

C. Tapers extrapolated from small l^*

In Sec. VII A we implicitly assumed knowledge of the DPSS taper. In practice, a DPSS taper might be hard to compute classically for large values of p . Similarly,

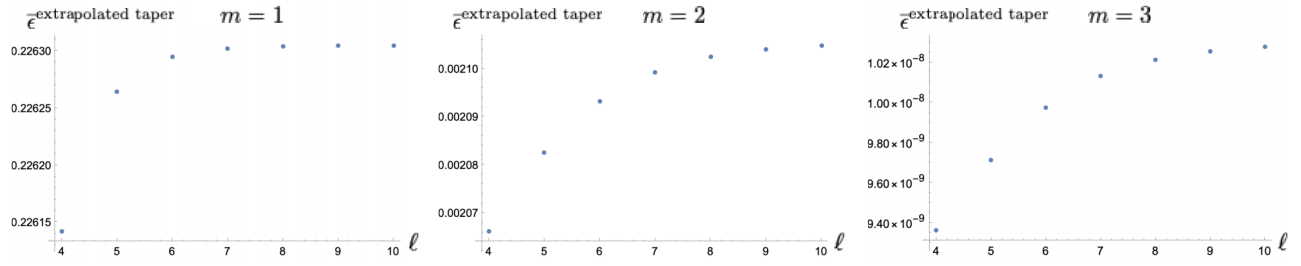


FIG. 8. Demonstration of the performance of the tapers from Sec. VIIC. Plotted are the average errors of tapers for $m = 1, 2, 3$ and $\ell = 4, \dots, 10$ that have been obtained with classical computations for the problem size with fixed $\ell^* = 4$. These tapers are optimal for only $\ell = 4$ by construction, but nonetheless perform quite well for larger ℓ .

the calculation of optimal bandwidth-limited tapers in Sec. VIIB might be expensive. Note that these are one-time classical calculations for each (ℓ, m) -pair. Nonetheless, in this section we present a method to classically efficiently generate approximate tapers that nevertheless perform very well.

The idea is to follow the prescription of Sec. VIIB (one could also follow Sec. VIIA with similar results) with some $\ell^* < \ell$ for which we can carry out the classical computation. The outcome of this calculation is a vector of dimension $2K + 1$. In those sections we turned this vector to an $N^* = 2^{\ell^*+m}$ dimensional one by appending zeros. Applying QFT then resulted in the optimal taper. Here, instead, we turn this vector to a $N = 2^{\ell+m}$ dimensional one and then apply QFT to get the new taper. This taper is suboptimal, since the $2K + 1$ dimensional vector was optimized for (ℓ^*, m) . Nevertheless, it performs very well as we can see in Fig. 8, even if $\ell \gg \ell^*$.

VIII. DISCUSSION

In this manuscript we studied the performance of the QPE algorithm as the input state of the ancilla register (called a taper here) is varied. We found that different objective criteria result in different optimal tapers. By combining QPE with phase randomization, we were able to show that the success probability of outputting an estimate within the desired error bars is maximized by the DPSS taper. The DPSS taper is optimal in all parameter regimes, not only asymptotically. Given the error probabilities shown in Fig. 4 it is likely that most applications will require $m \leq 4$ additional qubits beyond what is needed for any desired precision. We also showed how approximations to this taper with similar performance guarantees can be prepared efficiently on a quantum computer.

Quantum algorithms that require coherent phase estimation, such as the HHL algorithm [3] and Quantum Metropolis Sampling [21], also require that the estimate of the phase be uncomputed at some point. If this were not the case, incoherent phase estimation would be sufficient, as discarding the ancilla register after coherent QPE renders it equivalent to incoherent QPE. Because tQPE

outputs multiple estimates for a particular phase (unlike median boosted QPE that outputs the closest two estimates with high probability), one might worry that the uncomputation step might fail. In Sec. V we argued that this concern is unfounded: the error of any coherent QPE algorithm due to uncomputation is of the same magnitude as the error due to the approximation of the phase. For a precise statement of this result, see Theorem 4. Note that this theorem is proven in Appendix G for tQPE without randomization. The proof can be extended to the randomized method described in Sec. VI; however, the resulting error parameter ϵ in the upper bound remains a worst-case error rather than the average error $\bar{\epsilon}$ (a detailed proof is omitted). This distinction is not expected to have practical impact, as the upper bound will typically be dominated by δ in most applications.

We note that the authors of Ref. [63] studied continuous QPE and formulated a different optimization problem than ours. Interestingly, their optimal taper for continuous QPE also resulted in DPSS. So, it would be valuable to explore the similarities and differences between these two optimization problems, which we leave for future work.

We remark that the asymptotic scaling of the number of additional qubits for the DPSS taper compares favorably against the Kaiser taper which is known to also achieve a similar asymptotically scaling of m as our DPSS taper [47]. There are other additional factors inside the logarithm for the Kaiser taper which the DPSS taper does not have. Furthermore, for the case of the Kaiser taper, its K value scales as $O(\log(1/\epsilon)) + O(\log \log(1/\epsilon))$ with respect to ϵ in the asymptotic regime as shown in Appendix D of Ref. [47]. As far as we know, all qubit-count analysis for the number of qubits required by the Kaiser taper to boost the success probability for QPE has only been done using the asymptotic expression. On the other hand, our work provides a nonasymptotic minimum qubit count for the DPSS taper for boosting the success probability. For the asymptotic case, as shown in the proof for Theorem 3, the DPSS taper scales as $O(\log(1/\epsilon))$ in the asymptotic regime which implies that the DPSS taper has better frequency concentration around θ than the Kaiser taper in the regime where N is large and δ is small. This corresponds

well with existing classical signal processing literature where the DPSS taper is known to have maximal frequency concentration in the central lobe out of all taper functions and the Kaiser taper has originally been designed as an approximation to the optimal DPSS taper.

In addition to the asymptotic analysis, we also provide the explicit expressions for m in the nonasymptotic regime. This will be important for the practical implementation of the QPE algorithm in the early fault-tolerant regime. Finally, we would like to mention that we approach the QPE problem from a different angle, i.e., by framing optimization problems. Due to the optimality of the DPSS taper, all tapers must use $\Omega(\log \log(1/\epsilon))$ additional qubits to boost the average success probability to at least $1 - \epsilon$. This means that there cannot be any further improvements when using tapers for QPE. This result complements Thm. 1.3 of Ref. [14], which provides the same lower bound on the ϵ -dependence for any algorithm solving QPE in the worst case. However, while Thm. 1.3 of Ref. [14] is general and holds for any QPE algorithm, our lower bound only applies to tQPE with arbitrary tapers.

Due to the flexibility of our formulation, it allows us to find tapers satisfying different optimization criteria. For instance, we find that a sinusoidal taper minimizes the average-case error under the additional constraint that for $m = 0$ the error probability is zero whenever the true phase happens to be exactly in between two estimates. This taper has been obtained before in the literature as a result of a different optimization problem [33,64]: it is the taper that minimizes a cost function that penalizes estimates that are further away from the true value as $4 \sin^2((\theta - \theta_{\text{est}})/2)$. We have arrived at this taper from a completely different perspective.

Note added. During the completion of this manuscript, we became aware of a recent numerical study conducted by Greenaway *et al.*, concluding that QPE performed with the Kaiser taper significantly outperformed QSVT-QPE [16] (QPE performed with the QSVT framework) [65]. In particular, they observed that QPE with Kaiser tapers achieved success probabilities that are orders of magnitude higher than QSVT-QPE with better query complexity in the highly relevant regime where the desired success probability is very close to 1. Readers can find a more explicit comparison of the query complexity for the QSVT-QPE approach and the tQPE approach with the Kaiser taper in the figures in Ref. [65]. This finding further emphasizes the importance of utilizing tapers for the QPE problem.

ACKNOWLEDGMENTS

The authors thank Sam Sleazak for useful discussions related to the error analysis in Sec. V and anonymous reviewers for their suggestions that helped improve the presentation of this work. D.P. and S.J.S.T. were participants in the 2023 Quantum Computing Summer School

at LANL, sponsored by the LANL Information Science & Technology Institute. D.P. spent the following fall semester as a GRA at LANL during which he was supported by the Laboratory Directed Research and Development program of Los Alamos National Laboratory (LANL) under Project No. 20230049DR. This material is based upon work supported by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Quantum Science Center (QSC). Y.S. was funded by the QSC to perform the analytical and numerical analysis and to write the manuscript along with the other authors. A.T.S. acknowledges initial support from the LANL ASC Beyond Moore's Law project and subsequent support from the Laboratory Directed Research and Development program of Los Alamos National Laboratory (LANL) under project number 20230049DR.

DATA AVAILABILITY

No data were created or analyzed in this study.

APPENDIX A: APPROXIMATE TAPER PREPARATION

In this Appendix, we focus on preparing the tapers described in Sec. VII. We use a modification of the method for preparing broadband Gaussian states presented in Ref. [8] to prepare the taper state. The key idea is that the DPSS taper in the frequency domain, i.e., $|\hat{\phi}\rangle = U_{\text{QFT}}^\dagger |\phi\rangle$, is narrow band. By narrow band, we mean that it is highly concentrated on just a few grid points, implying that the state, $|\hat{\phi}\rangle$, has appreciable amplitudes on very few of its basis states. Therefore, we can prepare an approximate version of $|\hat{\phi}\rangle$ (say $|\hat{\phi}^*\rangle$) using standard methods presented in Ref. [8] with very little overhead. After that, we can perform a quantum Fourier transform on $|\hat{\phi}^*\rangle$ to obtain $|\phi^*\rangle$, which closely approximates the DPSS taper $|\phi\rangle$.

Lemma A1. The gate complexity of preparing the approximate tapers of Secs. VII A–VII C is

$$\mathcal{C}_\phi = O(\log^2(1/\bar{\epsilon}) + \log^2(1/\delta)). \quad (\text{A1})$$

Proof. We provide an explicit description of the taper state preparation method. The tapers of Secs. VII A–VII C all have in common the fact that in the frequency domain they have only $N' = 2K + 1$ nonzero entries. Since the complexity of state preparation only depends on this parameter, the analysis below applies to all these tapers. Using the results of Theorem 2, it is sufficient to choose $m = \lceil \log_2(x) \rceil + 1$ with $x = \lceil 175(\log(10/\bar{\epsilon}) + 1)^2 \rceil + 1$.

