arXiv:2306.07008v5 [quant-ph] 23 Dec 2024

Quantum Phase Estimation by Compressed Sensing

Changhao Yi[1,2,3], Cunlu Zhou[4,5], and Jun Takahashi[6,5]

[1]State Key Laboratory of Surface Physics, Department of Physics, and Center for Field Theory and Particle Physics, Fudan University, Shanghai, China

[2]Institute for Nanoelectronic Devices and Quantum Computing, Fudan University, Shanghai, China

[3]Shanghai Research Center for Quantum Sciences, Shanghai, China

[4]Department of Computer Science & Institut Quantique, Université de Sherbrooke, QC, Canada

[5]Center for Quantum Information and Control & Department of Physics and Astronomy, University of New Mexico, NM, USA

[6]Institute of Solid State Physics, University of Tokyo, Chiba, Japan

As a signal recovery algorithm, compressed sensing is particularly effective when the data has low complexity and samples are scarce, which aligns naturally with the task of quantum phase estimation (QPE) on early fault-tolerant quantum computers. In this work, we present a new Heisenberg-limited, robust QPE algorithm based on compressed sensing, which requires only sparse and discrete sampling of times. Specifically, given multiple copies of a suitable initial state and queries to a specific unitary matrix, our algorithm can recover the phase with a total runtime of $\mathcal{O}(\epsilon^{-1}\mathrm{poly}\log(\epsilon^{-1}))$, where ϵ is the desired accuracy. Additionally, the maximum runtime satisfies $T_{\max}\epsilon \ll \pi$, making it comparable to state-of-the-art algorithms. Furthermore, our result resolves the basis mismatch problem in certain cases by introducing an additional parameter to the traditional compressed sensing framework.

1 Introduction

Quantum phase estimation (QPE) [1] is one of the most useful subroutines in quantum computing, playing a crucial role in many promising quantum applications [2–4]. Given a unitary matrix U and one of its eigenvectors $|\Phi\rangle$ with eigenphase $e^{i2\pi\theta}$, the task of QPE is to estimate phase θ within a specified accuracy. Assuming the unitary matrix U represents the evolution operator under a Hamiltonian H, the task of QPE becomes equivalent to estimating a specific eigenenergy E_0 [5, 6]. Consequently, this subroutine has numerous applications in condensed matter physics, high-energy physics, and quantum chemistry. As a generalization, the problem of estimating multiple phases of U has been referred to as the quantum eigenvalue estimation problem (QEEP) [5–10].

While fully fault-tolerant quantum computers may still be years away, early fault-tolerant quantum computers with a limited number of logical qubits and gates are expected to emerge much sooner, capable of solving nontrivial tasks that demonstrate practical quantum advantages. Given the crucial role of QPE in many such tasks, it is imperative to design QPE algorithms specifically tailored for early fault-tolerant quantum computers. The standard textbook QPE algorithm [11] does not require an exact eigenstate as the

Changhao Yi: These authors contributed equally to this work.

Cunlu Zhou: These authors contributed equally to this work.

initial state and involves only one measurement, but it relies on a large number of ancilla qubits and controlled operations, making it experimentally demanding. Although Kitaev's original iterative QPE algorithm [1] uses just one ancilla qubit and a single controlled operation (see Fig. 1), it requires the initial state to be an exact eigenstate, which can be a challenging task itself. Therefore, neither approach is well-suited for early fault-tolerant quantum computers.

Most of the recent work [5, 12, 13] in QPE for early fault-tolerant quantum computers have focused on designing better protocols to improve various aspects of Kitaev's original QPE algorithm. More specifically, the following properties are desired when designing such algorithms:

- The quantum circuit should be simple, using at most one ancilla qubit and one controlled operation.
- The initial state is not necessarily an exact eigenstate of U.
- The total runtime achieves the Heisenberg limit, i.e., $\mathcal{O}(\epsilon^{-1}\,\mathrm{poly}\log(\epsilon^{-1}\delta^{-1}))$ for estimating the phase θ to accuracy ϵ with probability $1-\delta$.
- When the overlap of the initial state and the targeted eigenstate is large, the maximal runtime $T_{\max}$ (hence the maximum circuit depth) can be much smaller than π/ϵ.

In addition, the *experimental complexity* should also be considered carefully for early fault-tolerant quantum computers: It is desirable to have a small number of time samples with a regular choice of values. For early fault-tolerant experiments, one may still need to prepare quantum circuits for each time t by hand, and the total cost would be high if the number of different sample times is large. It is comparably easier to run the same quantum circuit multiple times, instead of running different quantum circuits a few times. Additionally, in quantum simulation algorithms [14, 15] the time evolution by $U(t)$ is approximated by short-time evolutions: $U(t) \approx U_{\mathrm{sim}}(\Delta t)^L$. Thus, it is more convenient to sample from a discrete set of times $\mathcal{T} = \{n\Delta t, n \in \mathbb{N}\}$ rather than to sample from a continuous region. Most state-of-the-art algorithms sample times from a continuous region [5, 6, 16] while only a few consider discrete sampling [17, 18].

For a Heisenberg-limited QPE algorithm with maximal runtime $T_{\max}$, if the size of the time samples needed is $\mathcal{O}(\mathrm{poly}\log T_{\max})$, the sampling is considered sparse in our paper. In this work, we design a non-adaptive algorithm that only requires discrete and sparse sampling of times without sacrificing performance. Informally, our main result can be stated as follows.

Theorem 1 (Informal)**.** *Suppose the target signal satisfies an approximately-on-grid assumption and its dominant part has frequency gap N^{-1}. Then our compressed sensing based QPE algorithm satisfies the following conditions:*

$$|\mathcal{T}| = \mathcal{O}(\mathrm{poly}\log N), \quad T_{\max} = \mathcal{O}(N), \quad T_{\mathrm{total}} = \mathcal{O}(N\,\mathrm{poly}\log N), \quad \epsilon = \mathcal{O}(\sigma_{\mathrm{off}} N^{-1}), \tag{1}$$

where $|\mathcal{T}|$ is the number of samples, $T_{\max}$ is the maximum time for each time evolution, T_{total} is the total runtime, and ϵ is the desired accuracy, and σ_{off} quantifies the size of the off-grid component. The classical runtime of the algorithm is $\mathcal{O}(\sigma_{\mathrm{off}}^{-1} N\,\mathrm{poly}\log N)$

A signal y is approximately-on-grid if there exists a "Fourier" matrix (with certain parameters) F such that $y = Fx, x \in \mathbb{R}^N$. Note that under this assumption, we do not require $p_0 \geq \frac{1}{2}$, where p_0 is the overlap between the initial state and the ground state.

The rest of the paper is organized as follows. We start with preliminaries about QEEP, sparse Fourier transformation and compressed sensing in Sec. 2. We then introduce our QPE algorithm based on compressed sensing in Sec. 3 and prove several analytical results, including its Heisenberg-limit scaling. We also numerically test the performance of our algorithm and compare it to previous works in Sec. 4. In the final section, we summarize several open problems and potential future research directions in Sec. 5.

