

Optimal Coherent Quantum Phase Estimation via Tapering

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

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

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

³Department of Computer Science, University of Maryland, College Park, MD, USA.

⁴Information Sciences, Los Alamos National Laboratory, Los Alamos, NM, USA.

(Dated: March 29, 2024)

Quantum phase estimation is one of the fundamental primitives that underpins many quantum algorithms, including quantum amplitude estimation, the HHL algorithm for solving linear systems of equations, and quantum principal component analysis. Due to its significance as a subroutine, in this work, we study 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. Unlike most existing phase estimation algorithms, which employ intermediary measurements steps that inevitably destroy coherence, only a couple of algorithms, including the well-known standard quantum phase estimation algorithm, consider this coherent setting. In this work, we propose an improved version of this standard algorithm that utilizes tapering/window functions. Our algorithm, which we call tapered quantum phase estimation algorithm, achieves the optimal query complexity (total number of calls to U and controlled- U) without requiring the use of a computationally expensive quantum sorting network for median computation, which the standard algorithm uses to boost the success probability arbitrarily close to one. We also show that the tapering functions that we use are optimal by formulating optimization problems with different optimization criteria. Beyond the asymptotic regime, we also provide non-asymptotic query complexity of our algorithm, as it is crucial for practical implementation. Finally, we also propose an efficient algorithm that prepares the quantum state corresponding to the optimal tapering function.

CONTENTS

I. Introduction	1
A. Contributions	2
B. Overview	3
C. Related work	4
II. A brief note on Discrete Prolate Spheroidal Sequences	4
III. Tapered Quantum Phase Estimation	5
IV. Optimal Tapers	6
A. The Optimization Problem	6
B. Ideal Case	6
1. Special Case: $\Delta = \pm 1/2N$	7
C. Average-case Optimal Tapers	7
D. Analysis of worst-case error	10
E. Numerics	11
V. Discussion	11
VI. Acknowledgements	13
References	13
A. Preparation of the DPSS Taper	15

B. Ideal Case	17
C. Relationship Between P-DPSS Tapers and Ideal Case Optimal Tapers	19
D. Average-Case	21
E. Classical Signal Analysis Derivation of DPSS as Optimal Tapers	22

I. INTRODUCTION

Quantum phase estimation (QPE) has been central to the field of quantum computing since its introduction [Kit95]. It has been used in Shor's algorithm for efficiently factoring large numbers [Sho94], in the Harrow-Hassidim-Lloyd (HHL) algorithm to solve a system of linear equations [HHL09], for quantum amplitude estimation [BHMT02], for quantum principal component analysis [LMR14], for fast Quantum Merlin-Arthur (QMA) amplification [NWZ09], and as a subroutine in many other applications [WBL12, LMR13, CSS18, WBD⁺21, ESP21, ALL⁺21, Ral20].

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 $\hat{\theta}$ that is δ -close to θ with a probability at least $1 - \epsilon$ for some $\delta, \epsilon > 0$.

* The first two authors contributed equally to this work.

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 complexity for a given circuit implementation of the unitary by simple multiplication.

Probably the simplest approach to the phase estimation 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 approximate 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 [Kit95]. A crucial insight for this quadratic speedup was to extract the phase value bit-by-bit using controlled- U^i for some suitable powers i , rather than simply using controlled- U in each iteration like in the Hadamard test. Interestingly, the authors of Ref. [MdW23] 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 $\{|\psi_{\theta_r}\rangle\}_r$ of U : $|\psi\rangle = \sum_r c_r |\psi_{\theta_r}\rangle$. Suppose that the eigenvalues corresponding to these eigenvectors are $\{e^{2\pi i\theta_r}\}_r$. A more practical goal is then to estimate phases coherently in superposition, i.e., prepare the state $\sum_r c_r |\omega_r^\delta\rangle |\psi_{\theta_r}\rangle$ with probability at least $1 - \epsilon$ such that $|\omega_r^\delta\rangle = \sum_s d_{r,s} |\bar{\theta}_{r,s}\rangle$ and $\bar{\theta}_{r,s}$ is δ -close to θ_r . Here, $|\bar{\theta}_i\rangle$ is a quantum state that corresponds to the binary fractional representation of $\bar{\theta}_i$. This coherent setting is more realistic 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 post-processing. 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 described by $\sum_r \sum_s |c_r d_{r,s}|^2 |\bar{\theta}_{r,s}\rangle |\psi_{\theta_r}\rangle \langle \bar{\theta}_{r,s}| \langle \psi_{\theta_r}|$.

To the best of our knowledge, only two algorithms exist in the literature that can perform phase estimation coherently [CEMM98, Ral21], with the standard QPE algorithm (also known as the textbook QPE algorithm) being

the most well-known [CEMM98]. 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 $4/\pi^2$ in a single run. Note that since this success probability is only constant, to boost this success probability to the desired $1 - \epsilon$, one way is to use the coherent median technique, where the algorithm is executed $O(\log(1/\epsilon))$ times in parallel and then the median of the outputs is computed coherently [NWZ09]. Therefore, the overall algorithm, i.e., the standard QPE combined with the coherent median computation for boosting success probability, has query complexity of $O(\delta^{-1} \log(1/\epsilon))$, which is optimal as it matches 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 [HNS02, Kla03, BBG+13], 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 $O(\delta^{-1} \epsilon^{-1})$ [CEMM98]. This is because the query complexity of the standard QPE algorithm grows exponentially with the number of these additional qubits used, i.e., $O(\delta^{-1} 2^m)$. 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 possibility of improving the standard QPE algorithm to achieve the optimal query complexity of $O(\delta^{-1} \log(1/\epsilon))$ without employing the median approach and the associated quantum sorting network.

A. Contributions

We propose a modified and 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 most-frequency-concentrated

discrete prolate spheroidal sequence (DPSS), which is a well-known *window/tapering* function used 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 $O((\delta\epsilon)^{-1})$. 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 an explicit algorithm for doing so, which we describe in detail in App. 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.

Similarly to the standard QPE algorithm, the tQPE 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. This can be seen as follows. In Ref. [HH02], 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 $O\left(\left(\log\left(\frac{1}{\delta}\right) + \log\left(\frac{1}{\epsilon}\right)\right) \log\left(\log\left(\frac{1}{\delta}\right) + \log\left(\frac{1}{\epsilon}\right)\right)\right)$ because the number of ancilla qubits required for this algorithm is $p = O\left(\log\left(\frac{1}{\delta}\right) + \log\left(\frac{1}{\epsilon}\right)\right)$. On the other hand, the gate complexity of this transform in the case of tQPE is $O\left(\left(\log\left(\frac{1}{\delta}\right) + \log\log\left(\frac{1}{\epsilon}\right)\right) \log\left(\log\left(\frac{1}{\delta}\right) + \log\log\left(\frac{1}{\epsilon}\right)\right)\right)$ because the ancilla qubits required for this algorithm is $p = O\left(\log\left(\frac{1}{\delta}\right) + \log\log\left(\frac{1}{\epsilon}\right)\right)$, 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 (19). 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.

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. 2, there are three main steps: 1) prepare the ancilla register in a state $|\phi\rangle$, 2) apply the controlled- U^j unitaries to the joint state, $|\phi\rangle|\psi_\theta\rangle$, for some suitable powers, j , and 3) apply the inverse QFT to the ancilla qubits. The standard QPE algorithm can be seen as a special case of tQPE, where the ancilla qubits are initialized to the uniform superposition state, i.e., $|\phi\rangle = 1/2^{p/2} (\sum_i |i\rangle)$. The application of the above three steps leads to the following transformation:

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

where $\hat{\phi}(\cdot)$ is the discrete-time Fourier transform, defined explicitly in (8), of $|\phi\rangle$.

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 goal is to prepare this final state such that the following holds:

$$\sum_{k: |\theta - k/2^p| \leq \delta} |\hat{\phi}(\theta - k/2^p)|^2 \geq 1 - \epsilon. \quad (2)$$

Intuitively, we want to choose a state, $|\phi\rangle$, of the ancilla register whose energy in the frequency domain is concentrated within the frequency band $[-\delta, \delta]$ and is at least $1 - \epsilon$. Fig. 1 shows an example of a taper whose energy 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

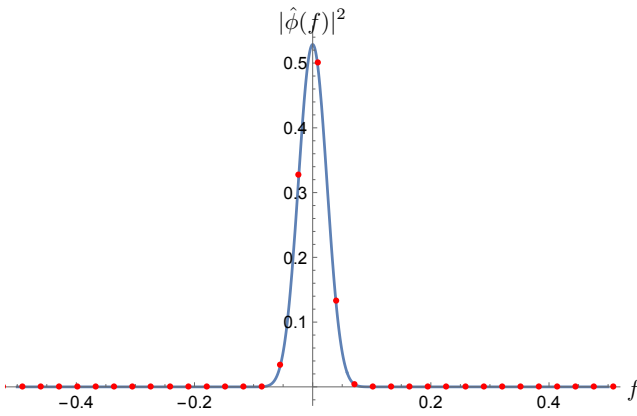


FIG. 1. For the DPSS taper with $p = 5$, we plot the absolute value squared of the taper in the frequency domain (blue curve) as well as the discrete values (red dots) at which it is evaluated for the expression of success probability in (2).

as the optimal solution in the average-case setting. Furthermore, we show that this DPSS also exhibits optimal performance with only a constant overhead in the worst-case setting. Finally, we find that a sinusoidal taper is obtained 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.

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 [Kit95, SHF13, NLY23, LT22, OTT19, WBC22, CBB20, DL23, GSP21, MdW23] are not suitable for the coherent setting that we are considering in this paper. This point is strongly emphasized in Ref. [Ral21]. As such, only two algorithms, such as those presented in Ref. [CEMM98] and Ref. [Ral21], provide coherent phase estimation algorithms and are thus suitable for use as subroutines in larger algorithms. The approach in Ref. [Ral21] 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 to enhance the probability of success.