We can upper-bound $N' = 2K + 1 = 2^m - 1$ as

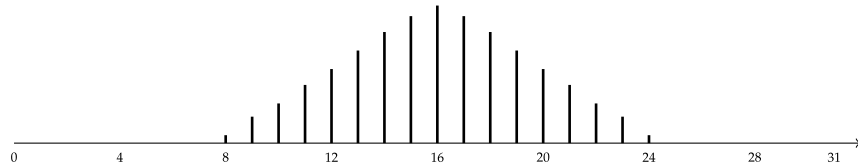
$$N' = 2^{\lceil \log_2(x) \rceil + 1} - 1 \quad (\text{A2})$$

$$\leq 4x. \quad (\text{A3})$$

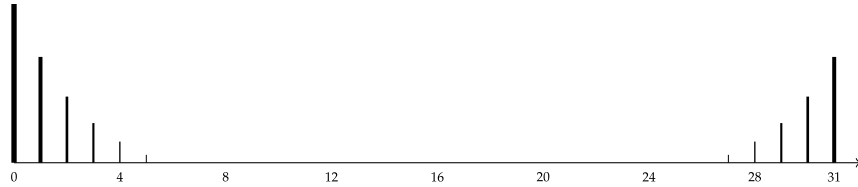
Hence $N' = O(\log^2(1/\epsilon))$. In the frequency domain the taper is centered around 0, and is band-limited. We use the $N' = 2^m - 1$ amplitudes in the central peak to construct a quantum state on a register of $\log(N' + 1) = m$ qubits. This state is centered at the computational-basis state 0 and symmetric around it modulo m . The amplitude at basis state 2^{m-1} is set to zero. This state is then prepared using the method described in Ref. [62]. Subsequently, we append ℓ qubits initialized to $|0\rangle^{\otimes \ell}$ to the ancilla register to form an $(\ell + m)$ -qubit ancilla register, in such a way that the added qubits represent the most significant digits of the computational basis. Now the taper is not centered around 0 anymore; all the nonzero amplitudes are on the computational-basis states in the interval $[0, 2^m - 1]$. To make the taper centered again it suffices to act with the following unitary:

$$\prod_{n=m+1}^p \text{CNOT}_{m \rightarrow n}. \quad (\text{A4})$$

This unitary leaves the computational-basis states in the interval $[0, 2^{m-1} - 1]$ unchanged and adds $2^p - 2^m$ to the basis states in the interval $[2^{m-1}, 2^m - 1]$. This is exactly the transformation needed to center the taper around 0. Finally, we act with U_{QFT} to obtain the taper in the time-domain which is centered around $N/2$ and $N/2 + 1$. This is the approximate taper.



Next, we transform the time-domain values to the frequency domain,



Now, we proceed with the truncation step where we choose to only keep the largest $2k + 1 = 2^{m-1} = 7$ values centered around 0. These are the frequencies that will be associated to the phase values that we hope to measure with high probability,

Now, we analyze the complexity of the taper state preparation protocol. Shende *et al.* [62] introduced a technique for preparing a state on a register of m qubits with a gate complexity of $O(2^m)$. Note that their technique requires some classical computation, which we disregard, along with the classical computation of the $2^m - 1$ amplitudes of the central lobe of $|\hat{\phi}\rangle$. The quantum circuit required to center the amplitudes consists of ℓ CNOT gates. Last, we require an application of QFT which has been shown in Ref. [66] to have gate complexity $O(\log^2 N) = O(p^2) = O((\ell + m)^2)$. This gives us an overall gate complexity of $O(2^m + \ell^2)$ for the state preparation protocol. We can state the gate complexity of the state preparation protocol $\mathcal{C}_\phi$ in terms of δ and ϵ as

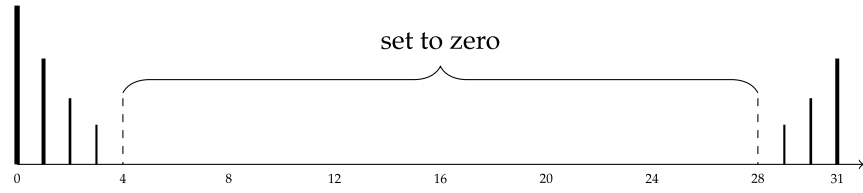
$$\mathcal{C}_\phi = O(\log^2(1/\epsilon) + \log^2(1/\delta)). \quad (\text{A5})$$

■

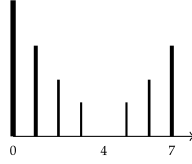
We note that the state preparation protocol does not require any additional ancilla qubits other than the p -qubits for the taper state register.

1. Taper preparation circuit example

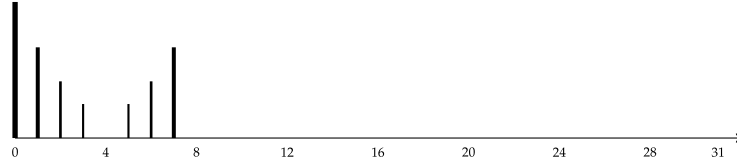
In this Appendix, we provide a concrete example of how our DPSS taper can be prepared using the truncation method described in Appendix A. In this example, we assume that $\ell = 2$ and $m = 3$. Recall that the truncation method requires us to start with the classical computation of the DPSS taper in the time domain. In the figure below, we illustrate the time-domain values of the DPSS taper,



Once we have truncated away the unwanted frequencies, we proceed to the part where the task would be performed on the quantum computer. We begin with m qubits and prepare the following state by using standard approaches for state synthesis:



Note that we are only using 3 qubits to encode the amplitudes of the 7 frequencies that we chose to keep. We now add 2 qubits in the $|0\rangle^{\otimes 2}$ state while adhering with the convention that the added qubits correspond to the two most significant bits in the computational state basis. The new state is then given by



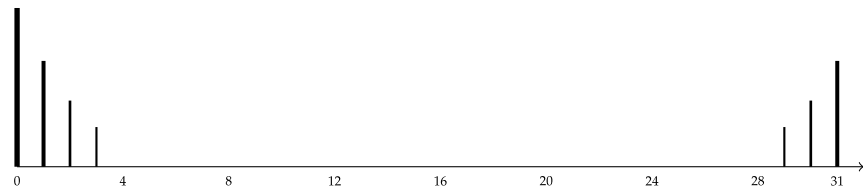
Next, we apply the following unitary that comprises of several CNOTs:

$$\prod_{n=m+1}^p \text{CNOT}_{m,n} = \text{CNOT}_{3,5}, \text{CNOT}_{3,4}. \quad (\text{A6})$$

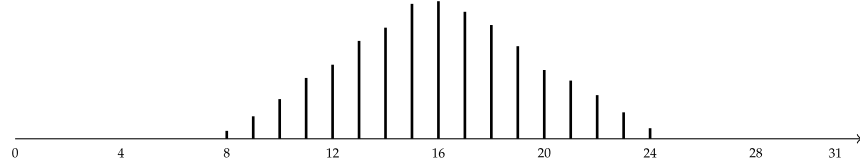
The effect of this set of gates on computational-basis states is as follows:

$$\begin{aligned} |0\rangle &= |00000\rangle \rightarrow |11100\rangle = |0\rangle, \\ |1\rangle &= |00001\rangle \rightarrow |11100\rangle = |1\rangle, \\ |2\rangle &= |00010\rangle \rightarrow |11100\rangle = |2\rangle, \\ |3\rangle &= |00011\rangle \rightarrow |11100\rangle = |3\rangle, \\ |4\rangle &= |00100\rangle \rightarrow |11100\rangle = |28\rangle, \\ |5\rangle &= |00101\rangle \rightarrow |11101\rangle = |29\rangle, \\ |6\rangle &= |00110\rangle \rightarrow |11110\rangle = |30\rangle, \\ |7\rangle &= |00111\rangle \rightarrow |11111\rangle = |31\rangle. \end{aligned} \quad (\text{A7})$$

Note that the first four states remained unchanged under the CNOT gates. This is essential because we want to keep those four amplitudes on the left. In general, this unitary shifts the right half of the taper by an amount $2^p - 2^m$, thus leaving it all the way to the right (that is, $|2^p - 1\rangle$). In other words, we have reproduced the original truncated frequency profile on a quantum register with p qubits,



Finally, we apply QFT to our p -qubit register to transform back into the time domain to get the following:



Note that the taper we have obtained is extremely close to the DPSS taper as proven in Appendix A.

APPENDIX B: IDEAL CASE OPTIMAL TAPERS

The optimal taper for tQPE is the one that maximizes the probability of outputting the value of a phase estimate of the form k/N that is δ -close to θ . To simplify the analysis we have set $\delta = 2^{-(\ell+1)}$ for some integer ℓ . In this case the $2K + 1$ discrete frequencies closest to θ are indeed δ -close to the true phase for $K = 2^{m-1}$, where $p = \ell + m$ is the total number of ancilla qubits. At most a single discrete frequency that is δ -close is left out, and that frequency has the smallest probability of being output by the algorithm. Thus without introducing much error we formulate the optimization problem in terms of maximizing the probability of outputting the closest $2K + 1$ discrete estimates,

see Eq. (18). Since this is a constrained problem, we used the Lagrangian formulation to solve it. For ease of reference, we restate the corresponding Lagrangian Eq. (19) from the main text below:

$$\mathcal{L}(|\phi\rangle, \lambda) = \sum_{j=-K}^K \left| \hat{\phi} \left(\Delta + \frac{j}{N} \right) \right|^2 - \lambda \left(\sum_{n=0}^{N-1} |\phi[n]|^2 - 1 \right). \quad (\text{B1})$$

To find the optimal taper for the original optimization problem, given by Eq. (18), we need to find the stationary point of $\mathcal{L}$ that maximizes the objective function of Eq. (18). But before doing so, we expand the expression on the right-hand side of the above equation by plugging the definition of $\hat{\phi}$, given by Eq. (12),

$$\mathcal{L}(\phi, \lambda) = \frac{1}{N} \sum_{j=-K}^K \sum_{n,n'=0}^{N-1} \phi[n] \phi^*[n'] e^{2\pi i \left(\Delta + \frac{j}{N} \right) (n-n')} - \lambda \left(\sum_{n=0}^{N-1} \phi[n] \phi^*[n] - 1 \right) \quad (\text{B2})$$

$$= \frac{1}{N} \sum_{j=-K}^K \sum_{n,n'=0}^{N-1} e^{2\pi i \Delta (n-n')} \phi[n] \phi^*[n'] e^{-2\pi i (j/N)(n'-n)} - \lambda \left(\sum_{n=0}^{N-1} \phi[n] \phi^*[n] - 1 \right). \quad (\text{B3})$$