2 Main idea

2.1 Setup

The QEEP can be formulated as a sparse signal recovery problem. Given an initial state $|\Phi\rangle$ and a Hamiltonian with spectrum decomposition $H = \sum_{\ell=0}^{D-1} E_\ell P_\ell$, where $\{E_\ell\}_{\ell=0}^{D-1}$ are energy levels and $\{P_\ell = |\phi_\ell\rangle\langle\phi_\ell|\}_{\ell=0}^{D-1}$ are projectors onto the corresponding eigenstates, the time domain signal in QEEP can be written as

$$y^0(t) = \langle\Phi|e^{-\mathrm{i}Ht}|\Phi\rangle = \sum_{\ell=0}^{D-1} |\langle\Phi|\phi_\ell\rangle|^2 e^{-\mathrm{i}E_\ell t}. \tag{2}$$

In QEEP we assume that $|\Phi\rangle$ has the following decomposition:

$$|\Phi\rangle = \sum_{\ell\in\mathcal{L}_{\mathrm{dom}}} \sqrt{p_\ell}|\phi_\ell\rangle + \sum_{\ell\in\mathcal{L}_{\mathrm{res}}} \sqrt{p_\ell}|\phi_\ell\rangle, \quad \sum_{\ell\in\mathcal{L}_{\mathrm{dom}}} p_\ell \approx 1, \quad |\mathcal{L}_{\mathrm{dom}}| \ll D, \tag{3}$$

where $\mathcal{L}_{\mathrm{dom}}$ denotes the dominant component of the signal, and $\mathcal{L}_{\mathrm{res}}$ is the residue component. Under this assumption, we can regard $y^0(t)$ as a sparse signal. The formal definition of sparsity will be given in the main text. In particular, when $|\mathcal{L}_{\mathrm{dom}}| = 1$, the task becomes QPE. For QPE, without loss of generality[i], we will mainly discuss the estimation of the ground energy E_0, i.e., the smallest eigenvalue of H.

The sparsity assumption applies to a wide range of situations. For instance, if we regard $|\Phi\rangle$ as the ground state of a perturbed Hamiltonian $H+V$, then the overlap $|\langle\Phi|\phi_\ell\rangle|$ tends to decay exponentially with the energy difference $|E_0 - E_\ell|$ (Because $|\langle\phi_\ell|V|\phi_0\rangle|$ decays exponentially with it. See [19] for details). Therefore, $|\Phi\rangle$ almost has no overlap with excited states with high energies, and $\mathcal{L}_{\mathrm{dom}}$ only contains a small amount of energy levels.

The objective of a QEEP algorithm is to estimate $\mathcal{L}_{\mathrm{dom}}$ within a specified accuracy ϵ using estimations of y_t^0 on a time set $\mathcal{T}$. An algorithm of this type can be separated into the quantum part and the classical post-processing part. Usually, the quantum part is a combination of Hamiltonian simulation [14] and the Hadamard test (see Fig. 1). Hamiltonian simulation algorithms are used to prepare the evolution operator $U(t)$. Longer evolution time requires more quantum gates, and the best-known circuit complexity for running $e^{-\mathrm{i}Ht}$ without ancilla qubit is almost linear in $\|Ht\|$ (t can be negative) [15, 20]. The total runtime T_{total} is thus

$$T_{\mathrm{total}} = \sum_{t\in\mathcal{T}} M_{\mathrm{H}}|t|. \tag{4}$$

where M_{H} is the number of Hadamard tests required for each y_t^0. Usually, we set M_{H} to be in order $\mathcal{O}(\log(|\mathcal{T}|)\sigma_{\mathrm{H}}^{-2})$, so that $\{y_t^0\}_{t\in\mathcal{T}}$ can all be estimated within error σ_{H}.

[i] In this work, we do not consider the hardness of state preparation. From the viewpoint of phase estimation, there is nothing special about the ground state energy compared to other eigenvalues as long as one can prepare an initial state that is close enough to the target eigenstate.

The total runtime T_{total} reflects the total circuit depth for running the algorithm. If $T_{\text{total}} = \mathcal{O}(\epsilon^{-1}\text{poly}(\epsilon^{-1}))$, we say that the algorithm satisfies the Heisenberg limit. Another important metric of complexity is the maximal runtime $T_{\max} = \max_{t\in\mathcal{T}} |t|$, which reflects the maximum circuit depth. Due to the difficulty in constructing large-size quantum circuits, the restriction on the maximal runtime is particularly important for early fault-tolerant quantum computers.

The notations frequently used in the main text is summarized in Table 1.

Table 1: Notations

Notation	Meaning
$y(t)$	noisy time domain signal
$y^0(t)$	ideal time domain signal
$z(t)$	noise in time domain signal
$x(k)$	ideal frequency domain signal
τ	unit time step
$\mathcal{T}$	sample of times divided by τ
N	signal length, $T_{\max}/\tau$
ϵ	accuracy on energy level
σ	noise tolerance parameter
$\mathcal{F}$	set of frequencies
S	sparsity

2.2 Previous work

The classical aspect of QPE and QEEP involves estimating frequencies from statistically sampled sparse signals, a process akin to the objectives of sparse Fourier transformation (SFT) algorithms [21]. Based on the data types, SFT algorithms can be classified into discrete setting algorithms and continuous setting algorithms. A discrete SFT algorithm performs discrete Fourier transformation, where both the time and the frequency of the signal are restricted on a discrete set:

$$y_t = \sum_{k=0}^{N-1} x_k e^{-\mathrm{i}2\pi kt/N}, \quad \{y_t\}_{t=0}^{N-1} \xrightarrow{\text{discrete SFT}} \{x_k\}_{k=0}^{N-1}. \tag{5}$$

A continuous SFT algorithm [22–24] aims to accomplish a more general task:

$$y(t) = \sum_{f\in\mathcal{F}} p_f e^{-\mathrm{i}2\pi ft} \xrightarrow{\text{continuous SFT}} x(k) = \sum_{f\in\mathcal{F}} p_f \delta(k-f), \tag{6}$$

where $y(t)$ is the time domain signal, and $x(k)$ is the frequency domain signal. For QPE, we do not assume that the frequencies (energies) live in a discrete space. Thus, continuous SFT algorithms are more appropriate. In both setups, sparsity S represents the number of distinct frequencies.

There are several aspects of evaluating the performance of an SFT algorithm. Its runtime complexity, sample complexity, and resolution are all important ingredients to consider. Here the runtime complexity refers to how long the algorithm takes on a classical computer, the sample complexity measures the number of time domain signal samples required in the algorithm, and the resolution quantifies the differences between the true frequencies and their estimates. For example, the Fast Fourier Transformation algorithm

[25] has runtime complexity $\mathcal{O}(N \log N)$ with sample complexity $\mathcal{O}(N)$. So far, the best runtime complexity is $\mathcal{O}(S \log^c(N) \log(N/S))$ with $c > 2$ [26], and the most sample-efficient algorithm requires only $\mathcal{O}(S \log S \log N)$ samples [27]. In practical scenarios, we most likely have noisy data, necessitating the need for algorithmic robustness. Given the unique characteristics of our quantum setting, we prioritize the sample complexity, resolution, and robustness of an algorithm.

Several continuous SFT algorithms have been applied to QEEP. To the best of our knowledge, [7] was the first work to address QEEP using Hadamard tests, treating it as a time-series analysis problem. Later, [5] emphasized the importance of the Heisenberg-limited scaling. By applying Fourier-filter function techniques, it presented a Heisenberg-limited QPE algorithm for early fault-tolerant quantum computers. The algorithm was further improved by the other follow-up works [6, 9], where the Gaussian derivative filter function was used to reduce the maximal runtime of the algorithm. In [6], $T_{\max}$ was reduced to a "constant" depth, i.e., a quantity that only depends on the spectral gap, at the expense of increasing T_{total} from $O(\epsilon^{-1}\text{poly}\log(\epsilon^{-1}))$ to $O(\epsilon^{-2}\text{poly}\log(\epsilon^{-1}))$, which made the algorithm not Heisenberg-limited.