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 [BDM99]. The authors of [BW00] 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. [RIK22], 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 [HHL09]. Bump functions have found application in spectral estimates based on time series analysis [Som19]. Tapers have also been explored within quantum spectral filtering methods aimed at efficient state initialization [FGML17]. Furthermore, approximate Gaussian tapers have been utilized in quantum algorithms for spectral density estimation [Rog20] and thermal state preparation [CKBG23, CKG23]. The Kaiser taper [CF⁺48, Kai66, KRD21, MGB22] has been considered in QPE [BSG⁺22] and also achieves asymptotically optimal scaling. We will further comment on this work and Kaiser tapers relation to the DPSS taper in Sec. V. However, none of these tapers are optimally frequency-concentrated, resulting in side-lobes that warrant further reduction.

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 this paper to refer to this taper.

In a series of seminal papers in the field of signal analysis [SP61, LP61, LP62, Sle64, Sle78], 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 non-zero 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 [Sle78]. In this paper, the authors 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. (They additionally showed how to stably compute it numerically.) For completeness, we provide a detailed derivation of this in App. E. Due to its ability to concentrate the energy of the Fourier transform in a small interval in the frequency domain, while also maintaining a finite support in the time domain, this maximally concentrated taper 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 [Sle78, Tho82, PLVI87, MP99, HTR09, SY12]. 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

Consider a unitary, U , acting on a d -dimensional Hilbert space, $\mathcal{H}$. Let $\{|\psi_{\theta_r}\rangle\}_{r=0}^{d-1}$ be a set of orthogonal eigenvectors of U , and $\{e^{2\pi i\theta_r}\}_{r=0}^{d-1}$ be the set of corresponding eigenvalues, where $\theta_r \in [0, 1)$ for all $r \in \{0, \dots, d-1\}$. It is assumed that we are given query access to U and controlled- U , as well as sample access to an arbitrary state, $|\psi\rangle = \sum_r c_r |\psi_{\theta_r}\rangle$, or alternatively, query access to the unitary that prepares $|\psi\rangle$. Given $\delta, \epsilon > 0$, the goal of coherent phase estimation is then to prepare the state $\sum_r c_r |\omega_r^\delta\rangle |\psi_{\theta_r}\rangle$ with probability at least $1 - \epsilon$ such that $|\omega_r^\delta\rangle = \sum_s d_{r,s} |\tilde{\theta}_{r,s}\rangle$ and $\tilde{\theta}_{r,s}$ is δ -close to θ_r .

In what follows, we focus on a particular eigenvector $|\psi_\theta\rangle$ of U with eigenvalue $e^{2\pi i\theta}$ for simplicity 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. 2, we prepare an ancilla register in a state $|\phi\rangle$ defined as

$$|\phi\rangle := \sum_{n=0}^{N-1} \phi[n] |n\rangle, \quad (3)$$

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

$$\sum_{n=0}^{N-1} |n\rangle\langle n| \otimes U^n \quad (4)$$

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

$$\sum_{n=0}^{N-1} \phi[n] e^{2\pi i\theta n} |n\rangle |\psi_\theta\rangle. \quad (5)$$

Subsequently, we apply the inverse QFT ($\text{QFT}^{-1} = \text{QFT}^\dagger$) on the ancilla register. This transforms the basis

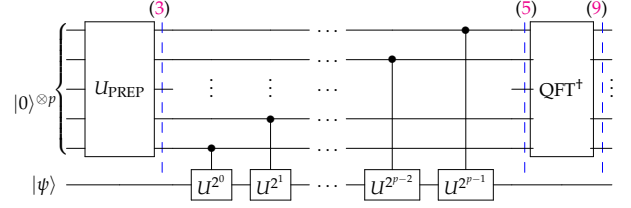


FIG. 2. Tapered QPE quantum circuit. The system is initialized in the state $|0\rangle^{\otimes p} |\psi\rangle$, where $|\psi\rangle$ is an arbitrary input state. U_{PREP} is the unitary that prepares the taper state $|\phi\rangle$ (see App. A for U_{PREP} that prepares the DPSS taper approximately). Dashed blue lines labeled by equation numbers denote the corresponding state at each step.

$\{|n\rangle\}_{n=0}^{N-1}$ as

$$|n\rangle \rightarrow \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} e^{-2\pi i n k / N} |k\rangle, \quad (6)$$

giving us the following final state:

$$\sum_{k=0}^{N-1} \left(\frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \phi[n] e^{2\pi i n (\theta - k/N)} \right) |k\rangle |\psi_\theta\rangle. \quad (7)$$

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

$$\hat{\phi}(f) := \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \phi[n] e^{2\pi i n f}, \quad (8)$$

of the time-limited signal, $\phi[n]$, evaluated at the frequency, $\theta - k/N$. Using this, the final state of the algorithm, given by (7), can be expressed more concisely as:

$$\sum_{k=0}^{N-1} \hat{\phi}(\theta - k/N) |k\rangle |\psi_\theta\rangle, \quad (9)$$

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 $|\hat{\phi}(\theta - k/N)|^2$.

Recall that the goal of tQPE is to prepare the final state, given by (9), such that the following holds:

$$\sum_{k: |\theta - k/N| \leq \delta} |\hat{\phi}(\theta - k/N)|^2 \geq 1 - \epsilon. \quad (10)$$

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 phase estimation success probability), the tQPE algorithm using the tophat taper outputs a δ -close phase estimate with probability at least $4/\pi^2 \approx 0.405$. 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. [CEMM98] 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 reduce the probability of outputting a phase estimate that differs from the true value by more than δ , it also increases the computational cost of the algorithm, since the number of queries to U is proportional to $2^p = 2^{\ell+m}$ (as shown in Fig. 2). It also increases the size of the inverse-QFT circuit because as mentioned above, the size of this circuit is of the order $O(p \log p)$. Thus, it is of utmost importance to choose our taper, $|\phi\rangle$, optimally so as to minimize m for a given ϵ .

IV. OPTIMAL TAPERS

In what follows, we derive optimal tapers for the case where the relative position of the true phase with respect to the grid of estimates is known and optimal tapers for the average-case settings which we define later. We then study the worst-case error of these tapers.

A. The Optimization Problem

We now state the optimization problem more formally:

$$\begin{aligned} \max_{|\phi\rangle \in \mathcal{H}_t} \quad & \sum_{k: |\theta - \frac{k}{N}| \leq \delta} \left| \hat{\phi} \left(\theta - \frac{k}{N} \right) \right|^2 \\ \text{subject to} \quad & \sum_{n=0}^{N-1} |\phi[n]|^2 = 1. \end{aligned} \quad (11)$$

Here, $\mathcal{H}_t$ denotes the N -dimensional Hilbert space corresponding to the taper register. The objective function above represents the probability of obtaining a phase estimate that is δ -close to the exact value, θ . 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 exactly 2^m estimates that are δ -close to θ ¹. Depending on the exact value of θ , the index, k , in the objective function of (11) can range over different values. This significantly complicates the optimization problem. In order to simplify the above expression further, we define the following:

$$\Delta := \theta - \frac{k^*}{N}, \quad (12)$$

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 $\hat{\phi}(\Delta - k/N)$ 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 leaves out at most one phase estimate that is furthest away from the true value (two if $\Delta = \pm 1/2N$). Finally, all tapers of practical interest we know of have Fourier transforms that are symmetric, have a peak at 0, and decay rapidly, meaning their value at $\Delta \pm (K+1)$ 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 as:

$$\begin{aligned} \max_{|\phi\rangle \in \mathcal{H}_t} \quad & \sum_{k=-K}^K \left| \hat{\phi} \left(\Delta - \frac{k}{N} \right) \right|^2 \\ \text{subject to} \quad & \sum_{n=0}^{N-1} |\phi[n]|^2 = 1. \end{aligned} \quad (13)$$

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

$$\begin{aligned} \mathcal{L}(|\phi\rangle, \lambda) = \quad & \sum_{j=-K}^K \left| \hat{\phi} \left(\Delta - \frac{j}{N} \right) \right|^2 \\ & - \lambda \left(\sum_{n=0}^{N-1} |\phi[n]|^2 - 1 \right). \end{aligned} \quad (14)$$

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

B. Ideal Case

Finding the optimal taper is equivalent to finding the stationary point of $\mathcal{L}$ that maximizes the objective function of (13). The stationary points of $\mathcal{L}$ can be found by

¹ There are two exceptions: when θ is exactly on the grid there are $2^m + 1$ δ -close estimates for $m > 0$, and when θ is exactly between two grid points there are two δ -close estimates for $m = 0$.

setting all the partial derivatives of $\mathcal{L}$ to zero. Doing so, we get the following two conditions (see App. B for a detailed derivation):

$$\frac{1}{N} \sum_{n=0}^{N-1} e^{2\pi i \Delta(n-m)} \left(\frac{\sin(\pi(m-n)(2K+1)/N)}{\sin(\pi(m-n)/N)} \right) \phi[n] = \lambda \phi[m]; \quad \forall m \in \{0, \dots, N-1\}, \quad (15)$$

$$\sum_{n=0}^{N-1} |\phi[n]|^2 = 1. \quad (16)$$

The first equation is an eigenvalue equation, while the second equation is the normalization constraint. By substituting all the stationary points $(|\phi\rangle, \lambda)$ satisfying the above two conditions into the objective function of (13), we obtain:

$$\sum_{j=-K}^K \left| \hat{\phi} \left(\Delta + \frac{j}{N} \right) \right|^2 = \lambda. \quad (17)$$

This implies that the stationary point, $(|\phi\rangle, \lambda)$, that maximizes the objective function is the eigenvector with maximum eigenvalue, λ . Since the objective function is the probability of outputting one of the $2K+1$ phase estimates closest to θ , using the eigenvector with the maximum eigenvalue (see (15)) as our taper will maximize this probability.