Then, to find the stationary points of $\mathcal{L}$, we set all partial derivatives to zero. First, we differentiate $\mathcal{L}$ with respect to $\phi^*[m]$ for all $m \in \{0, \dots, N-1\}$ then set the expression to zero,

$$\frac{\partial \mathcal{L}}{\partial \phi^*[m]} = \frac{1}{N} \sum_{j=-K}^K \sum_{n=0}^{N-1} e^{2\pi i \Delta (m-n)} \phi[n] e^{-2\pi i (j/N)(m-n)} - \lambda \phi[m] = 0 \quad (\text{B4})$$

$$\Rightarrow \frac{1}{N} \sum_{j=-K}^K \sum_{n=0}^{N-1} e^{2\pi i \Delta (n-m)} \phi[n] e^{-2\pi i (j/N)(m-n)} = \lambda \phi[m] \quad (\text{B5})$$

$$\Rightarrow \frac{1}{N} \sum_{n=0}^{N-1} e^{2\pi i \Delta (n-m)} \left(\sum_{j=-K}^K e^{-2\pi i (j/N)(m-n)} \right) \phi[n] = \lambda \phi[m] \quad (\text{B6})$$

$$\Rightarrow \frac{1}{N} \sum_{n=0}^{N-1} e^{2\pi i \Delta (n-m)} \left(\frac{\sin(\pi(m-n)(2K+1)/N)}{\sin(\pi(m-n)/N)} \right) \phi[n] = \lambda \phi[m]. \quad (\text{B7})$$

As is evident, the last equation is an eigenvalue equation with eigenvalue λ and eigenvector $|\phi\rangle$. We rewrite this equation more concisely as follows:

$$\mathcal{B}_\Delta |\phi\rangle = \lambda |\phi\rangle, \quad (\text{B8})$$

where we define the matrix $\mathcal{B}_\Delta$ as

$$\mathcal{B}_\Delta[m, n] := \frac{1}{N} e^{2\pi i \Delta(m-n)} \left(\frac{\sin(\pi(m-n)(2K+1)/N)}{\sin(\pi(m-n)/N)} \right). \quad (\text{B9})$$

Next, we differentiate $\mathcal{L}$ with respect to the Lagrange multiplier, λ , and also set it to zero,

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \sum_{n=0}^{N-1} \phi[n] \phi^*[n] - 1 = 0 \quad (\text{B10})$$

$$\implies \sum_{n=0}^{N-1} \phi[n] \phi^*[n] = 1, \quad (\text{B11})$$

reproducing the normalization constraint.

Now, by plugging all the stationary points $(|\phi\rangle, \lambda)$ that satisfy the conditions Eqs. (B7) and (B11) into the objective function of the original optimization problem Eq. (18), we find

$$\sum_{j=-K}^K \left| \hat{\phi} \left(\Delta + \frac{j}{N} \right) \right|^2 = \lambda. \quad (\text{B12})$$

In other words, this means that the stationary point $(|\phi\rangle, \lambda)$ that maximizes the objective function is the eigenvector of $\mathcal{B}_\Delta$ with the maximum eigenvalue λ , and this value corresponds to the success probability of the algorithm.

APPENDIX C: RELATIONSHIP BETWEEN P-DPSS TAPERS AND IDEAL CASE OPTIMAL TAPERS

We compare the operator $\mathcal{B}_\Delta$ defined in Eq. (B9) to the band-limiting operator $\mathcal{B}_K : \mathbb{C}^N \rightarrow \mathbb{C}^N$ described in Zhu *et al.* [60]. Suppose $K \in \mathbb{N}$ and $2K+1 < N$. The band-limiting operator, $\mathcal{B}_K$, is defined such that it zeros out the discrete frequencies of a signal $\phi \in \mathbb{C}^N$ that lie outside of the range $\{-K \equiv N-K \pmod{N}, \dots, 0, \dots, K\}$. Formally, this operator does the following:

$$(\mathcal{B}_K(\phi))[m] := \frac{1}{\sqrt{N}} \sum_{n \in S} \hat{\phi}[n] e^{\frac{2\pi i m n}{N}}, \quad (\text{C1})$$

where $m \in \{0, \dots, N-1\}$ and $S = \{0, 1, \dots, K\} \cup \{N-K, \dots, N-1\}$. In the time domain this can be written as

$$(\mathcal{B}_K(\phi))[m] = \frac{1}{N} \sum_{n=0}^{N-1} \frac{\sin(\pi(m-n)(2K+1)/N)}{\sin(\pi(m-n)/N)} \phi[n]. \quad (\text{C2})$$

Note that this expression only differs from the matrix description stated in Eq. (B7) by a factor of $e^{2\pi i \Delta(m-n)}$ in each summand. We now introduce the time-limiting operator $\mathcal{T}_N : \mathbb{C}^N \rightarrow \mathbb{C}$ that is defined similarly as the following:

$$(\mathcal{T}_N(\phi))[n] := \begin{cases} \phi[n], & n \in \{0, 1, \dots, N-1\}, \\ 0, & \text{else.} \end{cases} \quad (\text{C3})$$

In Ref. [60], it was shown that the P-DPSSs are the eigensequences of the operator $\mathcal{T}_N \mathcal{B}_K \mathcal{T}_N$. In other words, the P-DPSSs are discrete sequences that contain a particular eigensequence that not only fulfills the time-limiting constraints of having nonzero values only on the indices $0, 1, \dots, N-1$, but also has maximal frequency concentration within our window of interest, $2K+1$. Notice that the time-limiting property and the property of having maximal frequency concentration around $2K+1$ lattice points are the same properties that we want the optimal tapers to possess. These similarities make it interesting for us to study the eigenvalue spectrum of the P-DPSSs because it may suggest an upper bound on what we can do with our ideal tapers. The bounds for the eigenvalues of the P-DPSS's with respect to the operator $\mathcal{T}_N \mathcal{B}_K \mathcal{T}_N$, proven in Ref. [60], are stated in the following theorem, which we restate with our notation.

Theorem C1 ([Theorem 1][60]). Suppose $K, N \in \mathbb{N}$ and $W = (2K+1)/2N < \frac{1}{2}$. Also, suppose that $\lambda^{(i)}$ is the i th largest eigenvalue of $\mathcal{T}_N \mathcal{B}_K \mathcal{T}_N$. Then for any $\epsilon \in (0, \frac{1}{2})$, we have

$$\lambda^{(2\lfloor NW \rfloor - \lfloor R(N, \epsilon) \rfloor)} \geq 1 - \epsilon, \quad (\text{C4})$$

where

$$R(N, \epsilon) = \left(\frac{4}{\pi^2} \log(8N) + 6 \right) \log \left(\frac{16}{\epsilon} \right) + 2 \max \left(\frac{-\log \left(\frac{\pi}{32} \left(\left(\frac{N}{N-1} \right)^2 - 1 \right) \epsilon \right)}{\log \left(\frac{N}{N-1} \right)}, 0 \right). \quad (\text{C5})$$

Although Theorem C1 is stated for the matrix that corresponds to $\mathcal{T}_N \mathcal{B}_K \mathcal{T}_N$, the authors of Ref. [60] showed that the eigenvalues of $\mathcal{T}_N \mathcal{B}_K \mathcal{T}_N$ and $\mathcal{B}_K$ are the same. Note that the matrix representation of the eigenvalue equation given by Eq. (B7) differs from the matrix representation of $\mathcal{B}_K$ by some complex phase. However, we demonstrate in the following proposition that they share the same spectrum by showing that we can adapt the eigensequences of $\mathcal{B}_K$ to obtain the eigensequences that satisfy Eq. (B7). In other words, we can use the analytical bounds for the eigenvalues of the P-DPSSs to bound the eigenvalues for

our ideal tapers. Below, we refer to ϕ_i^Δ , the eigenvector of the operator described in Eq. (B7) with eigenvalue $\lambda^{(i)}$, as the QP-DPSS, where $-(1/2N) \leq \Delta \leq 1/2N$.

Proposition C1. Suppose $\{(\lambda^{(i)}, \psi_i)\}_i$ represents the set of eigenvalues and eigensequences of $\mathcal{B}_K$. Then, for some $-(1/2N) \leq \Delta \leq (1/2N)$, $\{(\lambda^{(i)}, \phi_i^\Delta)\}_i$ is the set of eigenvalues and eigensequences of B_Δ given in Eq. (B9), where

$$\phi_i^\Delta[n] = e^{2\pi i \Delta n} \psi_i[n]. \quad (\text{C6})$$

Proof. We first restate the eigenvalue equation described in Eq. (B7) with the proposed eigensequence, $\phi_i^\Delta[n]$,

$$\begin{aligned} \frac{1}{N} \sum_n e^{2\pi i \Delta(m-n)} \left(\frac{\sin(\pi(m-n)(2K+1)/N)}{\sin(\pi(m-n)/N)} \right) \phi_i^\Delta[n] \\ = \frac{1}{N} \sum_n e^{2\pi i \Delta(m-n)} \left(\frac{\sin(\pi(m-n)(2K+1)/N)}{\sin(\pi(m-n)/N)} \right) \\ \times e^{2\pi i \Delta n} \psi_i[n] \end{aligned} \quad (\text{C7})$$

$$= \frac{e^{2\pi i \Delta m}}{N} \sum_n \left(\frac{\sin(\pi(m-n)(2K+1)/N)}{\sin(\pi(m-n)/N)} \right) \psi_i[n] \quad (\text{C8})$$

$$= e^{2\pi i \Delta m} \lambda^{(i)} \psi_i[m] \quad (\text{C9})$$

$$= \lambda^{(i)} \phi_i^\Delta[m], \quad (\text{C10})$$

where the second to last equality follows from the fact that ψ_i is an eigensequence of $\mathcal{B}_K$ with eigenvalue $\lambda^{(i)}$. This concludes the proof. ■

Note that the eigensequences $\{\phi_i^\Delta\}_i$ are what we refer to as QP-DPSSs in the main text. It is clear from the development above that transforming the tophat taper in the way shown in Eq. (C6) would give us a taper that would succeed with probability 1. In fact, it is known that the P-DPSSs have up to $2K+1$ eigensequences with eigenvalue 1. While such a transformation would be amazing, we remark that it requires explicit knowledge of Δ which

is not something that we can reasonably assume to possess (i.e., it would assume knowledge of the phase that we are trying to determine). Nevertheless, we provide some further discussion below to allow the reader to understand the deeper connections between P-DPSSs and optimal quantum tapers. In particular, we note that it may be possible to construct optimal tapers with greater efficiency by taking linear combinations of optimal tapers due to the possibility of the existence of a much more efficient description of the tapers. In fact, if one is willing to give up on the goal of achieving a success probability of 1, there are possibly other tapers with a much more efficient representation. From the development above in Theorem C1 and Proposition C1, we have the following result.