Two recent QPE algorithms [12, 13], inspired by Robust Phase Estimation (RPE) [28–30], can also efficiently reduce the maximal runtime. These recent algorithms also improved the relation between $T_{\max}$, the initial overlap p_0, and the final accuracy ϵ. When the overlap p_0 is large, [12] reduces the prefactor τ_c in the maximum runtime scaling $T_{\max} = \tau_c/\epsilon$ by using a subroutine called the quantum complex exponential least squares (QCELS). In contrast to [5] in which the prefactor τ_c is at least π, the prefactor in [12] can be arbitrarily close to 0 as $p_0 \to 1$. In [16] and [17], the last two QPE algorithms have been extended to the QEEP setup.

Another recent work [18] proposed an efficient and versatile phase estimation algorithm named Quantum Multiple Eigenvalue Gaussian Filtered Search (QMEGS), which has most of the good properties mentioned above. Here we would like to emphasize its connection to a signal processing algorithm named Orthogonal Matching Pursuit (OMP) [31]. OMP is a greedy algorithm that searches for the dominant frequencies of a signal by maximizing the overlaps step by step. QMEGS can be regarded as an OMP algorithm with a modified time sampling procedure to reduce the maximal and total runtime. The OMP algorithm has a strong connection with compressed sensing and can be potentially combined with our algorithm.

2.3 QPE by compressed sensing

Our main contribution is a simple and robust classical post-processing algorithm for QPE based on compressed sensing [32–34]. Our algorithm only requires sparse sampling of times from a discrete set.

Compressed sensing is a prominent signal-processing algorithm with wide applications in various domains such as time-frequency analysis, image processing, and quantum state and process tomography [35, 36, 36–38]. It aims to solve special types of underdetermined linear inverse problems, i.e., given $y \in \mathbb{R}^M$ and $A \in \mathbb{R}^{M\times N}$ with $M \ll N$, finding the unique sparse solution to $Ax = y, x \in \mathbb{R}^N$. Certainly, the solution is not unique without further restrictions. If we assume x is S-sparse and A satisfies the restricted isometry property (RIP) [33] over $\mathcal{O}(S)$-sparse signals, then x can be uniquely recovered by solving a linear programming problem:

$$\min_{x\in\mathbb{R}^N} \|x\|_1, \quad \text{s.t.} \quad Ax = y. \tag{7}$$

If we set x as the frequency domain signal, y as the signal on the time samples, and A as the partial Fourier transformation operator, then this compressed sensing subroutine can be used for discrete SFT. It has been proved that with $\mathcal{O}(S \operatorname{poly} \log N)$ number of samples, one can successfully recover the frequency domain signal x with high probability [32]. For noisy situations, the signal can still be recovered by solving the following quadratic programming problem [33]:

$$\min_{x \in \mathbb{R}^N} \|x\|_1, \quad \text{s.t.} \quad \|Ax - y\|_2 \leq b. \tag{8}$$

The small number of required samples and the robustness against noisy sampling make compressed sensing an appealing post-processing algorithm.

Unfortunately, there is a significant drawback of compressed sensing: it only works for discrete SFT, not for continuous SFT. In other words, frequencies are assumed to be on a grid in the sense that $f = n/N, n \in [N]$. The on-grid assumption is unnatural for many signals in practice. The gap between the continuous world and the discrete model is formally termed as *basis mismatch* in signal analysis. Although off-grid compressed sensing algorithms have been proposed for solving the basis mismatch problem [39], the performance in our numerical test is not ideal. We show that with a slight modification, the vanilla compressed sensing algorithm can be used for special types of continuous SFT tasks. In other words, our algorithm partially solves the basis mismatch problem.

An overview of our algorithm is described as follows. For signal vectors with size N, when the frequencies are all nearly on-grid ($f \approx n/N, n \in \mathbb{Z}$) and the noise for each sample is bounded by a constant, the convex relaxation algorithm can recover the frequencies with only $\mathcal{O}(\operatorname{poly} \log N)$ samples, which satisfies the Heisenberg limit. With no prior knowledge about f (i.e., f could be off-grid), we introduce a grid shift parameter ν such that after shifting the signal by $e^{-\mathrm{i}2\pi f t} \to e^{-\mathrm{i}2\pi(f-\nu/N)t}$, the dominant frequencies of the new signal become nearly on-grid. This step requires an assumption on the signal, but we will show that a wide range of signals satisfy such an assumption. For each trial of ν, we run the compressed sensing subroutine on the data set $\{y_t\}_{t \in \mathcal{T}}$ to obtain a trial solution s_ν. The optimal ν is the one with the smallest $\|s_\nu\|_1$. By searching for the optimal grid-shift parameter in a finite set $\mathcal{V}$, the accuracy of the dominant frequencies is $\mathcal{O}(\sigma_{\text{off}}/N)$, where σ_{off} is the maximal entry of the minimal off-grid component. This quantity is related to the noise, the frequency gap, and the residual part of the signal. In terms of the maximum runtime $T_{\max}$, since the samples of the compressed sensing algorithm are integers in $[1, N]$, $T_{\max}$ scales linearly in N, and T_{total} is $\mathcal{O}(N \operatorname{poly} \log N)$.

3 Main results

3.1 Algorithm

In this subsection, we present an overview of our algorithm for QPE using compressed sensing. The quantum part of the algorithm is represented as follows.

For each time $t \in \mathcal{T}$, $y^0(t) = \langle \Phi | e^{-\mathrm{i}Ht} | \Phi \rangle$ can be estimated from averaging over the Hadamard tests. More precisely, by choosing $\mathbf{W} = I$ in Fig. 1, the measurement outcome of the ancilla qubit is a random variable $h_x(t)$ such that

$$\Pr[h_x(t) = +1] = \frac{1}{2}[1 + \operatorname{Re}(y^0(t))], \quad \Pr[h_x(t) = -1] = \frac{1}{2}[1 - \operatorname{Re}(y^0(t))]. \tag{9}$$

Similarly, when $\mathbf{W} = \mathbb{S}^\dagger$ ($\mathbb{S}$ is the phase gate), the measurement outcome is another random

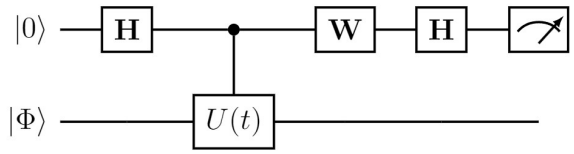

Figure 1: The one-ancilla quantum circuit used in Kitaev-type QPE algorithms. The final measurement is done in the Z basis. In terms of the measurement outcome, we regard the $|0\rangle$ state as obtaining value $+1$, and the $|1\rangle$ state as obtaining value -1. $\mathbf{H}$ is the Hadamard gate; $\mathbf{W}$ has two choices: when $\mathbf{W} = I$, the measurement outcome is ± 1 with probability $(1 \pm \mathrm{Re}(\langle\Phi|U(t)|\Phi\rangle))/2$ respectively. When $\mathbf{W} = \mathbb{S}^\dagger$, the complex conjugation of the phase gate, the measurement outcome is ± 1 with probability $(1 \pm \mathrm{Im}(\langle\Phi|U(t)|\Phi\rangle))/2$ instead. After averaging over many test outcomes, we obtain an estimate of the true signal $\langle\Phi|U(t)|\Phi\rangle$.