In App. C, we establish an explicit connection between the eigenvectors of (15) and the periodic discrete prolate spheroidal sequences (P-DPSS) [ZKD⁺17]. We call the former quantum periodic discrete prolate spheroidal sequences (QP-DPSS), which depend on the value of Δ . Then, combining results from App. C with known results for the P-DPSS in [XC84], we observe that the eigenvector of (15) 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 (12)). 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. Now, let's consider the case where $0 < \Delta \leq \frac{1}{2N}$ or $-\frac{1}{2N} \leq \Delta < 0$. This case can be converted into the $\Delta = 0$ case by shifting the grid of possible phase estimates by Δ . To accomplish this, we apply a unitary operator parameterized by Δ to the tophat taper. This operation shifts all 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.

We find that even for $K = 0$, unit success probability can be achieved with the optimal taper given by:

$$\phi^\Delta[n] = \frac{e^{2\pi i \Delta n}}{\sqrt{N}}. \quad (18)$$

As a check, we note that for $\Delta = 0$, we recover the tophat taper which is known to have zero error probability 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 App. 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. Eq. (18) 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 tapers: one shown in Eq. (18) 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

$$\phi^{\sin}[n] = \frac{\sin(\pi n/N)}{\sqrt{N/2}}, \quad (19)$$

$$\phi^{\cos}[n] = \frac{\cos(\pi n/N)}{\sqrt{N/2}}. \quad (20)$$

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 coherent phase estimation, 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

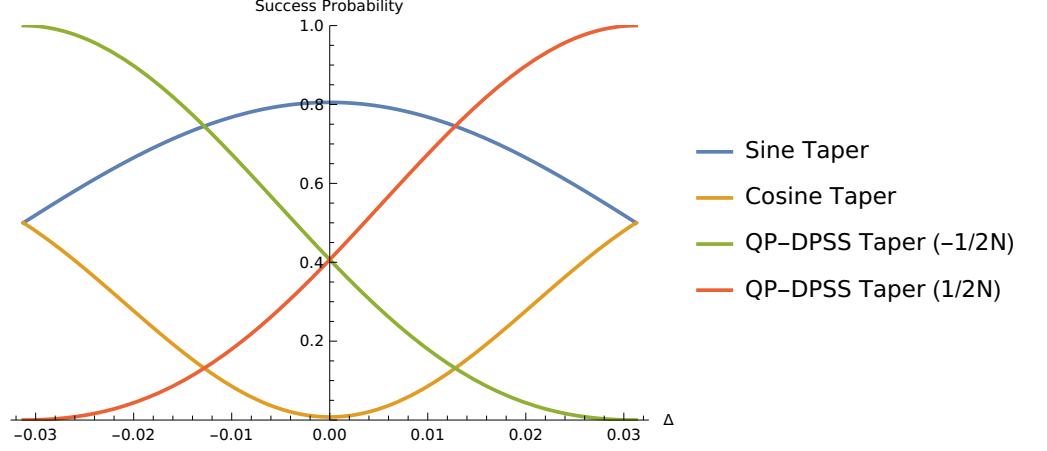


FIG. 3. We plot the probability of the $\phi^{-1/2N}$ (green), $\phi^{1/2N}$ (red), $\phi^{\sin}$ (blue), and $\phi^{\cos}$ (orange) tapers to output the closest phase estimate as a function of Δ , for $N = 2^5$. Both cosine and sine tapers achieve 0.5 at $\Delta = \pm 1/2N$. This means with unit probability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

that works best on average. The Lagrangian, $\mathcal{L}_{\text{avg}}$, for this modified optimization problem is similar to the one given by (14), except that the first term is now an expectation, resulting in the average success probability:

$$\mathcal{L}_{\text{avg}}(|\phi\rangle, \lambda) = \mathbb{E}_{\Delta \sim \mathcal{D}} \left[\sum_{j=-K}^K \left| \hat{\phi} \left(\Delta + \frac{j}{N} \right) \right|^2 \right] - \lambda \left(\sum_{n=0}^{N-1} |\phi[n]|^2 - 1 \right), \quad (21)$$

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

$$\frac{1}{N} \sum_{n=0}^{N-1} \frac{\sin(\pi(m-n)(2K+1)/N)}{\pi(m-n)} \phi[n] = \lambda \phi[m]; \quad \forall m \in \{0, \dots, N-1\}. \quad (22)$$

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, $\mathcal{L}_{\text{avg}}$ (i.e., average success probability), is the one with maximum eigenvalue. Thus, the optimal eigenvector is the DPSS with the maximum eigenvalue. Note that when we mention DPSS from here on, we refer to the DPSS with the maximum eigenvalue. For completeness, we provide a brief overview of DPSSs from the classical signal analysis point of view in App. E.

Remark 1. An important thing to note is that the derivation of DPSSs from the classical signal processing point of view (provided in App. 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 [KRD21] that provides (to date) the most stringent non-asymptotic bounds on the eigenvalues of the kernel in (22). The key idea is to use this theorem to derive a non-asymptotic bound on the number of ancilla qubits, i.e., m , needed to achieve the average success probability of at least $1 - \epsilon$, for some $\epsilon > 0$.

Theorem 1 ([KRD21, Corollary 1]). *For any $N, K \in \mathbb{N}$ such that $K \in [0, \dots, N/2 - 1]$, the maximum eigenvalue λ_{max} of the kernel in (22) satisfies*

$$\lambda_{\text{max}} \geq 1 - \min \left\{ 8 \exp \left[-\frac{2K-1}{\frac{2}{\pi^2} \log(4N)} \right], 10 \exp \left[-\frac{2K-6}{\frac{2}{\pi^2} \log(100K+75)} \right] \right\}. \quad (23)$$

As the maximum eigenvalue of the DPSS kernel in (22) is equal to the average success probability, the bound on this eigenvalue from the theorem given above naturally bounds the average success probability. Therefore, one way to infer the above theorem statement is as follows: when the maximal DPSS is used as a taper, it 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 (23).

Applying Thm. 1 to 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 non-asymptotic regime.

Theorem 2 (Non-asymptotic). *To ensure that the output of the tQPE algorithm is one of the $2K + 1$ phase estimates closest to the true phase θ with an average success probability of at least $1 - \epsilon$, we only require*

$$m = \left\lceil \log_2 \left(\left\lceil 175 (\log(10/\epsilon) + 1)^2 \right\rceil + 1 \right) \right\rceil + 1 \quad (24)$$

additional qubits when $K \geq 1$.

Proof. We use the bound from Thm. 1 to lower-bound the average success probability for outputting one of the $2K + 1$ p -bit phase estimates closest to θ . Subsequently, if this bound is at least $1 - \epsilon$, we can say that the average success probability is also at least $1 - \epsilon$. Formally, this is expressed as follows:

$$\epsilon \geq \min \left\{ 8 \exp \left[-\frac{2K-1}{\frac{2}{\pi^2} \log(4N)} \right], 10 \exp \left[-\frac{2K-6}{\frac{2}{\pi^2} \log(100K+75)} \right] \right\}. \quad (25)$$

Without loss of generality, we choose to work with the second argument of the min function in the equation above, i.e.,

$$\epsilon \geq 10 \exp \left[-\frac{2K-6}{\frac{2}{\pi^2} \log(100K+75)} \right]. \quad (26)$$

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

$$\log(10/\epsilon) \leq \frac{2K-6}{\frac{2}{\pi^2} \log(100K+75)}. \quad (27)$$

By assuming that $K \geq 1$, we first provide a lower bound for $\frac{2K-6}{\frac{2}{\pi^2} \log(100K+75)}$ in the following way:

$$\frac{2K-6}{\frac{2}{\pi^2} \log(100K+75)} \geq \frac{2K-6}{\frac{2}{\pi^2} \log(175K)} \quad (28)$$

$$\geq \frac{K-3}{\log(175K)} \quad (29)$$

$$= \frac{1}{175} \left(\frac{175K}{\log(175K)} \right) - \frac{3}{\log(175K)} \quad (30)$$

$$\geq \frac{1}{175} \left(\frac{175K}{\log(175K+1)} \right) - 1 \quad (31)$$

$$\geq \frac{\sqrt{175K}}{175} - 1 \quad (32)$$

$$= \sqrt{\frac{K}{175}} - 1, \quad (33)$$

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

$$\frac{x}{\log(x+1)} \geq \sqrt{x+1} \text{ for } x \geq -1. \quad (34)$$

Now, by enforcing the following inequality,

$$\log(10/\epsilon) \leq \sqrt{\frac{K}{175}} - 1, \quad (35)$$

we obtain $K \geq 175 (\log(10/\epsilon) + 1)^2$. In other words, as long as K is at least $\left\lceil 175 (\log(10/\epsilon) + 1)^2 \right\rceil$, we are guaranteed that (26) holds and the error is bounded by ϵ . Recall from above that 2^m of the p -bit phase estimates are δ -close. Letting $K = 2^{m-1} - 1$, we only need $m = \left\lceil \log_2 \left(\left\lceil 175 (\log(10/\epsilon) + 1)^2 \right\rceil + 1 \right) \right\rceil + 1$ additional qubits to ensure that our tQPE algorithm has at least $1 - \epsilon$ success probability. $\square$

Remark 2. Although the analytical proof for Thm. 2 implies that K needs to be of order $\log^2(1/\epsilon)$ to achieve an average success probability of $1 - \epsilon$, for all practical purposes K only needs to be of order $\log(1/\epsilon)$. For example, for all $\epsilon \geq 10^{-81}$, we have that $K \leq 192$, and as a result, we can replace K inside the logarithm in (27) with 192 and find that $K = \lceil \log(10/\epsilon) \rceil + 3$ is sufficient.