Corollary C1. Given $N = 2^{l+m}$, where l is the number of qubits needed to encode the target phase and m is the number of extra qubits necessary to guarantee that with probability at least $1 - \epsilon$, the desired precision will be met. Using the QP-DPSS taper with the largest eigenvalue, we have

$$m = O(\log \log 1/\delta + \log \log 1/\epsilon). \quad (\text{C11})$$

Proof. We are interested in the largest eigenvalue, i.e., $\lambda^{(0)}$. Therefore, from Eq. (C4), we first set $2\lfloor NW \rfloor - \lceil R(N, \epsilon) \rceil = 0$ to obtain $\lceil R(N, \epsilon) \rceil = 2\lfloor NW \rfloor \leq 2NW$. For all $N \geq 2$, the first term in Eq. (C5) dominates and therefore, Eq. (C5) reduces to

$$R(N, \epsilon) = O\left(\log\left(\frac{1}{\epsilon}\right) \log N\right). \quad (\text{C12})$$

This naturally implies that

$$NW = O\left(\log\left(\frac{1}{\epsilon}\right) \log N\right). \quad (\text{C13})$$

Recall that m , the additional qubits required for a given precision, can be expressed as $\log(NW)$ because $W = 2^m/2^{l+m}$ is the fraction of the lattice for which the taper has significant support. This gives

$$2^m = O\left(\log\left(\frac{1}{\epsilon}\right) \log N\right) \quad (\text{C14})$$

$$= O((m+l) \log(1/\epsilon)) \quad (\text{C15})$$

$$= O(m \log 1/\epsilon + \log 1/\delta \log 1/\epsilon) \quad (\text{C16})$$

$$\implies m = O(\log(m \log 1/\epsilon + \log 1/\delta \log 1/\epsilon)) \quad (\text{C17})$$

$$= O\left(\log(\log 1/\delta \log 1/\epsilon) + \frac{m \log 1/\epsilon}{\log 1/\delta \log 1/\epsilon}\right) \quad (\text{C18})$$

$$= O\left(\log(\log 1/\delta \log 1/\epsilon) + \frac{m}{\log 1/\delta}\right) \quad (\text{C19})$$

$$\implies m - O\left(\frac{m}{\log 1/\delta}\right) = O(\log(\log 1/\delta \log 1/\epsilon)) \quad (\text{C20})$$

$$\implies m = \frac{O(\log(\log 1/\delta \log 1/\epsilon))}{1 - O\left(\frac{1}{\log 1/\delta}\right)} \quad (\text{C21})$$

$$= O(\log \log 1/\delta + \log \log 1/\epsilon). \quad (\text{C22})$$

The second equality follows by substituting $N = 2^{l+m}$. The fifth equality follows from the fact that $\log(x+y) \leq \log(x) + y/x$ for all $x, y > 0$. This concludes the proof. $\blacksquare$

APPENDIX D: AVERAGE-CASE OPTIMAL TAPERS

For ease of reference, we restate the Lagrangian corresponding to the average case from the main text below:

$$\bar{\mathcal{L}}(|\phi\rangle, \lambda) = \mathbb{E}_{\Delta \sim \mathcal{D}} \left[\sum_{j=-K}^K \left| \hat{\phi}\left(\Delta + \frac{j}{N}\right) \right|^2 \right] - \lambda \left(\sum_{n=0}^{N-1} |\phi[n]|^2 - 1 \right). \quad (\text{D1})$$

As mentioned before, we focus on the case where Δ is uniformly distributed over the interval $[-1/2N, 1/2N]$, i.e., $\mathcal{D}$ is the uniform distribution over this interval. It is reasonable to make this assumption since it is unlikely that the phase values are influenced in a way such that they lie away or close to the lattice points in a systematic way. However, it is possible to generalize the subsequent analysis for other probability distributions, such as the Gaussian distribution.

Please note that the analysis below follows a similar approach to that outlined in Appendix B. Now, consider the following:

$$\bar{\mathcal{L}}(|\phi\rangle, \lambda) = \int_{-\frac{1}{2N}}^{\frac{1}{2N}} d\Delta N \sum_{j=-K}^K \left| \hat{\phi}\left(\Delta + \frac{j}{N}\right) \right|^2 - \lambda \left(\sum_{n=0}^{N-1} |\phi[n]|^2 - 1 \right) \quad (\text{D2})$$

$$= \frac{1}{N} \int_{-\frac{1}{2N}}^{\frac{1}{2N}} d\Delta N \sum_{j=-K}^K \sum_{n,n'=0}^{N-1} \phi[n] \phi^*[n'] e^{2\pi i (\Delta + \frac{j}{N})(n-n')} - \lambda \left(\sum_{n=0}^{N-1} \phi[n] \phi^*[n] - 1 \right) \quad (\text{D3})$$

$$= \sum_{n,n'=0}^{N-1} \phi[n] \phi^*[n'] \int_{-\frac{1}{2N}}^{\frac{1}{2N}} d\Delta e^{2\pi i \Delta (n-n')} \sum_{j=-K}^K e^{-2\pi i (j/N)(n'-n)} - \lambda \left(\sum_{n=0}^{N-1} \phi[n] \phi^*[n] - 1 \right) \quad (\text{D4})$$

$$= \sum_{n,n'=0}^{N-1} \phi[n] \phi^*[n'] \frac{\sin(2\pi (1/2N)(n'-n))}{\pi(n'-n)} \frac{\sin(\pi(n'-n)(2K+1)/N)}{\sin(\pi(n'-n)/N)} - \lambda \left(\sum_{n=0}^{N-1} \phi[n] \phi^*[n] - 1 \right) \quad (\text{D5})$$

$$= \sum_{n,n'=0}^{N-1} \phi[n] \phi^*[n'] \underbrace{\frac{\sin(\pi(n'-n)(2K+1)/N)}{\pi(n'-n)}}_{=:C(n,n')} - \lambda \left(\sum_{n=0}^{N-1} \phi[n] \phi^*[n] - 1 \right). \quad (\text{D6})$$

Then, to find stationary points of $\bar{\mathcal{L}}$, we set all of its partial derivatives to zero. First, we differentiate $\bar{\mathcal{L}}$ with respect to $\phi^*[m]$ for all $m \in \{0, \dots, N-1\}$ and set it to zero,

$$\frac{\partial \bar{\mathcal{L}}}{\partial \phi^*[m]} = \sum_{n=0}^{N-1} \phi[n] C(n, m) - \lambda \phi[m] = 0 \quad (\text{D7})$$

$$\implies \sum_{n=0}^{N-1} \phi[n] C(n, m) = \lambda \phi[m]. \quad (\text{D8})$$

Note that the above equation is an eigenvalue equation, and its eigenvectors are DPSSs, previously studied in classical signal processing [52]. Next, differentiating $\mathcal{L}$ with respect to the Lagrange multiplier λ and setting it to zero leads to the normalization constraint for the taper,

$$\frac{\partial \bar{\mathcal{L}}}{\partial \lambda} = - \sum_{n=0}^{N-1} \phi[n] \phi^*[n] + 1 = 0 \quad (\text{D9})$$

$$\implies \sum_{n=0}^{N-1} \phi[n] \phi^*[n] = 1. \quad (\text{D10})$$

Next, by plugging all the stationary points $(|\phi\rangle, \lambda)$ that satisfy the conditions Eqs. (D8) and (D10) into the objective function of $\bar{\mathcal{L}}$, we get the following:

$$\mathbb{E}_{\Delta \sim \mathcal{D}} \left[\sum_{j=-K}^K \left| \hat{\phi} \left(\Delta + \frac{j}{N} \right) \right|^2 \right] = \lambda. \quad (\text{D11})$$

In other words, this means that the stationary point $(|\phi\rangle, \lambda)$ that maximizes the objective function is the eigenvector of C with the maximum eigenvalue λ . This corresponds to the DPSS with the maximum eigenvalue, which additionally corresponds to the success probability of the algorithm.

APPENDIX E: CLASSICAL SIGNAL ANALYSIS DERIVATION OF DPSS AS OPTIMAL TAPERS

Following Refs. [45, 52], the discrete time Fourier transform (DTFT), $\hat{x} \in L_2([- \frac{1}{2}, \frac{1}{2}])$, of a discrete signal, $x \in l_2(\mathbb{Z})$, is given by

$$\hat{x}(f) := \sum_{n=-\infty}^{\infty} x[n] e^{-2\pi i f n}, \quad f \in \left[-\frac{1}{2}, \frac{1}{2}\right], \quad (\text{E1})$$

with the inverse transform given by

$$x[n] = \int_{-\frac{1}{2}}^{\frac{1}{2}} \hat{x}(f) e^{2\pi i f n} df. \quad (\text{E2})$$

From these definitions, we can see that $x, x' \in l_2(\mathbb{Z})$ satisfy the Parseval-Plancherel inequality, $\langle x, x' \rangle_{l_2(\mathbb{Z})} = \langle \hat{x}, \hat{x}' \rangle_{L_2([-1/2, 1/2])}$. Here, we denote $\langle \cdot, \cdot \rangle_{\mathcal{A}}$ as the inner product defined on the vector space $\mathcal{A}$. We say that $x \in l_2(\mathbb{Z})$ is time-limited to $n \in \{0, \dots, N-1\}$ if $x[n] = 0$ for all $n \in \mathbb{Z} \setminus \{0, \dots, N-1\}$. Furthermore, we say that $x \in l_2(\mathbb{Z})$ is band-limited to $|f| \leq W$ if $\hat{x}(f) = 0$ for $|f| > W$, where $W \in (0, 1/2)$.

The aim here is to find discrete functions $x[n]$ that are time-limited to $n \in \{0, \dots, N-1\}$ and are maximally band-limited to the frequency band $|f| \leq W$. We can formulate this as an optimization problem in the following

way:

$$\max_{x \in l_2(\mathbb{Z})} \int_{-W}^W |\hat{x}(f)|^2 df \quad (\text{E3})$$

$$\text{subject to } \|x\|_{l_2(\mathbb{Z})}^2 = 1, \quad (\text{E4})$$

$$x[n] = 0 \text{ for all } n \in \mathbb{Z} \setminus \{0, \dots, N-1\}. \quad (\text{E5})$$

We can simplify things conceptually by defining the following two operators: the time-limiting operator $\mathcal{T}_N$ as

$$(\mathcal{T}_N x)[n] := \begin{cases} x[n] & \text{if } n \in \{0, \dots, N-1\}, \\ 0 & \text{if } n \in \mathbb{Z} \setminus \{0, \dots, N-1\}, \end{cases} \quad (\text{E6})$$

and the band-limiting operator $\mathcal{B}_W$ as

$$(\mathcal{B}_W x)[n] := \sum_{l=-\infty}^{\infty} \frac{\sin[2W(l-n)]}{\pi(l-n)} x[l] \quad \text{for } n \in \mathbb{Z}. \quad (\text{E7})$$

Note that the DTFT of the band-limiting operator is $\widehat{\mathcal{B}_W x}(f) = \hat{x}(f)$ for $|f| \leq W$ and $\widehat{\mathcal{B}_W x}(f) = 0$ for $|f| > W$.