Algorithm 1 Signal estimation by Hadamard test

Input: Set of sampled integers $\mathcal{T}$, unit time step τ, Hamiltonian H, initial state $|\Phi\rangle$, error tolerance parameter σ_{H}, failure probability δ.

Output: $\{y(n\tau), n \in \mathcal{T}\}$.

1: **for** $n \in \mathcal{T}$ **do**
2: Prepare the initial state $|\Phi\rangle$ and unitary operator $e^{-\mathrm{i}Hn\tau}$;
3: Perform Hadamard tests for $\mathcal{O}(\log(|\mathcal{T}|/\delta)\sigma_{\mathrm{H}}^{-2})$ times;
4: Compute the average value of the test outcomes as $y(n\tau)$.
5: **end for**

variable $h_y(t)$ such that

$$\Pr[h_y(t) = +1] = \frac{1}{2}[1 + \mathrm{Im}(y^0(t))], \quad \Pr[h_y(t) = -1] = \frac{1}{2}[1 - \mathrm{Im}(y^0(t))]. \tag{10}$$

The summation of the two gives us the estimate of $y^0(t)$:

$$\mathbb{E}[h_x(t) + \mathrm{i}h_y(t)] = y^0(t). \tag{11}$$

After sampling the random variables $h_x(t), h_y(t)$ for M_{H} times, we obtain a noisy signal:

$$y(t) = \overline{h_x(t) + \mathrm{i}h_y(t)} = y^0(t) + z(t). \tag{12}$$

Here the overline represents the average over samples, and the noise $z(t)$ originates from the statistical uncertainty of the Hadamard tests. Hoeffding's inequality ensures that with probability $1 - \delta'$, we have

$$|z(t)| = \mathcal{O}\left(\sqrt{\frac{1}{M_{\mathrm{H}}}\log\frac{1}{\delta'}}\right). \tag{13}$$

In the rest of the paper, the meanings of $z(t)$ are not identical, but they always represent the noisy part of the signal. Let σ_{H} be the noise tolerance parameter. To guarantee $|z(t)| < \sigma_{\mathrm{H}}, \forall t \in \mathcal{T}$, with probability at least $1 - \delta$, we require $\delta = \mathcal{O}(\delta'|\mathcal{T}|^{-1})$ so that $M_{\mathrm{H}} = \Omega(\log(|\mathcal{T}|/\delta)/\sigma_{\mathrm{H}}^2)$. For a more rigorous proof, see Appendix A of [12]. The total runtime is thus

$$T_{\mathrm{total}} = \sum_{t\in\mathcal{T}} M_{\mathrm{H}}|t| = \mathcal{O}\left(\log(|\mathcal{T}|\delta^{-1})\sigma_{\mathrm{H}}^{-2}\sum_{t\in|\mathcal{T}|}|t|\right). \tag{14}$$

The quantum part of the algorithm is shown in Algorithm 1.

Algorithm 2 Quantum phase estimation by compressed sensing

Input: Signal length N, signal sparsity S, sampling ratio r, unit time step τ, Hamiltonian H, initial state $|\Phi\rangle$, noise tolerance parameter $\sigma_{\mathrm{H}}, \sigma_{\mathrm{test}}, \sigma$, failure probability δ, size of the trial set J, amplitude lower bound $p_{\min}$.

Output: $\nu_* = \nu_{j^*},\ s_{\nu_*} = s_{\nu_{j^*}},\ E^* = 2\pi(\min \mathcal{K} + \nu_{j^*})/N$.

Sample integers from $[N]$ with sampling ratio r. Denote the set of the samples by $\mathcal{T}$.
Apply Algorithm 1 with input $(\mathcal{T}, \tau, H, |\Phi\rangle, \sigma_{\mathrm{H}}, \delta)$. Denote the output by $\{y_n\}_{n\in\mathcal{T}_1}$.
for $j = 0, 1, \cdots, J-1$ **do**
    Set $\nu_j = -1/2 + j/J$.
    Solve

$$\min_{s\in\mathbb{R}^N} \|s\|_1, \quad \text{s.t.} \quad \|F_{\nu_j,\mathcal{T}} s - y_{\mathcal{T}}\|_2 \le \sqrt{|\mathcal{T}|}\sigma$$

    to obtain s_{ν_j}. If there is no feasible solution, set $s_{\nu_j} = (1, 1, \cdots, 1)$.
    Record (ν_j, s_{ν_j}) as the j-th solution.
end for
Sample integers from $[N]$ with sampling ratio r. Denote the set of samples by $\mathcal{T}_2$.
Apply Algorithm 1 with input $(\mathcal{T}_2, \tau, H, |\Phi\rangle, \sigma_{\mathrm{H}}, \delta)$. Denote the output by $\{y'_n\}_{n\in\mathcal{T}_2}$.
for $j = 0, 1, \cdots, J-1$ **do**
    Apply Algorithm 3 with input $(\nu_j, s_{\nu_j}, \{y'_n\}_{n\in\mathcal{T}_2}, \sigma_{\mathrm{test}})$. Denote the output by o_j.
    if $\mathrm{o}_j = 0$ **then**
        $\ell_j = N + 1$.
    else
        $\ell_j = \|s_{\nu_j}\|_1$.
    end if
end for
Let $j^* = \arg\min \ell_j$
Find all entries of $s_{\nu_{j^*}}$ satisfying $s_{\nu_{j^*}} \ge p_{\min}$. Denote the set of indices as $\mathcal{K}$.

For the classical post-processing part, the goal is to recover the dominant frequencies of y_t^0 with the noisy samples $\{y(t)\}_{t\in\mathcal{T}}$. To write the QEEP in Eq. (2) in the form of a compressed sensing problem, we first reformulate the problem on a "grid". Introduce a unit time step τ such that

$$y_n^0 = \sum_{f\in\mathcal{F}} p_f e^{-\mathrm{i}2\pi f n}, \quad \mathcal{F} = \left\{ \frac{E_\ell \tau}{2\pi} : \ \ell \in \mathcal{L}_{\mathrm{dom}} \right\}. \tag{15}$$

The dominant eigenenergy set $\mathcal{L}_{\mathrm{dom}}$ defined in Eq. (3) determines the frequency support $\mathcal{F}$. To keep the order of energy levels unchanged, we require that $E_\ell \tau \in [0, 2\pi),\ \forall \ell$. This condition can always be satisfied by adding a constant to the Hamiltonian H and choosing τ properly. The actual data to be processed is thus $\{y_n = y_n^0 + z_n\}_{n\in\mathcal{T}}$. Recall that the choice of M_H guarantees that $|z_n| \le \sigma_{\mathrm{H}}\ \forall n \in \mathcal{T}$, with high probability.

As discussed in Sec. 2.3, because we cannot always assume $f \approx n/N,\ \forall f \in \mathcal{F}$, the regular compressed sensing algorithm is not guaranteed to work. Our algorithm significantly relaxes the assumption by introducing a grid-shift parameter ν. As a simple instance, suppose the frequency support $\mathcal{F}$ satisfies

$$f = \frac{n+\nu}{N}, \quad n \in [N],\ \forall f \in \mathcal{F}. \tag{16}$$

Then y_n^0 can be transformed into an on-grid signal in a new basis as $y_n^0 = F_\nu x, x \in \mathbb{R}^N$,

Algorithm 3 Test of another sampling

Input: Parameter ν, frequency-domain signal s_ν, sampled data $\{y_m\}_{m\in\mathcal{T}_2}$, mean squared error threshold σ_{test}.