Since Thm. 1 holds for any $N \in \mathbb{N}$, our result described in Thm. 2 also holds for any N including the asymptotic regime where $N \rightarrow \infty$ and $\delta \rightarrow 0$. However, the bound on the maximum eigenvalue shown in Thm. 1 can be made tighter when we restrict our analysis to the asymptotic case with large N and small δ . In what follows, we state the result for this asymptotic case.

Theorem 3 (Asymptotic). *In the regime where N is large and δ is small, the tQPE algorithm with the DPSS taper outputs one of the $2K + 1$ phase estimates closest to the true phase θ with an average success probability of at least $1 - \epsilon$ when*

$$m = \left\lceil \log_2 \log \left(\frac{1}{\epsilon} \right) \right\rceil \quad (36)$$

additional qubits are used in the tQPE algorithm.

Proof. Let $W = 2\delta = \frac{1}{2^l}$. Then, $NW = 2^m$ because we know that $N = 2^{l+m}$. In the original paper on DPSS [Sle78], Slepian provided an asymptotic expression for the eigenvalues $\{\lambda_k\}_k$ of the DPSSs (i.e., when N is large or $\delta, W \rightarrow 0$),

$$1 - \lambda_k = \pi^{1/2} (k!)^{-1} 2^{(14k+9)/4} \times \alpha^{(2k+1)/4} (2 - \alpha)^{-(k+1/2)} N^{k+1/2} e^{-\gamma N}, \quad (37)$$

where α and γ are defined as follows:

$$\alpha = 1 - \cos 2\pi W \quad (38)$$

$$\gamma = \log \left(1 + \frac{2\sqrt{\alpha}}{\sqrt{2} - \sqrt{\alpha}} \right). \quad (39)$$

From the definition of α , it is clear that $\alpha \rightarrow 0$ as $W \rightarrow 0$. To this end, we Taylor expand the cos function, to get the following expression of α :

$$\alpha = \frac{(2\pi W)^2}{2} + O((2\pi W)^4). \quad (40)$$

Substituting this in the definition of γ , we get

$$\gamma = \log \left(1 + \frac{2\sqrt{\frac{(2\pi W)^2}{2} + O((2\pi W)^4)}}{\sqrt{2} - \sqrt{\frac{(2\pi W)^2}{2} + O((2\pi W)^4)}} \right) \quad (41)$$

$$\sim \log(1 + 2\pi W) \quad (42)$$

$$\sim 2\pi W. \quad (43)$$

Also, $\alpha \sim 2\pi^2 W^2$. Plugging the values of α , γ , and also $k = 0$ (because we are interested in the maximum eigenvalue) into the first equation, we get

$$\begin{aligned} 1 - \lambda_0 &\sim \pi^{1/2} 2^{9/4} (2\pi^2 W^2)^{1/4} \underbrace{(2 - 2\pi^2 W^2)^{-1/2}}_{\sim 2^{-1/2}} N^{1/2} e^{-2\pi W N} \\ &= \pi^{1/2} 2^{9/4} (2\pi^2 W^2)^{1/4} 2^{-1/2} N^{1/2} e^{-2\pi W N} \end{aligned} \quad (44)$$

$$= 4\pi(NW)^{1/2} e^{-2\pi NW}. \quad (45)$$

$$= 4\pi(NW)^{1/2} e^{-2\pi NW}. \quad (46)$$

We require that $1 - \lambda_0 \leq \epsilon$. This implies

$$4\pi(NW)^{1/2} e^{-2\pi NW} \leq \epsilon \quad (47)$$

Using the fact that $4\pi x^{1/2} e^{-2\pi x} \leq e^{-x}$ for $x \geq 1$, it suffices to choose

$$m = \lceil \log_2 \log(1/\epsilon) \rceil \quad (48)$$

Please note that this result can be further tightened, but it is sufficient for our purpose. $\square$

Remark 3. Now, using the fact that $K = 2^{m-1} - 1$, we further obtain:

$$K = 2^{\lceil \log_2 \log(1/\epsilon) \rceil - 1} - 1 \leq \log(1/\epsilon) - 1. \quad (49)$$

While we do not gain an improvement in m with respect to ϵ in the asymptotic regime, we observe an improvement in the scaling of K with respect to ϵ in this regime. In particular, we showed how the upper bound on K decreases from $175(\log(10/\epsilon) + 1)^2 + 1$ to $\log(1/\epsilon) - 1$ when we utilize the asymptotic bound that Slepian provides in [Sle78] which holds in the regime where N is large and δ is small. Recall that the tQPE algorithm succeeds in outputting one of the $2K + 1$ phase estimates closest to the true phase θ with an average success probability of at least $1 - \epsilon$. Thus, a smaller K value would imply that the DPSS taper has greater frequency concentration around θ and endows a smaller set of closest phase estimates with the same success probability of $1 - \epsilon$.

D. Analysis of worst-case error

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 the largest error probability. However, in general, different tapers have different worst-case Δ^* , which makes the corresponding optimization problem too hard. 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 (22):

$$\begin{aligned} \frac{1}{N} \sum_{n=0}^{N-1} \frac{\sin(\pi(m-n)(2K+1)/N)}{\pi(m-n)} \phi[n] \\ = \lambda \phi[m]; \quad \forall m \in \{0, \dots, N-1\}. \end{aligned} \quad (50)$$

Recall that the DPSS taper is the eigensequence with the largest eigenvalue that satisfies the DPSS kernel. 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 App. 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:

$$\begin{aligned} \sum_{j=-K}^K \left| \hat{\phi} \left(\Delta + \frac{j}{N} \right) \right|^2 \\ = \frac{1}{N} \sum_{n,m=0}^{N-1} e^{2\pi i(m-n)\Delta} \frac{\sin \left(\pi(m-n) \left(\frac{2K+1}{N} \right) \right)}{\sin(\pi(m-n)/N)} \phi^*[n] \phi[m] \end{aligned} \quad (51)$$

$$\begin{aligned} = \frac{1}{N} \sum_{n,m=0}^{N-1} \cos(\pi(m-n)/N) \\ \frac{\sin \left(\pi(m-n) \left(\frac{2K+1}{N} \right) \right)}{\sin(\pi(m-n)/N)} \phi^*[n] \phi[m] \end{aligned} \quad (52)$$

where the last equality comes from substituting $\Delta = \frac{1}{2N}$ 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 (52).

From Thm. 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 (51) 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. Therefore, DPSS is optimal asymptotically. While the DPSS taper may not achieve the same $1 - \epsilon$ success probability as it would in the average case for the non-asymptotic case, i.e., when N is finite, the fact that the worst-case DPSS success probability converges to the average-case DPSS success probability for large values of N implies that the worst-case success probability of the DPSS taper would share a similar dependence on ϵ . We provide numerical evidence in Figs. 4a, 4b, and 4c that shows that the worst-case success probability of the DPSS taper never falls below $1 - 4\epsilon$ for several different parameters.

E. Numerics

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. 5a, 5b, 5c, and 5d, we plot the success probability of each taper as a function of 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 Figs. 4a, 4b, and 4c, we again plot the success probability of the DPSS taper as a function of Δ for various values of m . In contrast to Figs. 5a, 5b, 5c, and 5d, we do not plot the other tapers; instead we display a horizontal line at $1 - 4\epsilon_{\text{ave}}$, where ϵ_{ave} is the average error associated with the DPSS taper (alternatively, $1 - \epsilon_{\text{ave}}$ is the largest eigenvalue of the DPSS kernel). We see that the success probability is above $1 - 4\epsilon_{\text{ave}}$ for all Δ . Thus, our numerics suggest that the DPSS taper has a worst-case error that is at most four times its average-case error. Since the worst-case error cannot be smaller than the average-case, the worst-case error of DPSS is at most a factor of four worse than the best possible worst-case error. This means that although we have not been able to find the optimal taper for the worst-case, the DPSS taper cannot be too far from it.

V. DISCUSSION

We note that the authors of Ref. [RDD17] 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 K 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 [BSG⁺22]. 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. [BSG⁺22]. As shown in the proof for Thm. 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 DPSS.

In addition to the asymptotic analysis, we also provide the explicit expressions for m in the non-asymptotic 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 [MdW23], 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 [MdW23] 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 [LP96, BDM99]: 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.