With these definitions, we see that the integral in Eq. (E3) can be represented as

$$\int_{-W}^W |\hat{x}(f)|^2 df = \langle \hat{x}, \widehat{\mathcal{B}_W x} \rangle_{L_2([-1/2, 1/2])}. \quad (\text{E8})$$

From the Parseval-Plancherel theorem, we have

$$\langle \hat{x}, \widehat{\mathcal{B}_W x} \rangle_{L_2([-1/2, 1/2])} = \langle x, \mathcal{B}_W x \rangle_{l_2(\mathbb{Z})}. \quad (\text{E9})$$

And then

$$\langle x, \mathcal{B}_W x \rangle_{l_2(\mathbb{Z})} = \langle \mathcal{T}_N x, \mathcal{B}_W \mathcal{T}_N x \rangle_{l_2(\mathbb{Z})} \quad (\text{E10})$$

$$= \langle x, \mathcal{T}_N \mathcal{B}_W \mathcal{T}_N x \rangle_{l_2(\mathbb{Z})}, \quad (\text{E11})$$

with the last expression, Eq. (E11), holding because $\mathcal{T}_N$ is self-adjoint.

From Eqs. (E3), (E8), and (E11), we can see that the eigenvector, ϕ_0 , corresponding to the maximum eigenvalue, λ_0 , of the matrix given by

$$[\mathcal{T}_N \mathcal{B}_W \mathcal{T}_N]_{l,n} = \frac{\sin(2\pi W(l-n))}{\pi(l-n)}; \quad l, n \in \{0, \dots, N-1\} \quad (\text{E12})$$

solves the maximization problem. This classical signal processing result gives the same optimal tapers that we find above for the quantum average-case optimal taper.

APPENDIX F: RELATIONSHIP BETWEEN TAPERS AND THE FOURIER CONVOLUTION THEOREM

Here, we aim to show a conceptually useful way to understand tapering, and hence tQPE, originating in the classical signal processing community. We begin by deriving the Fourier convolution theorem in a form useful for our discussion below. Starting with two functions of time, ψ and x , we have that their pointwise product (denoted by “.”)

$$\psi(t) \cdot x(t) = \left(\int_{-\infty}^{\infty} \hat{\psi}(g) e^{-2\pi i g t} dg \right) \cdot \left(\int_{-\infty}^{\infty} \hat{x}(f) e^{-2\pi i f t} df \right) \quad (F1)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{\psi}(g) \hat{x}(f) e^{-2\pi i (g+f)t} dg df \quad (F2)$$

$$= \int_{-\infty}^{\infty} \underbrace{\left(\int_{-\infty}^{\infty} \hat{\psi}(g) \hat{x}(h-g) dg \right)}_{(\hat{\psi} * \hat{x})} e^{-2\pi i h t} dh, \quad (F3)$$

where $*$ denotes a convolution.

Switching notation and denoting the continuous Fourier transform as $\mathcal{F}$, we see from above

$$\psi(t) \cdot x(t) = \mathcal{F}^{-1} \{ \mathcal{F}\{\psi\} * \mathcal{F}\{x\} \} \quad (F4)$$

$$\rightarrow \mathcal{F}\{\psi(t) \cdot x(t)\} = \mathcal{F}\{\psi\} * \mathcal{F}\{x\}. \quad (F5)$$

Note that here, to be consistent with the signs of the complex exponentials, we use the opposite convention for our transform than is usual.

For tQPE, the amplitudes of the state encoded on the ancilla register before the inverse QFT are $\phi[n]e^{2\pi i \theta n}$. Here, $\theta \in [0, 2\pi]$, but the taper, ϕ is a time-limited sequence with discrete support at equally spaced times, $n\Delta t$. We can represent this sequence as a continuous distribution, $\phi(t) \cdot \text{III}(t)$, in the limit as $n \rightarrow \infty$, and $\Delta t \rightarrow 0$, where $\text{III}(t) = \sum_{n=0}^{N-1} \delta(t - n\Delta t)$ is a Dirac comb. In this limit, $\phi[n]e^{2\pi i \theta n} \rightarrow ((\phi(t) \cdot \text{III}(t)) \cdot e^{2\pi i \theta t}) = ((\phi(t) \sum_{n=0}^{N-1} \delta(t - n\Delta t)) \cdot e^{2\pi i \theta t}) = ((\sum_{n=0}^{N-1} \phi(t) \delta(t - n\Delta t)) \cdot e^{2\pi i \theta t})$. To apply the Fourier convolution theorem, we take $x(t) = e^{2\pi i \theta t}$ giving $\mathcal{F}\{x\}(f) = \delta(\theta - f)$, and the discrete and finite in time distribution, $\psi(t) = \sum_{n=0}^{N-1} \phi(t) \delta(t - n\Delta t)$. This then gives us

$$\mathcal{F}\{\psi\}(f) = \int_{-\infty}^{\infty} \sum_{n=0}^{N-1} \phi(t) \delta(t - n\Delta t) e^{-2\pi i f t} dt \quad (F6)$$

$$= \sum_{n=0}^{N-1} \int_{-\infty}^{\infty} \phi(t) e^{-2\pi i f t} \delta(t - n\Delta t) dt \quad (F7)$$

$$= \sum_{n=0}^{N-1} \phi[n] e^{-2\pi i n \Delta t f}. \quad (F8)$$

Rescaling $f \Delta t \rightarrow f$ and similarly for θ , this gives us

$$\mathcal{F}\{\psi(t) \cdot x(t)\}(f) = \{ \mathcal{F}\{\psi\} * \mathcal{F}\{x\} \}(f) \quad (F9)$$

$$= \left\{ \left(\sum_{n=0}^{N-1} \phi[n] e^{-2\pi i n f} \right) * \delta(\theta - f) \right\}(f) \quad (F10)$$

$$= \sum_{n=0}^{N-1} \phi[n] e^{-2\pi i n (\theta - f)} \quad (F11)$$

$$= \mathcal{F}\{\phi\}(\theta - f). \quad (F12)$$

Switching back to our previous notation, for tQPE, we have

$$\mathcal{F}\{\psi(t) \cdot x(t)\} = \hat{\phi}(\theta - f), \quad (F13)$$

the coefficient in Eq. (13). tQPE samples this with the QFT at discrete frequencies, k , as $\text{QFT}^\dagger[\psi(t) \cdot x(t)] = \hat{\phi}_k(\theta - k/N)$.

APPENDIX G: PROOF OF THEOREM 4

In this Appendix we prove Theorem 4 of Sec. V. For convenience we reproduce the theorem statement here. For the definitions refer to Sec. V.

Theorem G1. The following holds:

$$\frac{1}{2} \left\| \mathcal{T}_{CA} - \text{Tr}_B \circ \mathcal{Q}_{BA}^\dagger \circ \tilde{\mathcal{T}}_{CB} \circ \mathcal{Q}_{BA} \circ \mathcal{A}_{CA} \right\|_\diamond \leq 2L_T \delta + 4\epsilon. \quad (G1)$$

Here, Tr_B denotes the partial-trace channel, where the register B is traced out.

Proof. Consider that

$$\begin{aligned} & \frac{1}{2} \left\| \mathcal{T}_{CA} - \text{Tr}_B \circ \mathcal{Q}_{BA}^\dagger \circ \tilde{\mathcal{T}}_{CB} \circ \mathcal{Q}_{BA} \circ \mathcal{A}_{CA} \right\|_\diamond \\ &= \sup_{\rho_{RCA}} \frac{1}{2} \left\| (\mathcal{I}_R \otimes \mathcal{T}_{CA})(\rho_{RCA}) \right. \\ & \quad \left. - (\mathcal{I}_R \otimes \text{Tr}_B \circ \mathcal{Q}_{BA}^\dagger \circ \tilde{\mathcal{T}}_{CB} \circ \mathcal{Q}_{BA} \circ \mathcal{A}_{CA})(\rho_{RCA}) \right\|_1, \end{aligned} \quad (G2)$$

where R is a reference system of arbitrary dimension and $\mathcal{I}_R$ is the identity channel acting on R . Expanding the first term of the right-hand side of the above equation, we get

$$(\mathcal{I}_R \otimes \mathcal{T}_{CA})(\rho_{RCA}) = (I_R \otimes T_{CA})\rho_{RCA}(I_R \otimes T_{AC}^\dagger) \quad (\text{G3})$$

$$= \left(I_R \otimes \sum_{j_1} T(\theta_{j_1})_C \otimes |\psi_{j_1}\rangle\langle\psi_{j_1}|_A \right) \rho_{RCA} \left(I_R \otimes \sum_{j_2} T^\dagger(\theta_{j_2})_C \otimes |\psi_{j_2}\rangle\langle\psi_{j_2}|_A \right) \quad (\text{G4})$$

$$= \left(\sum_{j_1} I_R \otimes T(\theta_{j_1})_C \otimes |\psi_{j_1}\rangle\langle\psi_{j_1}|_A \right) \rho_{RCA} \left(\sum_{j_2} I_R \otimes T^\dagger(\theta_{j_2})_C \otimes |\psi_{j_2}\rangle\langle\psi_{j_2}|_A \right). \quad (\text{G5})$$

This implies that the Kraus operator, which we denote by V_{RCA} , of the above channel is

$$V_{RCA} = \sum_{j_1} I_R \otimes T(\theta_{j_1})_C \otimes |\psi_{j_1}\rangle\langle\psi_{j_1}|_A. \quad (\text{G6})$$

Therefore, this channel can be rewritten as

$$(\mathcal{I}_R \otimes \mathcal{T}_{CA})(\rho_{RCA}) = V_{RCA}\rho_{RCA}V_{RCA}^\dagger. \quad (\text{G7})$$