Output: 0 if the data fails the test; 1 if the data passes the test.

1: Compute the total empirical error with respect to the new set

$$\mathcal{E} = \sum_{m\in\mathcal{T}_2} |(F_\nu s_\nu)_m - y_m|^2.$$

2: **if** $\mathcal{E} \geq |\mathcal{T}_2|\sigma_{\text{test}}^2$ **then**
3: Return 0.
4: **else** Return 1.
5: **end if**

where F_ν is the shifted Fourier transformation:

$$(F_\nu)_{nk} \equiv e^{-\mathrm{i}2\pi(k+\nu)n/N} \quad n,k\in[N]. \tag{17}$$

The signal can then be recovered by solving

$$\min_{s\in\mathbb{R}^N} \|s\|_1, \quad \text{s.t.} \quad \|\mathcal{P}_\mathcal{T}(F_\nu s - y)\|_2 \leq \sqrt{|\mathcal{T}|}\sigma, \tag{18}$$

where $\mathcal{P}_\mathcal{T}$ represents the projector

$$(\mathcal{P}_\mathcal{T})_{ij} = \delta_{ij}1_{i\in\mathcal{T}}. \tag{19}$$

Define $F_{\nu,\mathcal{T}} \equiv \mathcal{P}_\mathcal{T}F_\nu$ and $y_\mathcal{T} \equiv \mathcal{P}_\mathcal{T}y$ for later use. One can argue that a general signal does not satisfy the condition in Eq. (16), even approximately. We will address this issue in the next section and validate the universality of our algorithm with numerical evidence.

In the case that the condition in Eq. (16) is indeed satisfied, we can find ν by a brute-force search. Let $\mathcal{V}$ be a trial set of the grid shift parameters. For each $\nu\in\mathcal{V}$[ii], let s_ν be the solution of the compressed sensing subroutine and s_{ν_*} be the final solution with the smallest 1-norm:

$$\nu_* \equiv \arg\min_{\nu\in\mathcal{V}} \|s_\nu\|_1. \tag{20}$$

To bound the algorithmic error properly, we need a rough bound for $|\nu_* - \mathfrak{u}|$ first, where $\mathfrak{u}$ is the optimal parameter introduced in the next section. This can be guaranteed by the test of another sampling (Algorithm 3). The intuition is as follows. The output of a compressed sensing subroutine always matches the true signal on $\mathcal{T}$. If the solution is "good" in the sense that $|\nu_* - \mathfrak{u}|$ is smaller than a critical value, then the recovered signal should be close to the true signal on another random sample set $\mathcal{T}_2$. If the solution is "bad", then the difference between the two signals will be very large on $\mathcal{T}_2$. The classical post-processing part of the algorithm is stated in Algorithm 2 with Algorithm 3 as a subroutine.

3.2 Statement of the main theorem

The error bound and runtime of our algorithm are given in this subsection. Suppose the original signal is

$$y_t^0 = \sum_{f\in\mathcal{F}} p_f e^{-\mathrm{i}2\pi ft}, \quad f = \frac{n_f+\nu_f}{N}, \tag{21}$$

[ii]Throughout this paper, we require that $|\nu| \leq 1/2$.

where n_f (ν_f) is the integer (decimal) part of frequency f. Given a Fourier matrix F_ν, y^0 can be uniquely decomposed as $y^0 = F_\nu(x_{\text{re}} + \mathrm{i}x_{\text{im}})$, $x_{\text{re}}, x_{\text{im}} \in \mathbb{R}^N$, and $F_\nu x_{\text{re}}$ ($\mathrm{i}F_\nu x_{\text{im}}$) is termed as the on-grid (off-grid) component of y^0 with respect to F_ν. Define

$$\mathfrak{u} \equiv \arg\min_{\nu} \|F_\nu x_{\text{im}}\|_2, \tag{22}$$

and denote the corresponding grid decomposition as

$$y^0 = F_{\mathfrak{u}}(x_{\text{on}} + \mathrm{i}x_{\text{off}}) \quad x_{\text{on}}, x_{\text{off}} \in \mathbb{R}^N, \tag{23}$$

which we call the optimal grid decomposition of y^0.

Let $\sigma_{\text{off}} \equiv \|F_{\mathfrak{u}} x_{\text{off}}\|_\infty$. We assume that the signal y_t^0 satisfies the following conditions:

$$\begin{gathered} \mathcal{F} = \mathcal{F}_{\text{dom}} \bigsqcup \mathcal{F}_{\text{res}}, \quad \min_{f \in \mathcal{F}_{\text{dom}}} p_f \geq p_{\min}, \quad \sum_{f \in \mathcal{F}_{\text{res}}} p_f = \mathcal{O}(\sigma_{\text{off}}), \\ |\mathcal{F}_{\text{dom}}| = S, \quad \min_{f_1 \neq f_2 \in \mathcal{F}_{\text{dom}}} |f_1 - f_2| > \frac{1}{N}. \end{gathered} \tag{24}$$

In contrast to other QPE algorithms [12, 13] where $p_0 > \frac{1}{2}$ is required, our algorithm outputs the dominant on-grid approximation of the signal ($y_t \approx \sum_n p_n e^{-\mathrm{i}2\pi nt/N}$), instead of the dominant single-frequency approximation of the signal ($y_t \approx p e^{-\mathrm{i}2\pi ft}, f \in [0,1)$). Hence, even if $p_0 < \frac{1}{2}$, as long as the dominant part is large enough, the ground-state energy can still be well estimated.

Even signals with no frequency gap can be put into this framework. Indeed, when $|f_1 - f_2| \ll N^{-1}$, the two frequencies can be replaced by a single frequency f_3, and the tiny frequency gap is absorbed into the off-grid component. If f_3 can be approximated with high accuracy, it can be used as an estimate for f_1 because $|f_1 - f_3| \ll N^{-1}$.

In Appendix B.2, we prove the following lemma.

Lemma 1. *Given a length-N signal y^0 with optimal grid decomposition $y^0 = F_{\mathfrak{u}}(x_{\text{on}} + \mathrm{i}x_{\text{off}}), \sigma_{\text{off}} \equiv \|F_{\mathfrak{u}} x_{\text{off}}\|_\infty$. If y^0 satisfies the assumptions in Eq. (24), then for all $f \in \mathcal{F}_{\text{dom}}$, $|p_f(\nu_f - \mathfrak{u})| = \mathcal{O}(\sigma_{\text{off}})$.*

From Lemma 1, we know that if σ_{off} is small, then all f with large enough p_f are approximately on-grid simultaneously.

Define

$$\mathcal{D} \equiv \left\{ n : \exists f \in \mathcal{F}_{\text{dom}}, |Nf - n| \leq \frac{1}{2} \right\}. \tag{25}$$

Note that $|\mathcal{D}| = |\mathcal{F}_{\text{dom}}| = S$. In the next lemma, we show that the assumptions in Eq. (24) ensure that $\mathcal{D}$ is the dominant part of x_{on} as well. The proof is presented in Appendix B.3.

Lemma 2. *Suppose that the length-N ($N \geq 100$) signal satisfies Eq. (24). Then for any $f \in \mathcal{F}_{\text{dom}}, |Nf - n| \leq 1/2$, we have*

$$x_{\text{on},n} = p_f - \mathcal{O}(S\sigma_{\text{off}}), \tag{26}$$

and for all $n \notin \mathcal{D}$, we have

$$x_{\text{on},n} = \mathcal{O}(S\sigma_{\text{off}}). \tag{27}$$

If $\mathcal{D}$ can be recovered using Algorithm 2, and $\mathfrak{u}$ is estimated by ν_*, then all $f = (n+\mathfrak{u})/N \in \mathcal{F}_{\text{dom}}$ are estimated by $(n+\nu_*)/N$, and their error bounds are all $\mathcal{O}(|\nu_* - \mathfrak{u}|/N)$. The formal error bound of our protocol is guaranteed by the following arguments.