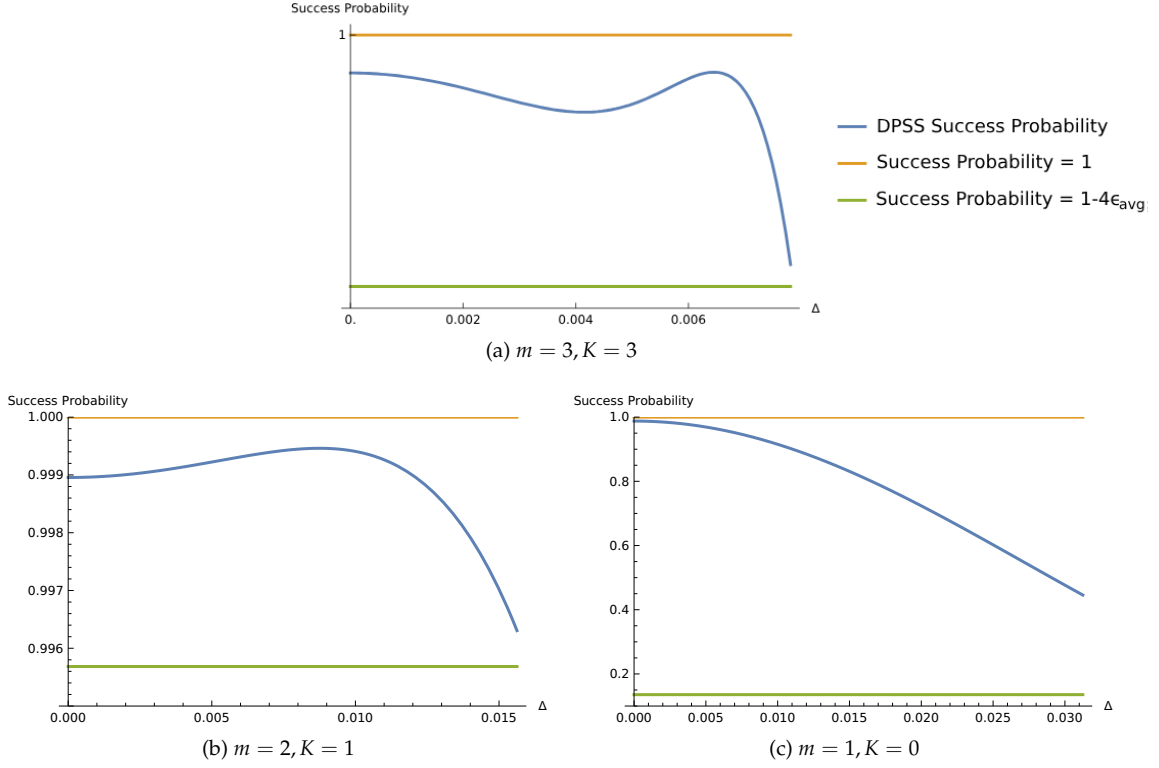


FIG. 4. Performance of the DPSS taper as a function of Δ from 0 to $1/2N$. (a) The figure is plotted for success probabilities ranging from 0.999999975 to 1. We have omitted labeling the tick with 0.999999975 to avoid cluttering the figure.

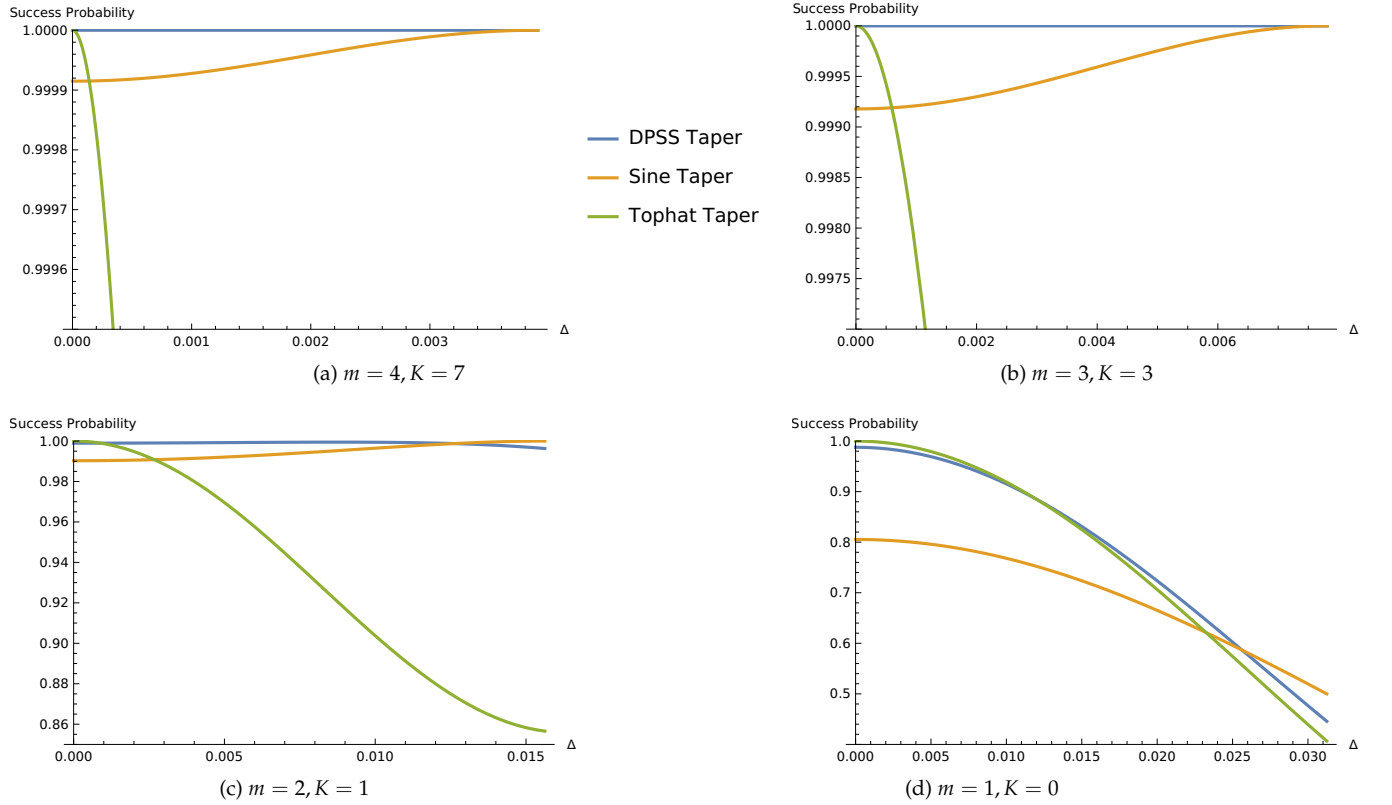


FIG. 5. Comparison between the DPSS taper, the sine taper, and the tophat taper.

It might be possible to construct tapers with slightly worse average case error but better worst case error by hybridizing some of the eigenvectors of the DPSS kernel with large eigenvalues. This is similar to how we constructed the sine taper by hybridizing two QP-DPSS tapers. Our numerics suggests that the worst case error of DPSS taper is at most $4\epsilon_{\text{ave}}$. It is an open question how close the worst case error can be made by hybridizing DPSS's.

VI. ACKNOWLEDGEMENTS

DP and SJST were participants in the 2023 Quantum Computing Summer School at LANL, sponsored by the LANL Information Science & Technology Insti-

tute. DP 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 number 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). YS was funded by the QSC to perform the analytical and numerical analysis and to write the manuscript along with the other authors. ATS 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.

-
- [ALL⁺21] 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.
- [BBG⁺13] Robert Beals, Stephen Brierley, Oliver Gray, Aram W Harrow, Samuel Kutin, Noah Linden, Dan Shepherd, and Mark Stather. Efficient distributed quantum computing. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 469(2153):20120686, 2013.
- [BDM99] Vladimir Bužek, Radoslav Derka, and Serge Massar. Optimal quantum clocks. *Physical review letters*, 82(10):2207, 1999.
- [BHMT02] Gilles Brassard, Peter Høyer, Michele Mosca, and Alain Tapp. Quantum amplitude amplification and estimation. *Contemporary Mathematics*, 305:53–74, 2002.
- [BSG⁺22] Dominic W Berry, Yuan Su, Casper Gyurik, Robbie King, Joao Basso, Alexander Del Toro Barba, Abhishek Rajput, Nathan Wiebe, Vedran Dunjko, and Ryan Babush. Quantifying quantum advantage in topological data analysis. *arXiv preprint arXiv:2209.13581*, 2022.
- [BW00] Dominic W Berry and Howard M Wiseman. Optimal states and almost optimal adaptive measurements for quantum interferometry. *Physical review letters*, 85(24):5098, 2000.
- [CBB20] François Chapeau-Blondeau and Etienne Belin. Fourier-transform quantum phase estimation with quantum phase noise. *Signal Processing*, 170:107441, May 2020.
- [CEMM98] Richard Cleve, Artur Ekert, Chiara Macchiavello, and Michele Mosca. Quantum algorithms revisited. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1969):339–354, 1998.
- [CF⁺48] George Ashley Campbell, Ronald M Foster, et al. *Fourier Integrals for Practical Applications*. D. van Nostrand, NY, 1948.
- [CKBG23] Chi-Fang Chen, MJ Kastoryano, FGSL Brandao, and A Gilyén. Quantum thermal state preparation. *arXiv preprint arXiv:2303.18224*, 10, 2023.
- [CKG23] Chi-Fang Chen, Michael J Kastoryano, and Andrés Gilyén. An efficient and exact noncommutative quantum gibbs sampler. *arXiv preprint arXiv:2311.09207*, 2023.
- [Cop94] D Coppersmith. An approximate fourier transform useful in quantum factoring. *IBM Research Report*, pages RC-19642, 1994.
- [CSS18] Anirban Narayan Chowdhury, Yiğit Subaşı, and Rolando D Somma. Improved implementation of reflection operators. *arXiv preprint arXiv:1803.02466*, 2018.
- [DL23] 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(2):020331, May 2023.
- [ESP21] Alexander Engel, Graeme Smith, and Scott E Parker. Linear embedding of nonlinear dynamical systems and prospects for efficient quantum algorithms. *Physics of Plasmas*, 28(6), 2021.
- [FGML17] François Fillion-Gourdeau, Steve MacLean, and Raymond Laflamme. Efficient state initialization by a quantum spectral filtering algorithm. *Physical Review A*, 95(4):042331, 2017.
- [GSP21] Valentin Gebhart, Augusto Smerzi, and Luca Pezzè. Bayesian quantum multiphase estimation algorithm. *Physical Review Applied*, 16(1):014035, 2021.
- [HH02] L. Hales and S. Hallgren. An improved quantum fourier transform algorithm and applications. In *Proceedings 41st Annual Symposium on Foundations of Computer Science*, SFCS-00. IEEE Comput. Soc, August 2002.
- [HHL09] Aram W Harrow, Avinatan Hassidim, and Seth Lloyd. Quantum algorithm for linear systems of equations. *Physical Review Letters*, 103(15):150502, 2009.
- [HNS02] Hoyer, Neerbek, and Shi. Quantum complexities of ordered searching, sorting, and element distinctness. *Algorithmica*, 34(4):429–448, 2002.
- [HTR09] Simon Haykin, David J Thomson, and Jeffrey H Reed. Spectrum sensing for cognitive radio. *Proceedings of the IEEE*, 97(5):849–877, 2009.
- [Kai66] James F Kaiser. Digital filters. In *System Analysis by Digital Computer*, pages 218–285. Wiley New York, NY, 1966.
- [Kit95] A Yu Kitaev. Quantum measurements and the Abelian stabilizer problem. *arXiv preprint quant-ph/9511026*, 1995.
- [Kla03] Hartmut Klauck. Quantum time-space tradeoffs for sorting. In *Proceedings of the thirty-fifth annual ACM sym-*