Similarly, we get the Kraus representation of the second channel in Eq. (G2). For this, consider the following:

$$\begin{aligned} & \mathcal{I}_R \otimes \text{Tr}_B \circ \mathcal{Q}_{BA}^\dagger \circ \tilde{\mathcal{T}}_{CB} \circ \mathcal{Q}_{BA} \circ \mathcal{A}_{CA}(\rho_{RCA}) \\ &= I_R \otimes \text{Tr}_B \circ \mathcal{Q}_{BA}^\dagger \circ \tilde{\mathcal{T}}_{CB} \circ \mathcal{Q}_{BA} \circ (\rho_{RCA} \otimes |0\rangle\langle 0|_B) \end{aligned} \quad (\text{G8})$$

$$= I_R \otimes \text{Tr}_B \circ \mathcal{Q}_{BA}^\dagger \circ \tilde{\mathcal{T}}_{CB} \left(\mathcal{Q}_{BA}(\rho_{RCA} \otimes |0\rangle\langle 0|_B) \mathcal{Q}_{AB}^\dagger \right) \quad (\text{G9})$$

$$= I_R \otimes \text{Tr}_B \circ \mathcal{Q}_{BA}^\dagger \circ \tilde{\mathcal{T}}_{CB} \left(\sum_{j_1, x_1, y_1} f_{j_1 x_1 y_1} |x_1\rangle\langle y_1|_B \otimes |\psi_{j_1}\rangle\langle\psi_{j_1}|_A (\rho_{RCA} \otimes |0\rangle\langle 0|_B) \sum_{j_2, x_2, y_2} f_{j_2 x_2 y_2}^* |y_2\rangle\langle x_2|_B \otimes |\psi_{j_2}\rangle\langle\psi_{j_2}|_A \right) \quad (\text{G10})$$

$$= I_R \otimes \text{Tr}_B \circ \mathcal{Q}_{BA}^\dagger \circ \tilde{\mathcal{T}}_{CB} \left(\sum_{j_1, j_2, x_1, x_2} f_{j_1 x_1} f_{j_2 x_2}^* \langle\psi_{j_1}|\rho_{RCA}|\psi_{j_2}\rangle \otimes |x_1\rangle\langle x_2|_B \otimes |\psi_{j_1}\rangle\langle\psi_{j_2}|_A \right) \quad (\text{G11})$$

$$= I_R \otimes \text{Tr}_B \circ \mathcal{Q}_{BA}^\dagger \left(\sum_{j_1, j_2, x_1, x_2} f_{j_1 x_1} f_{j_2 x_2}^* T(g(x_1))_C \langle\psi_{j_1}|\rho_{RCA}|\psi_{j_2}\rangle T^\dagger(g(x_2))_C \otimes |x_1\rangle\langle x_2|_B \otimes |\psi_{j_1}\rangle\langle\psi_{j_2}|_A \right) \quad (\text{G12})$$

$$= I_R \otimes \text{Tr}_B \left(\sum_{j_1, j_2, x_1, x_2, y_3, y_4} f_{j_1 x_1} f_{j_2 x_2}^* f_{j_1 x_1 y_3}^* f_{j_2 x_2 y_4} T(g(x_1))_C \langle\psi_{j_1}|\rho_{RCA}|\psi_{j_2}\rangle T^\dagger(g(x_2))_C \otimes |y_3\rangle\langle y_4|_B \otimes |\psi_{j_1}\rangle\langle\psi_{j_2}|_A \right) \quad (\text{G13})$$

$$= I_R \otimes \left(\sum_{j_1, j_2, x_1, x_2, y_3} f_{j_1 x_1} f_{j_2 x_2}^* f_{j_1 x_1 y_3}^* f_{j_2 x_2 y_3} T(g(x_1))_C \langle\psi_{j_1}|\rho_{RCA}|\psi_{j_2}\rangle T^\dagger(g(x_2))_C \otimes |\psi_{j_1}\rangle\langle\psi_{j_2}|_A \right) \quad (\text{G14})$$

$$= \sum_{j_1, j_2} \left(\sum_{x_1, x_2, y_3} f_{j_1 x_1} 0 f_{j_2 x_2}^* f_{j_1 x_1 y_3}^* f_{j_2 x_2 y_3} (I_R \otimes T(g(x_1))_C) \langle \psi_{j_1} | \rho_{RCA} | \psi_{j_2} \rangle (I_R \otimes T^\dagger(g(x_2))_C) \otimes |\psi_{j_1}\rangle \langle \psi_{j_2}|_A \right) \quad (\text{G15})$$

$$= \sum_{y_3} \left(\underbrace{\sum_{j_1, x_1} f_{j_1 x_1} 0 f_{j_1 x_1 y_3}^* (I_R \otimes T(g(x_1))_C \otimes |\psi_{j_1}\rangle \langle \psi_{j_1}|_A)}_{=: K_{y_3}} \right) \rho_{RCA} \left(\underbrace{\sum_{j_2, x_2} f_{j_2 x_2}^* 0 f_{j_2 x_2 y_3} (I_R \otimes T^\dagger(g(x_2))_C \otimes |\psi_{j_2}\rangle \langle \psi_{j_2}|_A)}_{K_{y_3}^\dagger} \right) \quad (\text{G16})$$

$$= \sum_{y_3} K_{y_3} \rho_{RCA} K_{y_3}^\dagger. \quad (\text{G17})$$

Plugging Eqs. (G7) and (G17) into Eq. (G2) and removing the system labels for simplicity, we get

$$\frac{1}{2} \left\| \mathcal{T}_{AC} - \text{Tr}_B \circ \mathcal{Q}_{BA}^\dagger \circ \tilde{\mathcal{T}}_{CB} \circ \mathcal{Q}_{BA} \circ \mathcal{A}_{CA} \right\|_\diamond = \sup_\rho \frac{1}{2} \left\| V \rho V^\dagger - \sum_{y_3} K_{y_3} \rho K_{y_3}^\dagger \right\|_1. \quad (\text{G18})$$

In the perfect QPE case, uncomputation is exact, resulting in register B being perfectly reset to the all-zeros state after applying $\mathcal{Q}^\dagger$ (see Fig. 7). Consequently, in this scenario, there exists only a single Kraus operator (corresponding to $y_3 = 0$), and this Kraus operator is the unitary operator V . For the imperfect QPE case, we expect this operator to remain approximately unitary, as we will demonstrate next.

Continuing,

$$\frac{1}{2} \left\| \mathcal{T}_{AC} - \text{Tr}_B \circ \mathcal{Q}_{BA}^\dagger \circ \tilde{\mathcal{T}}_{CB} \circ \mathcal{Q}_{BA} \circ \mathcal{A}_{CA} \right\|_\diamond = \sup_\rho \frac{1}{2} \left\| V \rho V^\dagger - K_0 \rho K_0^\dagger - \sum_{y_3 \neq 0} K_{y_3} \rho K_{y_3}^\dagger \right\|_1 \quad (\text{G19})$$

$$\leq \sup_\rho \frac{1}{2} \left(\left\| V \rho V^\dagger - K_0 \rho K_0^\dagger \right\|_1 + \left\| \sum_{y_3 \neq 0} K_{y_3} \rho K_{y_3}^\dagger \right\|_1 \right) \quad (\text{G20})$$

$$= \sup_\rho \frac{1}{2} \left\| V \rho V^\dagger - K_0 \rho K_0^\dagger \right\|_1 + \sup_\rho \frac{1}{2} \left\| \sum_{y_3 \neq 0} K_{y_3} \rho K_{y_3}^\dagger \right\|_1, \quad (\text{G21})$$

where the first inequality follows from the triangle inequality.

Let us begin by bounding the first term of the above equation:

$$\sup_\rho \left\| V \rho V^\dagger - K_0 \rho K_0^\dagger \right\|_1 = \sup_\rho \left\| V \rho V^\dagger - K_0 \rho V^\dagger + K_0 \rho V^\dagger - K_0 \rho K_0^\dagger \right\|_1 \quad (\text{G22})$$

$$\leq \sup_\rho \left\| V \rho V^\dagger - K_0 \rho V^\dagger \right\|_1 + \sup_\rho \left\| K_0 \rho V^\dagger - K_0 \rho K_0^\dagger \right\|_1 \quad (\text{G23})$$

$$= \sup_\rho \left\| (V - K_0) \rho V^\dagger \right\|_1 + \sup_\rho \left\| K_0 \rho (V^\dagger - K_0^\dagger) \right\|_1 \quad (\text{G24})$$

$$\leq \sup_\rho \|V - K_0\| \|\rho\|_1 \|V^\dagger\| + \sup_\rho \|K_0\| \|\rho\|_1 \|V^\dagger - K_0^\dagger\| \quad (\text{G25})$$

$$= \|V - K_0\| + \|K_0\| \|V - K_0\| \quad (\text{G26})$$

$$= (1 + \|K_0\|) \|V - K_0\|, \quad (\text{G27})$$

where the first inequality follows from the triangle inequality and the second inequality follows from the Hölder's inequality. Next, we bound the quantities $\|K_0\|$ and $\|V - K_0\|$ from above one by one,

$$\|K_0\| = \left\| \sum_{j_1, x_1} f_{j_1 x_1 0} f_{j_1 x_1 0}^* (I_R \otimes T(g(x_1))_C \otimes |\psi_{j_1}\rangle\langle\psi_{j_1}|_A) \right\| \quad (\text{G28})$$

$$= \max_{j_1} \left\| \sum_{x_1} |f_{j_1 x_1 0}|^2 (I_R \otimes T(g(x_1))_C) \right\| \quad (\text{G29})$$

$$\leq \max_{j_1} \sum_{x_1} |f_{j_1 x_1 0}|^2 \|I_R \otimes T(g(x_1))_C\| \quad (\text{G30})$$

$$\leq 1. \quad (\text{G31})$$

The second equality follows from Lemma H1, the first inequality follows from the triangle inequality, and the last inequality follows from the fact that for every x_1 , we have $|f_{j_1 x_1 0}|^2 \leq 1$. We now bound $\|V - K_0\|$ from above:

$$\|V - K_0\| = \left\| \sum_{j_1} I_R \otimes T(\theta_{j_1})_C \otimes |\psi_{j_1}\rangle\langle\psi_{j_1}|_A - \sum_{j_1, x_1} f_{j_1 x_1 0} f_{j_1 x_1 0}^* (I_R \otimes T(g(x_1))_C \otimes |\psi_{j_1}\rangle\langle\psi_{j_1}|_A) \right\| \quad (\text{G32})$$

$$= \left\| \sum_{j_1} T(\theta_{j_1})_C \otimes |\psi_{j_1}\rangle\langle\psi_{j_1}|_A - \sum_{j_1, x_1} |f_{j_1 x_1 0}|^2 (T(g(x_1))_C \otimes |\psi_{j_1}\rangle\langle\psi_{j_1}|_A) \right\| \quad (\text{G33})$$

$$= \max_{j_1} \left\| T(\theta_{j_1})_C - \sum_{x_1} |f_{j_1 x_1 0}|^2 T(g(x_1))_C \right\| \quad (\text{G34})$$

$$= \max_{j_1} \left\| \sum_{x_1} |f_{j_1 x_1 0}|^2 T(\theta_{j_1})_C - \sum_{x_1} |f_{j_1 x_1 0}|^2 T(g(x_1))_C \right\| \quad (\text{G35})$$

$$\leq \max_{j_1} \sum_{x_1} |f_{j_1 x_1 0}|^2 \|T(\theta_{j_1})_C - T(g(x_1))_C\| \quad (\text{G36})$$

$$= \max_{j_1} \left(\sum_{x_1: |\theta_{j_1} - g(x_1)| \leq \delta} |f_{j_1 x_1 0}|^2 \|T(\theta_{j_1})_C - T(g(x_1))_C\| + \sum_{x_1: |\theta_{j_1} - g(x_1)| > \delta} |f_{j_1 x_1 0}|^2 \|T(\theta_{j_1})_C - T(g(x_1))_C\| \right) \quad (\text{G37})$$

$$\leq \max_{j_1} \left(\sum_{x_1: |\theta_{j_1} - g(x_1)| \leq \delta} |f_{j_1 x_1 0}|^2 (L_T \delta) + \sum_{x_1: |\theta_{j_1} - g(x_1)| > \delta} |f_{j_1 x_1 0}|^2 (2) \right) \quad (\text{G38})$$

$$\leq \max_{j_1} (L_T \delta + 2\epsilon) \quad (\text{G39})$$

$$= L_T \delta + 2\epsilon, \quad (\text{G40})$$

where the first inequality follows from the multiplicativity of the spectral norm under tensor products, the second equality follows from Lemma H1, the second inequality follows from the triangle inequality, the third inequality follows from Eq. (69), and the last inequality follows from the following fact about the error probability of the QPE algorithm used:

$$\sum_{x_1: |\theta_{j_1} - g(x_1)| > \delta} |f_{j_1 x_1 0}|^2 \leq \epsilon. \quad (\text{G41})$$