- Let ν_1 be the parameter in the trial set $\mathcal{V}$ that is closest to $\mathfrak{u}$, hence $|\nu_1 - \mathfrak{u}| = \mathcal{O}(|\mathcal{V}|^{-1})$. If $|\mathcal{V}|^{-1}$ is small enough, then all dominant frequencies of x_{on} are preserved in s_{ν_1}.
- Recall that the solution with the smallest 1-norm is s_{ν_*}. Using the test of another sampling, we can obtain a rough bound for $|\nu_* - \mathfrak{u}|$. Then we can prove that s_{ν_*} is close enough to s_{ν_1}, so that all the dominant frequencies of s_{ν_1} are preserved in s_{ν_*}.
- Because all dominant frequencies of x_{on} are preserved in s_{ν_*}, the accuracy of the final result is $\mathcal{O}(|\nu^* - \mathfrak{u}|/N)$.

Combining these arguments together, we obtain the following theorem regarding the accuracy and complexity of our algorithm. The formal proof can be found in Appendix B.

Theorem 2. *Suppose y^0 is a length-N ($N \geq 100$) signal with optimal grid decomposition $y^0 = F_{\mathfrak{u}}(x_{\text{on}} + \mathrm{i}x_{\text{off}}), \sigma_{\text{off}} \equiv \|F_{\mathfrak{u}} x_{\text{off}}\|_\infty$, and satisfies Eq. (24). If $\|x_{\text{on}}\|_2 \gg \|x_{\text{on}}\|_1/\sqrt{N}$, $E_0\tau/2\pi \in \mathcal{F}_{\text{dom}}$, and the parameters in Algorithm 2 satisfy*

$$4\pi\sqrt{S} < \sqrt{3}\log N, \quad r = \mathcal{O}(N^{-1}S\log^2(S)\,\text{poly}\log N), \quad \delta^{-1} = \mathcal{O}(\text{poly}\log N), \tag{28}$$

and $\exists \sigma_0 > 0$ such that

$$\begin{aligned} p_{\min} \gg S\sigma_0, \quad J \geq \left\lceil \frac{\log N}{\sqrt{S}\sigma_0} \right\rceil, \quad \sigma_0 \geq \sigma_{\text{off}}, \quad \sigma_0 \geq \sigma_{\text{H}}, \\ \sigma = \mathcal{O}(\sigma_0) \geq 2\sigma_0 \left(1 - \frac{4\pi\sqrt{S}}{\log N\sqrt{3}}\right)^{-1}, \end{aligned} \tag{29}$$

then there exists $\sigma_{\text{test}} = \mathcal{O}(\sigma_0)$ such that with probability at least $1 - 1/\text{poly}(N)$, the output of Algorithm 2 (ν_*, s_{ν_*}, E^*) satisfies

$$|\nu_* - \mathfrak{u}| = \mathcal{O}(\sigma_0), \quad \|s_{\nu_*} - x_{\text{on}}\|_2 = \mathcal{O}(\sigma_0), \quad |E^* - E_0| = \mathcal{O}\left(\frac{\sigma_0}{N p_{\min}}\right). \tag{30}$$

Note that the parameter σ_0 is essentially determined by σ_{off}. If $\sigma_{\text{off}} = 0$, meaning the y^0 is on-grid, then we can choose $\sigma_0 \to 0$, so that the error bound can approach 0 as well.

Remark. *The algorithm in [40] solves the compressed sensing subroutine within 1-norm error $\mathcal{O}(1)$ with runtime $\mathcal{O}(N \,\text{poly}\log N)$. In our numerical tests, we simply use the 'Minimize' subroutine in CVXPY.*

4 Numerical results

4.1 Previous algorithms

We begin with a brief review of three different early fault-tolerant QPE algorithms: ML-QCELS [12], MM-QCELS [16], and QMEGS [18], with the latter two algorithms applicable to QEEP.

The outlines of the first two algorithms are as follows. The ML-QCELS algorithm features a hierarchical structure. In each hierarchy, Hadamard tests are used to estimate the signal $y(t)$ at N_0 different time points. The algorithm then outputs an estimate for the dominant frequency by minimizing the cost function:

$$L(r, E) = \frac{1}{N_0} \sum_{n=1}^{N_0} \left| re^{-\mathrm{i}En\tau} - y(n\tau) \right|^2. \tag{31}$$

In the subsequent hierarchy, the search is refined to a narrower region, yielding a new estimate. This iterative process ultimately produces an accurate estimate of the dominant frequency. While ML-QCELS has proven effective for single-phase estimation, it is not suitable for multiple-phase estimation.

To address this limitation, the authors later introduced a multiple-phase version called MM-QCELS [16]. This algorithm modifies the original ML-QCELS in two significant ways: the time samples $\mathcal{T}$ are drawn from a probability distribution, and the cost function is redefined as

$$L\left(\{r_k, E_k\}_{k=1}^K\right) = \frac{1}{|\mathcal{T}|}\sum_{t\in\mathcal{T}}\left|\sum_{k=1}^K r_k e^{-\mathrm{i}E_k t} - y(t)\right|^2. \tag{32}$$

Additionally, when applying MM-QCELS to single-phase estimation, the hierarchical structure can be eliminated, rendering the algorithm non-adaptive.

A more recent algorithm, QMEGS [18], also samples time from a continuous probability distribution. Instead of estimating the frequencies by minimizing $L\left(\{r_k, E_k\}_{k=1}^K\right)$, the algorithm identifies the dominant frequency of the target signal by solving

$$\min_{f\in[0,1)}\sum_{t\in\mathcal{T}}\left|y(t)e^{\mathrm{i}2\pi f t}\right|^2. \tag{33}$$

Let the solution of this optimization be f^*. In the next step, the algorithm narrows its search to the region $[0,1]/(f^*-\delta f, f^*+\delta f)$ to estimate the sub-dominant frequency. By iterating this procedure, all dominant frequencies can eventually be determined.

Compressed sensing, in essence, is similar to non-adaptive MM-QCELS. Both methods aim to fit the sampled data using a signal ansatz. However, they differ in sampling strategies and optimization objectives. In MM-QCELS, times are sampled from a continuous probability distribution, and the cost function minimizes the total empirical error in the time domain. In contrast, compressed sensing samples times uniformly at random from a discrete set, and the cost function minimizes the 1-norm of the signal in the frequency domain.

4.2 Models and results

In this subsection, we present several numerical tests and compare with previous works.