- posium on Theory of computing, pages 69–76, 2003.
- [KRD21] Santhosh Karnik, Justin Romberg, and Mark A Davenport. Improved bounds for the eigenvalues of prolate spheroidal wave functions and discrete prolate spheroidal sequences. *Applied and Computational Harmonic Analysis*, 55:97–128, 2021.
- [LMR13] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost. Quantum algorithms for supervised and unsupervised machine learning. *arXiv preprint arXiv:1307.0411*, 2013.
- [LMR14] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost. Quantum principal component analysis. *Nature Physics*, 10(9):631–633, 2014.
- [LP61] Henry J Landau and Henry O Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty—II. *Bell System Technical Journal*, 40(1):65–84, 1961.
- [LP62] 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 System Technical Journal*, 41(4):1295–1336, 1962.
- [LP96] A Luis and J Peřina. Optimum phase-shift estimation and the quantum description of the phase difference. *Physical review A*, 54(5):4564, 1996.
- [LT22] Lin Lin and Yu Tong. Heisenberg-limited ground-state energy estimation for early fault-tolerant quantum computers. *PRX Quantum*, 3(1):010318, February 2022.
- [MdW23] Nikhil S Mande and Ronald de Wolf. Tight bounds for quantum phase estimation and related problems. *arXiv preprint arXiv:2305.04908*, 2023.
- [MGB22] Sam McArdle, András Gilyén, and Mario Berta. Quantum state preparation without coherent arithmetic. *arXiv preprint arXiv:2210.14892*, 2022.
- [MP99] Partha P Mitra and Bijan Pesaran. Analysis of dynamic brain imaging data. *Biophysical Journal*, 76(2):691–708, 1999.
- [NLY23] Hongkang Ni, Haoya Li, and Lexing Ying. On low-depth algorithms for quantum phase estimation. *Quantum*, 7:1165, November 2023.
- [NWZ09] Daniel Nagaj, Paweł Wocjan, and Yong Zhang. Fast amplification of qma. *Quantum Information & Computation*, 9(11):1053–1068, 2009.
- [OTT19] Thomas E O’Brien, Brian Tarasinski, and Barbara M Terhal. Quantum phase estimation of multiple eigenvalues for small-scale (noisy) experiments. *New Journal of Physics*, 21(2):023022, February 2019.
- [PLVI87] Jeffrey Park, Craig R Lindberg, and Frank L Vernon III. Multitaper spectral analysis of high-frequency seismograms. *Journal of Geophysical Research: Solid Earth*, 92(B12):12675–12684, 1987.
- [Ral20] Patrick Rall. Quantum algorithms for estimating physical quantities using block encodings. *Physical Review A*, 102(2):022408, 2020.
- [Ral21] Patrick Rall. Faster coherent quantum algorithms for phase, energy, and amplitude estimation. *Quantum*, 5:566, October 2021.
- [RDD17] Wojciech Rządowski and Rafał Demkowicz-Dobrzański. Discrete-to-continuous transition in quantum phase estimation. *Physical Review A*, 96(3):032319, 2017.
- [RIK22] Gumaro Rendon, Taku Izubuchi, and Yuta Kikuchi. Effects of cosine tapering window on quantum phase estimation. *Physical Review D*, 106(3):034503, 2022.
- [Rog20] Alessandro Roggero. Spectral-density estimation with the Gaussian integral transform. *Physical Review A*, 102(2):022409, 2020.
- [SBM05] Vivek V Shende, Stephen S Bullock, and Igor L Markov. Synthesis of quantum logic circuits. In *Proceedings of the 2005 Asia and South Pacific Design Automation Conference*, pages 272–275, 2005.
- [SHF13] Krysta M Svore, Matthew B Hastings, and Michael Freedman. Faster phase estimation. *arXiv preprint arXiv:1304.0741*, 2013.
- [Sho94] Peter W Shor. Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings 35th Annual Symposium on Foundations of Computer Science*, pages 124–134. IEEE, 1994.
- [Sle64] David Slepian. Prolate spheroidal wave functions, Fourier analysis and uncertainty—IV: extensions to many dimensions; generalized prolate spheroidal functions. *Bell System Technical Journal*, 43(6):3009–3057, 1964.
- [Sle78] David Slepian. Prolate spheroidal wave functions, Fourier analysis, and uncertainty—V: The discrete case. *Bell System Technical Journal*, 57(5):1371–1430, 1978.
- [Som19] Rolando D Somma. Quantum eigenvalue estimation via time series analysis. *New Journal of Physics*, 21(12):123025, 2019.
- [SP61] David Slepian and Henry O Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty—I. *Bell System Technical Journal*, 40(1):43–63, 1961.
- [SY12] Andrew Sornborger and Takeshi Yokoo. A multivariate, multitaper approach to detecting and estimating harmonic response in cortical optical imaging data. *Journal of Neuroscience Methods*, 203(1):254–263, 2012.
- [Tho82] David J Thomson. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9):1055–1096, 1982.
- [WBC22] Kianna Wan, Mario Berta, and Earl T. Campbell. Randomized quantum algorithm for statistical phase estimation. *Physical Review Letters*, 129(3):030503, July 2022.
- [WBD⁺21] Lewis Wright, Fergus Barratt, James Dborin, George H Booth, and Andrew G Green. Automatic post-selection by ancillae thermalization. *Physical Review Research*, 3(3):033151, 2021.
- [WBL12] Nathan Wiebe, Daniel Braun, and Seth Lloyd. Quantum algorithm for data fitting. *Physical Review Letters*, 109(5):050505, 2012.
- [XC84] Wen Yuan Xu and Christodoulos Chamzas. On the periodic discrete prolate spheroidal sequences. *SIAM Journal on Applied Mathematics*, 44(6):1210–1217, 1984.
- [ZKD⁺17] 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 Processing Letters*, 25(1):95–99, 2017.

Supplementary Material for “Optimal Coherent Quantum Phase Estimation via Tapering”

Appendix A: Preparation of the DPSS Taper

In this section, we focus on preparing the DPSS taper $|\phi\rangle$. We use a modification of the method for preparing broadband Gaussian states presented in [CSS18] to prepare the taper state. The key idea is that the DPSS taper in the frequency domain, i.e., $|\hat{\phi}\rangle = U_{\text{QFT}}|\phi\rangle$, 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 non-trivial amplitude 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 [CSS18] with very little overhead. After that, we can perform an inverse quantum Fourier transform on $|\hat{\phi}^*\rangle$ to obtain $|\phi^*\rangle$, which closely approximates the DPSS taper $|\phi\rangle$.

Before we provide an explicit description of the taper state preparation protocol, we show in the lemma below that we can prepare an approximate DPSS taper state that is ϵ -close to the DPSS taper state using $N' = O(\log^2(1/\epsilon)) \ll N$ parameters where ϵ is an average failure probability of the tQPE algorithm using the DPSS taper. In this case, we refer to the non-trivial amplitudes of the state $|\hat{\phi}\rangle$ as the N' parameters that are needed to characterize our approximate DPSS taper state.

Lemma A.1. *The number of parameters, N' , necessary to prepare a state, $|\phi^*\rangle$, that is $O(\epsilon)$ -close in Euclidean distance to $|\phi\rangle$ i.e., $\| |\phi^*\rangle - |\phi\rangle \|_2 = O(\epsilon)$ satisfies $N' = 2 \lceil 175 (\log(10/\epsilon) + 1)^2 \rceil + 1$.*

Proof. Recall from Thm. 1 and Thm. 2 that

$$\epsilon \geq 10 \exp \left[-\frac{2K-6}{\frac{2}{\pi^2} \log(100K+25)} \right], \quad (\text{A1})$$

where ϵ is the average probability outside of the central lobe of the most-concentrated DPSS taper in the frequency domain. Because the DPSS taper in the frequency domain has $2K+1$ grid points in the central lobe, the number of parameters N' needed to describe the frequency amplitudes in the central lobe is then $2K+1$. Note that N' may be determined with a classical numerical computation on a lattice of size N with central lobe bandwidth parameter K (as shown in Figure A1). From the derivation of the proof of Thm. 2, we observe that $K = \lceil 175 (\log(10/\epsilon) + 1)^2 \rceil$.

Since $N' = 2K+1$, we obtain $N' = 2 \lceil 175 (\log(10/\epsilon) + 1)^2 \rceil + 1$.