To this end, using the bounds given by Eqs. (G31) and (G40) in Eq. (G27), we get

$$\sup_{\rho} \|V \rho V^\dagger - K_0 \rho K_0^\dagger\|_1 \leq 2L_T \delta + 4\epsilon. \quad (\text{G42})$$

Now that we have bounded the first term of Eq. (G21), we focus on bounding its second term. For this, consider the following:

$$\left\| \sum_{y_3 \neq 0} K_{y_3}^\dagger K_{y_3} \right\| = \|I - K_0^\dagger K_0\| \quad (\text{G43})$$

$$= \|V^\dagger V - K_0^\dagger K_0\| \quad (\text{G44})$$

$$= \|V^\dagger V - K_0^\dagger V + K_0^\dagger V - K_0^\dagger K_0\| \quad (\text{G45})$$

$$\leq \|V^\dagger V - K_0^\dagger V\| + \|K_0^\dagger V - K_0^\dagger K_0\| \quad (\text{G46})$$

$$\leq \|V^\dagger - K_0^\dagger\| \|V\| + \|K_0^\dagger\| \|V - K_0\| \quad (\text{G47})$$

$$\leq 2 \|V - K_0\| \quad (\text{G48})$$

$$\leq 2L_T\delta + 4\epsilon, \quad (\text{G49})$$

where the first inequality follows from the triangle inequality, the second inequality follows from the submultiplicativity property of the spectral norm, the third inequality follows from Eq. (G31), and the last inequality follows from Eq. (G40). Next, we have the following equality:

$$\sup_{\rho} \left\| \sum_{y_3 \neq 0} K_{y_3} \rho K_{y_3}^\dagger \right\|_1 = \left\| \sum_{y_3 \neq 0} K_{y_3}^\dagger K_{y_3} \right\|. \quad (\text{G50})$$

This follows due to the following fact, where each A_i is some linear operator:

$$\left\| \sum_i A_i^\dagger A_i \right\| = \sup_{\rho} \text{Tr} \left[\sum_i A_i^\dagger A_i \rho \right] \quad (\text{G51})$$

$$= \sup_{\rho} \sum_i \text{Tr} [A_i^\dagger A_i \rho] \quad (\text{G52})$$

$$= \sup_{\rho} \sum_i \text{Tr} [A_i \rho A_i^\dagger] \quad (\text{G53})$$

$$= \sup_{\rho} \text{Tr} \left[\sum_i A_i \rho A_i^\dagger \right] \quad (\text{G54})$$

Proof.

$$\frac{1}{2} \|\mathcal{U} - \mathcal{V}\|_{\diamond} = \sup_{\rho} \frac{1}{2} \|(\mathcal{I} \otimes \mathcal{U})(\rho) - (\mathcal{I} \otimes \mathcal{V})(\rho)\|_1 \quad (\text{H3})$$

$$= \sup_{\rho} \frac{1}{2} \|(I \otimes U)\rho(I \otimes U^\dagger) - (I \otimes V)\rho(I \otimes V^\dagger)\|_1 \quad (\text{H4})$$

$$= \sup_{\rho} \left\| \sum_i A_i \rho A_i^\dagger \right\|_1. \quad (\text{G55})$$

Using the identity given by Eq. (G50), and the bound given in Eq. (G49), we get

$$\sup_{\rho} \left\| \sum_{y_3 \neq 0} K_{y_3} \rho K_{y_3}^\dagger \right\|_1 \leq 2L_T\delta + 4\epsilon. \quad (\text{G56})$$

Finally, combining the bounds of Eq. (G42) and Eq. (G56) to bound Eq. (G21) from above, we get

$$\frac{1}{2} \|\mathcal{T}_{AC} - \text{Tr}_B \circ \mathcal{Q}_{BA}^\dagger \circ \tilde{\mathcal{T}}_{CB} \circ \mathcal{Q}_{BA} \circ \mathcal{A}_{CA}\|_{\diamond} \leq 2L_T\delta + 4\epsilon. \quad (\text{G57})$$

This concludes the proof. $\blacksquare$

APPENDIX H: USEFUL LEMMAS

Lemma H1 (Norm of direct-sum of operators). Let $A = A_1 \oplus \dots \oplus A_n$, where each A_i is a linear operator and n is some natural number. Then the following is true:

$$\|A\| = \max_i \|A_i\|. \quad (\text{H1})$$

Proof. For each $i \in \{1, \dots, n\}$, consider the following singular value decomposition: $A_i = U_i D_i V_i^\dagger$, where U_i and V_i are unitaries and D_i is a diagonal matrix consisting of singular values. Then the singular value decomposition of $A = U D V^\dagger$, where $U = U_1 \oplus \dots \oplus U_n$, $D = D_1 \oplus \dots \oplus D_n$, and $V = V_1 \oplus \dots \oplus V_n$. Using this and the fact that the spectral norm of a matrix is equal to its maximum singular value, we get $\|A\| = \max_i \|A_i\|$. $\blacksquare$

Lemma H2 (Bound on diamond distance between unitary channels; Alternate simple proof of Lemma 12.6 of [67]). Let $\mathcal{U}(\cdot) := U(\cdot)U^\dagger$ and $\mathcal{V}(\cdot) := V(\cdot)V^\dagger$ be two unitary channels with unitaries U and V , respectively. Then the following holds:

$$\frac{1}{2} \|\mathcal{U} - \mathcal{V}\|_{\diamond} \leq \|U - V\|. \quad (\text{H2})$$

$$= \sup_{\rho} \frac{1}{2} \left\| (I \otimes U)\rho(I \otimes U^{\dagger}) - (I \otimes U)\rho(I \otimes V^{\dagger}) + (I \otimes U)\rho(I \otimes V^{\dagger}) - (I \otimes V)\rho(I \otimes V^{\dagger}) \right\|_1 \quad (\text{H5})$$

$$\leq \sup_{\rho} \frac{1}{2} \left\| (I \otimes U)\rho(I \otimes U^{\dagger}) - (I \otimes U)\rho(I \otimes V^{\dagger}) \right\|_1 + \frac{1}{2} \left\| (I \otimes U)\rho(I \otimes V^{\dagger}) - (I \otimes V)\rho(I \otimes V^{\dagger}) \right\|_1 \quad (\text{H6})$$

$$= \sup_{\rho} \frac{1}{2} \left\| (I \otimes U)\rho(I \otimes U^{\dagger} - I \otimes V^{\dagger}) \right\|_1 + \frac{1}{2} \left\| (I \otimes U - I \otimes V)\rho(I \otimes V^{\dagger}) \right\|_1 \quad (\text{H7})$$

$$\leq \sup_{\rho} \frac{1}{2} \|I \otimes U\| \|\rho\|_1 \|I \otimes U^{\dagger} - I \otimes V^{\dagger}\| + \frac{1}{2} \|I \otimes U - I \otimes V\| \|\rho\|_1 \|I \otimes V^{\dagger}\| \quad (\text{H8})$$

$$= \frac{1}{2} \|I \otimes U^{\dagger} - I \otimes V^{\dagger}\| + \frac{1}{2} \|I \otimes U - I \otimes V\| \quad (\text{H9})$$

$$= \frac{1}{2} \|I \otimes U - I \otimes V\| + \frac{1}{2} \|I \otimes U - I \otimes V\| \quad (\text{H10})$$

$$= \frac{1}{2} \|U - V\| + \frac{1}{2} \|U - V\| \quad (\text{H11})$$

$$= \|U - V\|. \quad (\text{H12})$$

The first inequality follows from the triangle inequality, the second inequality follows from the generalized Hölder's inequality, and the third inequality holds due to the multiplicativity property of the trace norm under tensor products. ■

-
- [1] A. Yu Kitaev, Quantum measurements and the Abelian stabilizer problem, [arXiv:quant-ph/9511026](#).
 - [2] Peter W. Shor, in *Proceedings 35th Annual Symposium on Foundations of Computer Science* (IEEE Computer Society, Los Alamitos, CA, USA, 1994), p. 124.
 - [3] Aram W. Harrow, Avinatan Hassidim, and Seth Lloyd, Quantum algorithm for linear systems of equations, *Phys. Rev. Lett.* **103**, 150502 (2009).
 - [4] Gilles Brassard, Peter Høyer, Michele Mosca, and Alain Tapp, Quantum amplitude amplification and estimation, *Contemp. Math.* **30**, 53 (2002).
 - [5] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost, Quantum principal component analysis, *Nat. Phys.* **10**, 631 (2014).
 - [6] Daniel Nagaj, Paweł Wocjan, and Yong Zhang, Fast amplification of QMA, *Quantum Inf. Comput.* **9**, 1053 (2009).
 - [7] Dong An, Noah Linden, Jin-Peng Liu, Ashley Montanaro, Changpeng Shao, and Jiasu Wang, Quantum-accelerated multilevel Monte Carlo methods for stochastic differential equations in mathematical finance, *Quantum* **5**, 481 (2021).
 - [8] Anirban Narayan Chowdhury, Yiğit Subaşı, and Rolando D. Somma, Improved implementation of reflection operators, [arXiv:1803.02466](#).
 - [9] Alexander Engel, Graeme Smith, and Scott E. Parker, Linear embedding of nonlinear dynamical systems and prospects for efficient quantum algorithms, *Phys. Plasmas* **28**, 062305 (2021).
 - [10] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost, Quantum algorithms for supervised and unsupervised machine learning, [arXiv:1307.0411](#).
 - [11] Patrick Rall, Quantum algorithms for estimating physical quantities using block encodings, *Phys. Rev. A* **102**, 022408 (2020).
 - [12] Lewis Wright, Fergus Barratt, James Dborin, George H. Booth, and Andrew G. Green, Automatic post-selection by ancillae thermalization, *Phys. Rev. Res.* **3**, 033151 (2021).
 - [13] Nathan Wiebe, Daniel Braun, and Seth Lloyd, Quantum algorithm for data fitting, *Phys. Rev. Lett.* **109**, 050505 (2012).
 - [14] Nikhil S. Mande and Ronald de Wolf, in *31st Annual European Symposium on Algorithms (ESA 2023)*, edited by Inge Li Gørtz, Martin Farach-Colton, Simon J. Puglisi, and Grzegorz Herman (Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2023), p. 81:1.
 - [15] Richard Cleve, Artur Ekert, Chiara Macchiavello, and Michele Mosca, Quantum algorithms revisited, *Proc. R. Soc. Lond. Ser. A: Math. Phys. Eng. Sci.* **454**, 339 (1998).
 - [16] Patrick Rall, Faster coherent quantum algorithms for phase, energy, and amplitude estimation, *Quantum* **5**, 566 (2021).
 - [17] Robert Beals, Stephen Brierley, Oliver Gray, Aram W. Harrow, Samuel Kutin, Noah Linden, Dan Shepherd, and Mark Stather, Efficient distributed quantum computing, *Proc. R. Soc.: Math. Phys. Eng. Sci.* **469**, 20120686 (2013).
 - [18] Hoyer, Neerbek, Shi, Quantum complexities of ordered searching, sorting, and element distinctness, *Algorithmica* **34**, 429 (2002).
 - [19] Hartmut Klauck, in *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, USA, 2003), p. 69.
 - [20] L. Hales and S. Hallgren, in *2013 IEEE 54th Annual Symposium on Foundations of Computer Science* (IEEE Comput. Society, Los Alamitos, CA, USA, 2000).