Given a Hamiltonian H, we first renormalize it to

$$\overline{H} = \frac{\pi H}{4\|H\|_2} \tag{34}$$

so that the spectra of $\overline{H}$ belong to $[-\pi/4, \pi/4]$. We further shift the Hamiltonian to $\overline{H}' = \overline{H} + \pi/2$ so that the spectrum of $\overline{H}'$ belong to $[\pi/4, 3\pi/4] \subset [0, 2\pi]$, as assumed in the initial setup of our algorithm. To have a better control of the parameters in the signal, we design the following family of initial states: choose a parameter $\alpha \in (0,1)$, and set the initial state as

$$|\Psi_\alpha\rangle \propto \sum_{\ell=0}^9 \sqrt{\alpha^\ell}|\phi_\ell\rangle, \tag{35}$$

where $|\phi_\ell\rangle$ is the ℓ-th eigenstate of $\overline{H}$ with eigenvalue E_ℓ. Then, the target signal is as follows:

$$y_n = \sum_{\ell=0}^9 p_\ell e^{-\mathrm{i}E_\ell n\tau}, \quad p_\ell = \frac{(1-\alpha)\alpha^\ell}{1-\alpha^{10}}. \tag{36}$$

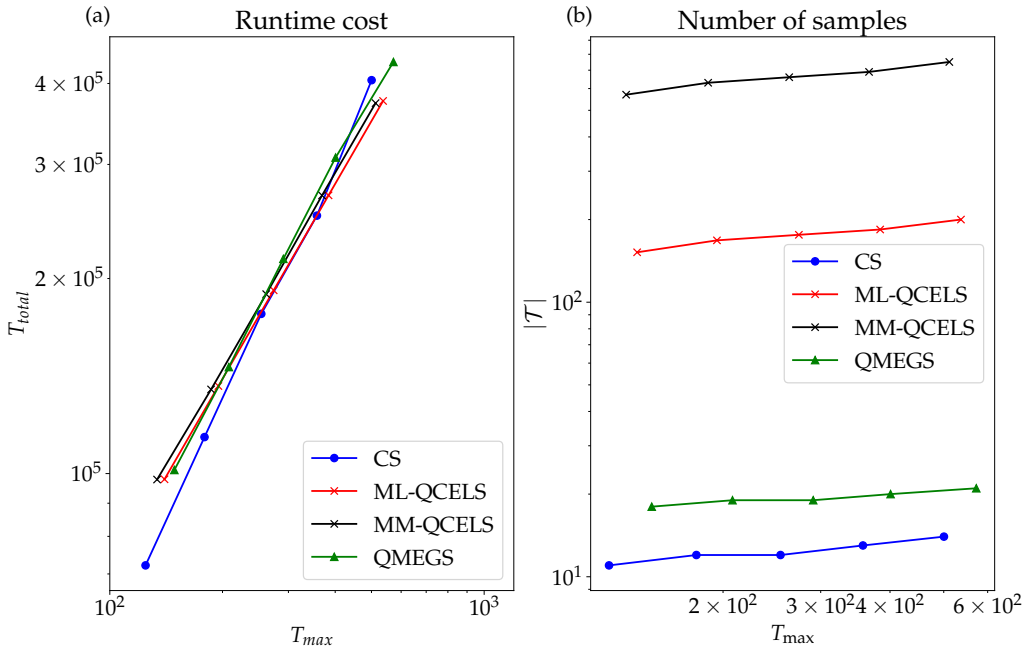

Figure 2: Comparison between the setups of the four algorithms. Here we set $T_n = \lfloor 100 \times (1.4)^n \rfloor, n = 1, 2, 3, 4, 5$. In Algorithm 2, we set $S = 1, r = 2.3 \ln T_n / T_n, \tau = 1, \sigma = 0.2\sqrt{2.3 \ln T_n}, J = 100, M_{\mathrm{H}} = 100$ and let $\mathcal{K} = \{\arg\max s_{\nu_*}\}$. In ML-QCELS, we set $N = 8, N_s = 50, J = \lfloor 4 \ln T_n \rfloor$ (the meanings of the parameters can be found in [12]). In MM-QCELS, we set $K = 2, N_T = 30, \gamma = 1, N_s = 100, J = \lfloor 4 \ln T_n \rfloor$ (the meanings of the parameters can be found in [16]). In QMEGS, we set $K = 10, \mathrm{dx} = 10^{-4}, \alpha = 5, N = 10 + 2\lfloor \ln(2T_n) \rfloor$ (the meanings of the parameters can be found in [18]). The codes are available online [41].

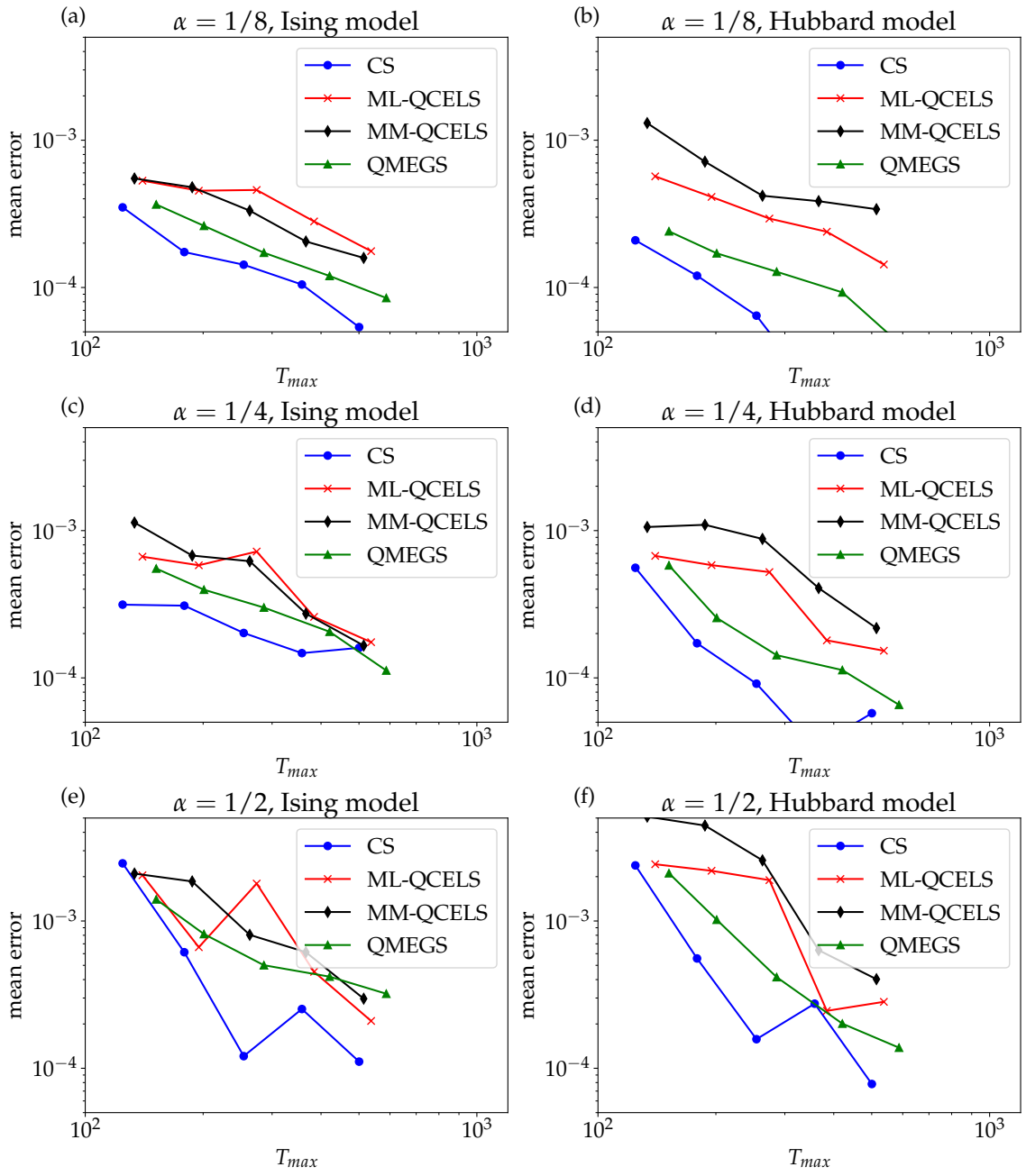

Figure 3: Comparison between Algorithm 2, ML-QCELS, MM-QCELS and QMEGS with respect to signals generated by Ising model (Eq. (37), left column) or Hubbard model (Eq. (38), right column). Subfigures (a),(b) represent $\alpha = 1/8$ ($p_0 \approx 0.88$); (c),(d) represent $\alpha = 1/4$ ($p_0 \approx 0.75$); (e),(f) represent $\alpha = 1/2$ ($p_0 \approx 0.5$).