Suppose we construct a state $|\hat{\phi}^*\rangle$ using the N' parameters as frequency amplitudes at the N^* positions inside the central lobe before appropriately normalizing the state. Let us denote the projection of $|\hat{\phi}^*\rangle$ onto the subspace spanned by the basis states with the N' parameters with $|\hat{\phi}^*\rangle_{N'}$. In other words, we have $|\hat{\phi}^*\rangle_{N'} = C|\hat{\phi}\rangle_{N'} = \frac{1}{\| |\hat{\phi}\rangle_{N'} \|_2} |\hat{\phi}\rangle_{N'}$ where C is the appropriate normalization constant. Likewise, we denote the complement of N' as $\bar{N}'$ which gives us $|\hat{\phi}^*\rangle_{\bar{N}'} = 0$ since $|\hat{\phi}^*\rangle$ has zero amplitude outside of the N' positions in the central lobe. Note that the act of projection onto these subspaces may result in sub-normalized states. We see that

$$\| |\hat{\phi}^*\rangle - |\hat{\phi}\rangle \|_2 = \| |\hat{\phi}^*\rangle_{\bar{N}'} + |\hat{\phi}^*\rangle_{N'} - (|\hat{\phi}\rangle_{\bar{N}'} + |\hat{\phi}\rangle_{N'}) \|_2 \quad (\text{A2})$$

$$\leq \| |\hat{\phi}^*\rangle_{\bar{N}'} - |\hat{\phi}\rangle_{\bar{N}'} \|_2 + \| |\hat{\phi}^*\rangle_{N'} - |\hat{\phi}\rangle_{N'} \|_2 \quad (\text{A3})$$

$$= \| |\hat{\phi}\rangle_{\bar{N}'} \|_2 + \| C|\hat{\phi}\rangle_{N'} - |\hat{\phi}\rangle_{N'} \|_2 \quad (\text{A4})$$

$$= O(\epsilon) + (C-1) \| |\hat{\phi}\rangle_{N'} \|_2 \quad (\text{A5})$$

$$= O(\epsilon) + (C-1)(1/C) \quad (\text{A6})$$

$$= O(\epsilon) + 1 - \| |\hat{\phi}\rangle_{N'} \|_2 \quad (\text{A7})$$

$$\leq O(\epsilon) + 1 - (1-\epsilon) \quad (\text{A8})$$

$$= O(\epsilon), \quad (\text{A9})$$

where we used the fact that $\| |\hat{\phi}\rangle_{\bar{N}'} \|_2 \leq \epsilon$ in the second equality and $\| |\hat{\phi}\rangle_{N'} \|_2 \geq 1 - \epsilon$ in the last inequality. These inequalities come from the bounds on the total probability inside and outside of the central lobe of the DPSS taper state. Lastly, using the Parseval-Plancherel identity, we obtain $\| |\phi^*\rangle - |\phi\rangle \|_2 = O(\epsilon)$ as desired. $\square$

We note that the lemma stated above is similar to Lemma 1 in [CSS18] which states that a Gaussian taper with the reciprocal standard deviation, $1/\sigma$, as the width of its central lobe, could also be used in place of a DPSS taper to concentrate phase measurements. However, it would not have the optimal concentration of probability in the central lobe that we would observe from the DPSS taper.

We now provide an explicit description of the taper state preparation method that closely follows App. B of [CSS18]. We first begin by classically computing the N' amplitudes that characterize the central lobe of $|\hat{\phi}\rangle$ using classical numerical methods. Next, we use these N' amplitudes to prepare a state on a register of $\log N'$ qubits using the preparation method described in [SBM05]. Subsequently, we append $p - \log N'$ qubits initialized to $|0\rangle^{\otimes(p-\log N')}$ to the ancilla register to form a p -qubit ancilla register. The N' amplitudes form an off-center lobe in the frequency domain, so we need to center these amplitudes on the p -qubit register to form a central lobe. In App. C of [CSS18], the authors provide an explicit circuit involving CNOT and X gates that performs the desired centering by permuting and mapping the original N' basis states $\{|j\rangle \mid 0 \leq j \leq N' - 1\}$ to the new N' basis states $\{|j\rangle \mid (N - N')/2 \leq j \leq (N + N')/2 - 1\}$ in the p -qubit ancilla register. Lastly, we perform the centered Fourier transform operator $U_{\text{C-QFT}}$ that was proposed in App. B of [CSS18].

Given $\epsilon > 0$, where ϵ is the failure probability of the tQPE algorithm, let us define $U_{\text{C-QFT}}$ as such:

$$U_{\text{C-QFT}} := \pi U_{\text{QFT}} \pi, \quad (\text{A10})$$

where π is a p -qubit cyclic permutation operator that maps $|j\rangle$ to $|j - N/2 \bmod N\rangle$, and U_{QFT} is the p -qubit Fourier transform operator. The cyclic permutation operator π can be implemented with $U_{\text{QFT}} Z U_{\text{QFT}}^{-1}$, where Z is a diagonal operator defined as $\text{diag}(Z) := [1, -1, 1, -1, \dots, 1, -1]$ as elaborated in App. B of [CSS18]. Therefore, the overall action of the centered Fourier transform operator $U_{\text{C-QFT}}$ can be understood as a shifted QFT that accounts for the new central lobe, and it outputs the approximate DPSS taper state $|\phi^*\rangle$ that is $O(\epsilon)$ -close to $|\phi\rangle$ in the ancilla register of the tQPE quantum circuit. Readers can find more details about the derivation in Sec. IV-D and App. B of [CSS18]. From Figure A1, we can see that with an $O(\epsilon)$ -close approximation of the DPSS taper, we can achieve the exponential decrease in sidelobe amplitudes described in Corollary 2.

Now, we analyze the complexity of the taper state preparation protocol. Ref. [SBM05] introduced a technique for preparing a state on a register of $\log N'$ qubits with a gate complexity of $O(N')$ from N' amplitudes. Note that their technique requires some classical computation, which we disregard, along with the classical computation of the N' amplitudes of the central lobe of $|\hat{\phi}\rangle$. For the quantum circuit required to center the amplitudes, it has been shown in App. C of [CSS18] that the number of two-qubit gates required is $\log\left(\frac{N-N'}{2}\right) = O(\log(N - N'))$. Lastly, we provide the gate complexity of $U_{\text{C-QFT}}$. We begin by providing its alternative expression:

$$U_{\text{C-QFT}} = \pi U_{\text{QFT}} \pi = \left(U_{\text{QFT}} Z U_{\text{QFT}}^{-1} \right) U_{\text{QFT}} \left(U_{\text{QFT}} Z U_{\text{QFT}}^{-1} \right) = U_{\text{QFT}} Z U_{\text{QFT}} Z U_{\text{QFT}}^{-1}. \quad (\text{A11})$$

Notice that we require three applications of QFT which has been shown in [Cop94] to have gate complexity $O(p^2) = O(\log^2 N)$. This gives us an overall gate complexity of $O(N' + \log(N - N') + \log^2 N) = O(N' + \log^2 N)$ for the state preparation protocol. Using the fact that $N' = O(\log^2(1/\epsilon))$ and $N = 2^p = 2^m \cdot 2^\ell = 2^{O(\log \log 1/\epsilon)} \cdot 2^{O(\log 1/\delta)}$, we can state the gate complexity of the state preparation protocol $\mathcal{C}_\phi$ in terms of δ and ϵ :

$$\mathcal{C}_\phi = O\left(\log^2(1/\epsilon) + [\log \log(1/\epsilon) + \log(1/\delta)]^2\right) \quad (\text{A12})$$

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

Having discussed how we can prepare an approximate DPSS taper $|\phi^*\rangle$ that is $O(\epsilon)$ -close to the DPSS taper $|\phi\rangle$, we now proceed to provide guarantees for the performance of $|\phi^*\rangle$. To do that we first write the approximate DPSS taper in terms of the DPSS taper:

$$|\phi^*\rangle = \sqrt{1 - O(\epsilon^2)} |\phi\rangle + O(\epsilon) |\phi^\perp\rangle, \quad (\text{A13})$$

where $\langle \phi | \phi^\perp \rangle = 0$.

Now, we consider the success probability of our tQPE algorithm when we use the approximate taper. We assume that the probability of error of the exact taper $|\phi\rangle$ is ϵ , that is

$$|\phi\rangle |\psi_\theta\rangle \xrightarrow{\text{QPE}} \sqrt{1 - \epsilon} \sum_i^{|\tilde{\theta}_i - \theta| \leq \delta} |\tilde{\theta}_i\rangle |\psi_\theta\rangle + \sqrt{\epsilon} \sum_i^{|\tilde{\theta}_i - \theta| > \delta} |\tilde{\theta}_i\rangle |\psi_\theta\rangle. \quad (\text{A14})$$

For $|\phi^\perp\rangle$ we assume the worst possibility, i.e. that it returns a δ -close phase estimate with zero probability. Then the approximate taper returns a δ -close estimate with probability at least

$$P_{\text{success}}^* \geq \left(1 - O(\epsilon^2)\right) (1 - \epsilon) = 1 - \epsilon - O(\epsilon^2) + O(\epsilon^3) = 1 - O(\epsilon). \quad (\text{A15})$$

Appendix B: Ideal Case

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 $\delta = 2^{-(l+1)}$ 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 (13). Since this is a constrained problem, we used the Lagrangian formulation to solve it. For ease of reference, we restate the corresponding Lagrangian (14) 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 (13), we need to find the stationary point of $\mathcal{L}$ that maximizes the objective function of (13). 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 (8):

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

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

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:

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

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

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

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

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

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

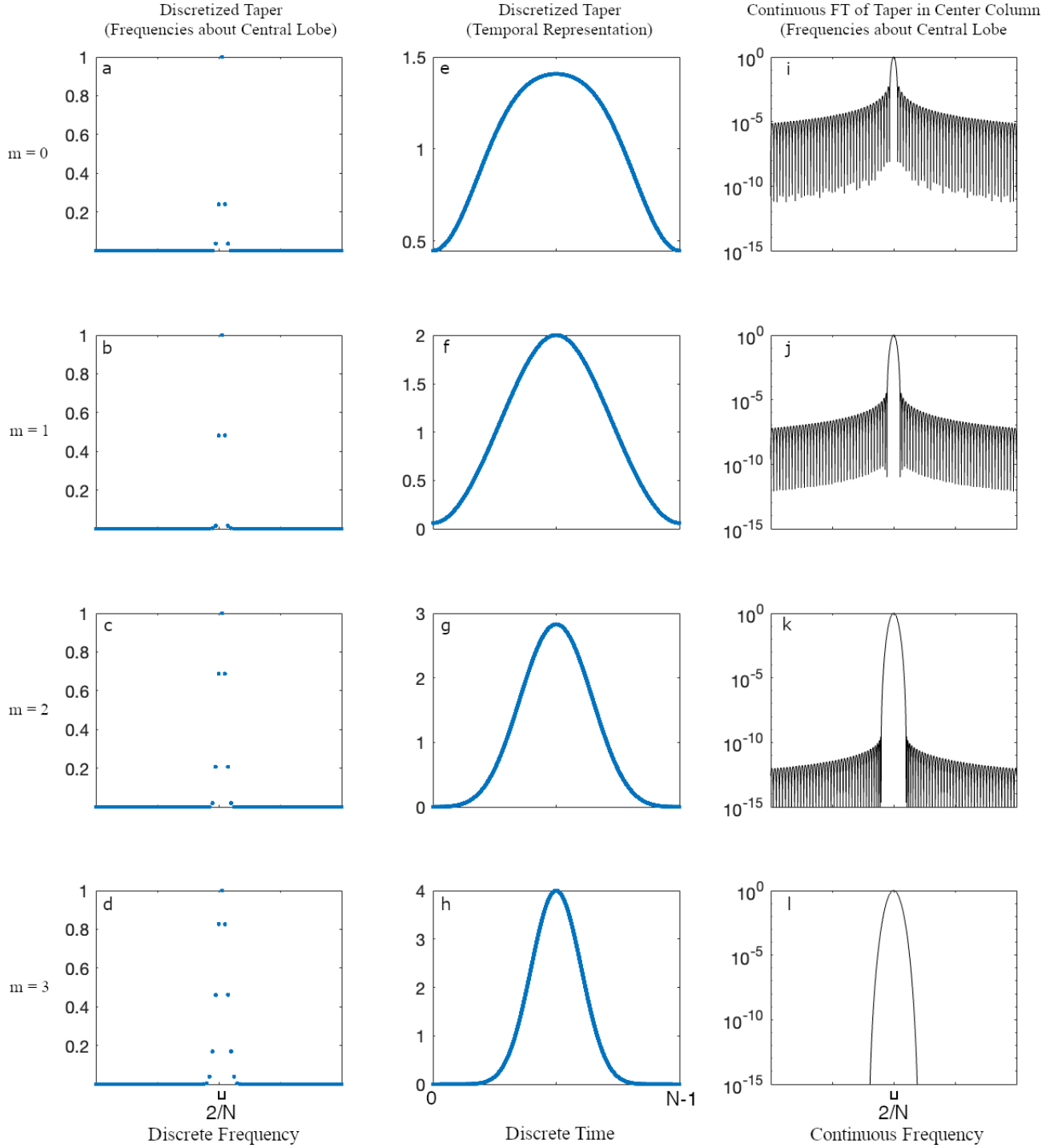


FIG. A1. Preparation of Taper States. Using the algorithm outlined in App. A, we plot discrete frequency amplitudes of DPSS tapers in the left hand column (a-d). The central lobe bandwidth of two lattice widths of the standard tophat taper is indicated by the bracket under panel d). There are ($N^* = 5, N^* = 7, N^* = 13, N^* = 21$) nonzero values within the central lobe needed for state preparation of ($m^* = 0, m^* = 1, m^* = 2, m^* = 3$) tapers, where $N^* \sim 2^{m^*}$. All values other than the central lobe amplitudes are set to zero. We then plot the discrete inverse Fourier transform of the values in the left hand column, transforming the DPSS tapers to the discrete time domain (e-h, center column). In the right hand column (i-l), we plot $\log_{10}$ of the measurement probability density (unnormalized). The key thing to note here is the ratio of the height of the central peak to the side lobe height, which decreases exponentially with N^* . The bandwidth of the standard tophat taper is indicated under panel l. Here, we see that using very few classically computed parameters, it is nonetheless possible to construct a DPSS taper whose side lobes decrease exponentially as a function of N^* .

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 (B7) and (B11) into the objective function of the original optimization problem (13), 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 B_Δ 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 B_Δ defined in (B9) to the bandlimiting operator $\mathcal{B}_K : \mathbb{C}^N \rightarrow \mathbb{C}^N$ described in Zhu, et al. [ZKD⁺17]. Suppose $K \in \mathbb{N}$ and $2K + 1 < N$. The bandlimiting 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 (B7) by a factor of $e^{2\pi i \Delta(m-n)}$ in each summand. We now introduce the timelimiting 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 [ZKD⁺17], 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 non-zero 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 [ZKD⁺17], are stated in the following theorem, which we restate with our notation.

Theorem C.1. [ZKD⁺17, Theorem 1] Suppose $K, N \in \mathbb{N}$ and $W = \frac{2K+1}{2N} < \frac{1}{2}$. 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

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

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, Thm. C.1 is stated for the matrix that corresponds to $\mathcal{T}_N \mathcal{B}_K \mathcal{T}_N$, the authors of [ZKD⁺17] 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 (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 (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 (B7) with eigenvalue $\lambda^{(i)}$, as the quantum periodic discrete prolate spheroidal sequence (QP-DPSS), where $-\frac{1}{2N} \leq \Delta \leq \frac{1}{2N}$.

Proposition C.1. *Suppose $\{(\lambda^{(i)}, \psi_i)\}_i$ represents the set of eigenvalues and eigensequences of $\mathcal{B}_K$. Then, for some $-\frac{1}{2N} \leq \Delta \leq \frac{1}{2N}$, $\{(\lambda^{(i)}, \phi_i^\Delta)\}_i$ is the set of eigenvalues and eigensequences of $\mathcal{B}_\Delta$ given in (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 (B7) with the proposed eigensequence, $\phi_i^\Delta[n]$.

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

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

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

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

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. $\square$

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 (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 Thm. C.1 and Prop. C.1, we have the following result.

Corollary C.1. *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 (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 (C5) dominates and therefore, (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 $W = \frac{2^m}{2^{l+m}}$ is the fraction of the lattice for which the taper has significant support. This gives

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

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

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

$$\Rightarrow m = O(\log(m \log 1/\epsilon + \log 1/\delta \log 1/\epsilon)) \quad (C17)$$

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

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

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

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

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

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. $\square$

Appendix D: Average-Case

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

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

As mentioned before, we focus on the case where Δ is uniformly distributed over the interval $[-\frac{1}{2N}, \frac{1}{2N}]$, 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 outlined in App. B. Now, consider the following:

$$\mathcal{L}_{\text{avg}}(|\phi\rangle, \lambda) \quad (D2)$$

$$= \int_{-\frac{1}{2N}}^{\frac{1}{2N}} d\Delta N \sum_{j=-K}^K \left| \hat{\phi}\left(\Delta + \frac{j}{N}\right) \right|^2 - \lambda \left(\sum_{n=0}^{N-1} |\phi[n]|^2 - 1 \right) \quad (D3)$$

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

$$= \sum_{n,n'=0}^{N-1} \phi[n] \phi^*[n'] \int_{-\frac{1}{2N}}^{\frac{1}{2N}} d\Delta e^{2\pi i \Delta (n-n')} \sum_j e^{-2\pi i (j/N)(n'-n)} - \lambda \left(\sum_{n=0}^{N-1} \phi[n] \phi^*[n] - 1 \right) \quad (D5)$$

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

$$= \sum_{n,n'=0}^{N-1} \phi[n] \phi^*[n'] \underbrace{\frac{\sin(\pi(n'-n)(2K+1)/N)}{\pi(n'-n)}}_{=:C(n,n')} - \lambda \left(\sum_{n=0}^{N-1} \phi[n] \phi^*[n] - 1 \right) \quad (D7)$$

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

$$\frac{\partial \mathcal{L}_{\text{avg}}}{\partial \phi^*[m]} = \sum_{n=0}^{N-1} \phi[n]C(n, m) - \lambda \phi[m] = 0 \quad (\text{D8})$$

$$\implies \sum_{n=0}^{N-1} \phi[n]C(n, m) = \lambda \phi[m]. \quad (\text{D9})$$

Note that the above equation is an eigenvalue equation, and its eigenvectors are DPSSs, previously studied in classical signal processing [Sle78]. Next, differentiating $\mathcal{L}$ with respect to the Lagrange multiplier λ and setting it to zero leads to the normalization constraint for the taper:

$$\frac{\partial \mathcal{L}_{\text{avg}}}{\partial \lambda} = - \sum_{n=0}^{N-1} \phi[n]\phi^*[n] + 1 = 0 \quad (\text{D10})$$

$$\implies \sum_{n=0}^{N-1} \phi[n]\phi^*[n] = 1, \quad (\text{D11})$$

Next, by plugging all the stationary points $(|\phi\rangle, \lambda)$ that satisfy the conditions (D9) and (D11) into the objective function of $\mathcal{L}_{\text{avg}}$, we get the following:

$$\mathbb{E}_{\Delta \sim \mathcal{D}} \left[\sum_{j=-K}^K \left| \hat{\phi} \left(\Delta + \frac{j}{N} \right) \right|^2 \right] = \lambda. \quad (\text{D12})$$

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 [Sle78, KRD21], 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 (E3) can be represented as

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

From the Parseval-Plancherel theorem, we have

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

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, (E11), holding because $\mathcal{T}_N$ is self-adjoint.

From (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.