- [21] Kristan Temme, Tobias J. Osborne, Karl G. Vollbrecht, David Poulin, and Frank Verstraete, Quantum metropolis sampling, *Nature* **471**, 87 (2011).
- [22] X. I. LU, Unbiased quantum phase estimation, *Quantum Inf. Comput.* **23**, 0016 (2023).
- [23] François Chapeau-Blondeau and Etienne Belin, Fourier-transform quantum phase estimation with quantum phase noise, *Signal Process.* **17**, 107441 (2020).
- [24] Zhiyan Ding and Lin Lin, Even shorter quantum circuit for phase estimation on early fault-tolerant quantum computers with applications to ground-state energy estimation, *PRX Quantum* **4**, 020331 (2023).
- [25] Valentin Gebhart, Augusto Smerzi, and Luca Pezzè, Bayesian quantum multiphase estimation algorithm, *Phys. Rev. Appl.* **16**, 014035 (2021).
- [26] Lin Lin and Yu Tong, Heisenberg-limited ground-state energy estimation for early fault-tolerant quantum computers, *PRX Quantum* **3**, 010318 (2022).
- [27] Hongkang Ni, Haoya Li, and Lexing Ying, On low-depth algorithms for quantum phase estimation, *Quantum* **7**, 1165 (2023).
- [28] Thomas E. O'Brien, Brian Tarasinski, and Barbara M. Terhal, Quantum phase estimation of multiple eigenvalues for small-scale (noisy) experiments, *New J. Phys.* **21**, 023022 (2019).
- [29] Krysta M. Svore, Matthew B. Hastings, and Michael Freedman, Faster phase estimation, *Quantum Inf. Comput.* **14**, 306 (2013).
- [30] Kianna Wan, Mario Berta, and Earl T. Campbell, Randomized quantum algorithm for statistical phase estimation, *Phys. Rev. Lett.* **129**, 030503 (2022).
- [31] András Gilyén, Yuan Su, Guang Hao Low, and Nathan Wiebe, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, USA, 2019), p. 193.
- [32] John M. Martyn, Zane M. Rossi, Andrew K. Tan, and Isaac L. Chuang, Grand unification of quantum algorithms, *PRX Quantum* **2**, 040203 (2021).
- [33] Vladimir Bužek, Radoslav Derka, and Serge Massar, Optimal quantum clocks, *Phys. Rev. Lett.* **82**, 2207 (1999).
- [34] Dominic W. Berry and Howard M. Wiseman, Optimal states and almost optimal adaptive measurements for quantum interferometry, *Phys. Rev. Lett.* **85**, 5098 (2000).
- [35] Gumaro Rendon, Taku Izubuchi, and Yuta Kikuchi, Effects of cosine tapering window on quantum phase estimation, *Phys. Rev. D* **106**, 034503 (2022).
- [36] Rolando D. Somma, Quantum eigenvalue estimation via time series analysis, *New J. Phys.* **21**, 123025 (2019).
- [37] François Fillion-Gourdeau, Steve MacLean, and Raymond Laflamme, Efficient state initialization by a quantum spectral filtering algorithm, *Phys. Rev. A* **95**, 042331 (2017).
- [38] Alessandro Roggero, Spectral-density estimation with the Gaussian integral transform, *Phys. Rev. A* **102**, 022409 (2020).
- [39] Chi-Fang Chen, M. J. Kastoryano, F. G. S. L. Brandao, and A. Gilyén, Quantum thermal state preparation, [arXiv:2303.18224](https://arxiv.org/abs/2303.18224).
- [40] Chi-Fang Chen, Michael J. Kastoryano, and András Gilyén, An efficient and exact noncommutative quantum Gibbs sampler, [arXiv:2311.09207](https://arxiv.org/abs/2311.09207).
- [41] Jonathan E. Moussa, Low-depth quantum metropolis algorithm, [arXiv:1903.01451](https://arxiv.org/abs/1903.01451).
- [42] Gumaro Rendon and Peter D. Johnson, Low-depth Gaussian state energy estimation, [arXiv:2309.16790](https://arxiv.org/abs/2309.16790).
- [43] George Ashley Campbell and Ronald M. Foster, *Fourier Integrals for Practical Applications* (D. van Nostrand, New York, 1948).
- [44] James F. Kaiser, in *System Analysis by Digital Computer* (Wiley, New York, NY, 1966), p. 218.
- [45] Santhosh Karnik, Justin Romberg, and Mark A. Davenport, Improved bounds for the eigenvalues of prolate spheroidal wave functions and discrete prolate spheroidal sequences, *Appl. Comput. Harmon. Anal.* **5**, 97 (2021).
- [46] Sam McArdle, András Gilyén, and Mario Berta, Quantum state preparation without coherent arithmetic, [arXiv:2210.14892](https://arxiv.org/abs/2210.14892).
- [47] Dominic W. Berry, Yuan Su, Casper Gyurik, Robbie King, Joao Basso, Alexander Del Toro Barba, Abhishek Rajput, Nathan Wiebe, Vedran Dunjko, and Ryan Babbush, Quantifying quantum advantage in topological data analysis, *PRX Quantum* **5**, 010319 (2024).
- [48] Noah Linden and Ronald de Wolf, Average-case verification of the quantum Fourier transform enables worst-case phase estimation, *Quantum* **6**, 872 (2022).
- [49] Henry J. Landau and Henry O. Pollak, Prolate spheroidal wave functions, Fourier analysis and uncertainty—II, *Bell Syst. Tech. J.* **40**, 65 (1961).
- [50] Henry J. Landau and Henry O. Pollak, Prolate spheroidal wave functions, Fourier analysis and uncertainty—III: The dimension of the space of essentially time- and band-limited signals, *Bell Syst. Tech. J.* **41**, 1295 (1962).
- [51] David Slepian, Prolate spheroidal wave functions, Fourier analysis and uncertainty—IV: Extensions to many dimensions; generalized prolate spheroidal functions, *Bell Syst. Tech. J.* **43**, 3009 (1964).
- [52] David Slepian, Prolate spheroidal wave functions, Fourier analysis, and uncertainty—V: The discrete case, *Bell Syst. Tech. J.* **57**, 1371 (1978).
- [53] David Slepian and Henry O. Pollak, Prolate spheroidal wave functions, Fourier analysis and uncertainty—I, *Bell Syst. Tech. J.* **40**, 43 (1961).
- [54] Donald B. Percival and Andrew T. Walden, *Spectral Analysis for Physical Applications* (Cambridge University Press, Cambridge, 1993).
- [55] Simon Haykin, David J. Thomson, and Jeffrey H. Reed, Spectrum sensing for cognitive radio, *Proc. IEEE* **97**, 849 (2009).
- [56] Partha P. Mitra and Bijan Pesaran, Analysis of dynamic brain imaging data, *Biophys. J.* **76**, 691 (1999).
- [57] Jeffrey Park, Craig R. Lindberg, and Frank L. Vernon III, Multitaper spectral analysis of high-frequency seismograms, *J. Geophys. Res.: Solid Earth* **92**, 12675 (1987).
- [58] Andrew Sornborger and Takeshi Yokoo, A multivariate, multitaper approach to detecting and estimating harmonic response in cortical optical imaging data, *J. Neurosci. Methods* **203**, 254 (2012).

- [59] David J. Thomson, Spectrum estimation and harmonic analysis, [Proc. IEEE](#) **70**, 1055 (1982).
- [60] Zhihui Zhu, Santhosh Karnik, Mark A. Davenport, Justin Romberg, and Michael B. Wakin, The eigenvalue distribution of discrete periodic time-frequency limiting operators, [IEEE Signal Process. Lett.](#) **25**, 95 (2017).
- [61] Wen Yuan Xu and Christodoulos Chamzas, On the periodic discrete prolate spheroidal sequences, [SIAM J. Appl. Math.](#) **44**, 1210 (1984).
- [62] Vivek V. Shende, Stephen S. Bullock, and Igor L. Markov, in [Proceedings of the 2005 Asia and South Pacific Design Automation Conference](#) (Association for Computing Machinery, New York, NY, USA, 2005), p. 272.
- [63] Wojciech Rządowski and Rafał Demkowicz-Dobrzański, Discrete-to-continuous transition in quantum phase estimation, [Phys. Rev. A](#) **96**, 032319 (2017).
- [64] A. Luis and J. Peřina, Optimum phase-shift estimation and the quantum description of the phase difference, [Phys. Rev. A](#) **54**, 4564 (1996).
- [65] Sean Greenaway, William Pol, and Sukin Sim, A case study against QSVT: Assessment of quantum phase estimation improved by signal processing techniques, [arXiv:2404.01396](#).
- [66] D. Coppersmith, An approximate Fourier transform useful in quantum factoring, IBM Research Report, 1994, p. RC-19642.
- [67] Dorit Aharonov, Alexei Kitaev, and Noam Nisan, in [Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing, STOC '98](#) (Association for Computing Machinery, New York, NY, USA, 1998), p. 20.