For example, when $\alpha = 1/2$, we have $p_0 \approx 1/2$. Clearly, p_0 decreases with α.

In the first set of numerical tests, H is the normalized Transverse Field Ising (TFI) model on 8 sites:

$$H_{\mathrm{TFI}} = -\sum_{j=1}^{7} Z_j Z_{j+1} - Z_8 Z_1 - 4\sum_{j=1}^{8} X_j. \tag{37}$$

In the second set of numerical tests, H is the Fermi-Hubbard (FH) model on 4 sites:

$$H_{\mathrm{FH}} = -\sum_{j=1}^{3}\sum_{\sigma=\uparrow,\downarrow} c_{j,\sigma}^{\dagger} c_{j+1,\sigma} + 10\sum_{j=1}^{4}\left(n_{j,\uparrow} - \frac{1}{2}\right)\left(n_{j,\downarrow} - \frac{1}{2}\right). \tag{38}$$

For each Hamilonian, we set $\alpha = 1/8, 1/4, 1/2$ separately.

To have a rather fair comparison between the algorithms, we deliberately choose the parameters to ensure that the runtimes $(T_{\max}, T_{\mathrm{total}})$ of the algorithms are approximately on the same line, which is demonstrated in Fig. 2. In the same figure, we also plot the number of distinct t used in these algorithms. As shown in Fig. 2, our algorithm requires a much smaller size of time samples $|\mathcal{T}|$, in contrast with that of ML-QCELS and MM-QCELS. The mean errors of the outputs are shown in Fig. 3. When the initial overlap is comparably large ($\alpha = 1/8$), the performance of compressed sensing based algorithm is better than the other methods. This observation supports the claim that Algorithm 2 is prominent in its sparse sampling of time and high level of accuracy.

5 Discussions

In this paper, we presented a simple and robust algorithm for QPE using compressed sensing. For the single eigenvalue estimation (i.e., QPE), we rigorously established its Heisenberg-limit scaling in Theorem 2 and numerically demonstrated its performance compared to other state-of-the-art QPE algorithms in Sec. 4. Our algorithm has a smaller average error when the initial overlap is large, provided that the runtime costs $(T_{\mathrm{total}}, T_{\max})$ are approximately the same. Similarly to QMEGS, our algorithm is non-adaptive, which means we can perform all the measurements first, and then focus on the classical post-processing part. As a comparison, RPE-inspired algorithms are usually adaptive [12, 13]. Our algorithm requires a rather small size of $\mathcal{T}$ on a discrete set. In Fig. 3, for a signal of length $N \in (100, 600)$, our algorithm only requires $\approx 10^1$ different time samples while MM-QCELS requires $\approx 10^2$ number of time samples from a continuous region. Our numerical tests also show that QMEGS only requires $\approx 10^1$ different samples to obtain an accurate estimation. As mentioned in the end of Sec. 2.3, there is a close connection between QMEGS and algorithm OMP, and we conjecture that it is possible to analyze QMEGS using the framework of compressed sensing.

Lastly, we list a few open questions:

- In discrete sampling protocols, would it be possible to shorten the maximal runtime by biased sampling of times? What is the limitation in the discrete scenario? Can we achieve a similar improvement to the Gaussian filter method in [18]?
- In our numerical experiments, the test of another sampling (Algorithm 3) is actually unnecessary. Is it possible to show this analytically as well?
- It is possible to find the optimal grid shift parameter by optimization instead of doing brute-force search in a trial set, which should significantly improve our current result.

6 Acknowledgement

We thank Tianyu Wang for the helpful discussions. C.Y. acknowledges support from the National Natural Science Foundation of China (Grant No. 92165109), National Key Research and Development Program of China (Grant No. 2022YFA1404204), and Shanghai Municipal Science and Technology Major Project (Grant No. 2019SHZDZX01). C.Z. and J.T. acknowledge support from the U.S. National Science Foundation under Grant No. 2116246, the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, and Quantum Systems Accelerator.

References

[1] Alexei Y Kitaev. "Quantum measurements and the Abelian stabilizer problem". quant-ph/9511026 (1995).

[2] Peter W Shor. "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer". SIAM review **41**, 303–332 (1999).

[3] Daniel S Abrams and Seth Lloyd. "Quantum algorithm providing exponential speed increase for finding eigenvalues and eigenvectors". Phys. Rev. Lett. **83**, 5162 (1999).

[4] Sam McArdle, Suguru Endo, Alán Aspuru-Guzik, Simon C Benjamin, and Xiao Yuan. "Quantum computational chemistry". Rev. Mod. Phys. **92**, 015003 (2020).

[5] Lin Lin and Yu Tong. "Heisenberg-limited ground-state energy estimation for early fault-tolerant quantum computers". PRX Quantum **3**, 010318 (2022).

[6] Guoming Wang, Daniel Stilck-França, Ruizhe Zhang, Shuchen Zhu, and Peter D Johnson. "Quantum algorithm for ground state energy estimation using circuit depth with exponentially improved dependence on precision". Quantum **7**, 1167 (2023).

[7] Rolando D Somma. "Quantum eigenvalue estimation via time series analysis". New J. Phys. **21**, 123025 (2019).

[8] Thomas E O'Brien, Brian Tarasinski, and Barbara M Terhal. "Quantum phase estimation of multiple eigenvalues for small-scale (noisy) experiments". New J. Phys. **21**, 023022 (2019).

[9] Ruizhe Zhang, Guoming Wang, and Peter Johnson. "Computing ground state properties with early fault-tolerant quantum computers". Quantum **6**, 761 (2022).

[10] Alicja Dutkiewicz, Barbara M Terhal, and Thomas E O'Brien. "Heisenberg-limited quantum phase estimation of multiple eigenvalues with few control qubits". Quantum **6**, 830 (2022).

[11] Michael A Nielsen and Isaac L Chuang. "Quantum computation and quantum information". Cambridge university press. (2010).

[12] Zhiyan Ding and Lin Lin. "Even shorter quantum circuit for phase estimation on early fault-tolerant quantum computers with applications to ground-state energy estimation". PRX Quantum **4**, 020331 (2023).

[13] Hongkang Ni, Haoya Li, and Lexing Ying. "On low-depth algorithms for quantum phase estimation". Quantum **7**, 1165 (2023).

[14] Iulia M Georgescu, Sahel Ashhab, and Franco Nori. "Quantum simulation". Rev. Mod. Phys. **86**, 153 (2014).

[15] Andrew M Childs, Yuan Su, Minh C Tran, Nathan Wiebe, and Shuchen Zhu. "Theory of Trotter error with commutator scaling". Phys. Rev. X **11**, 011020 (2021).

[16] Zhiyan Ding and Lin Lin. "Simultaneous estimation of multiple eigenvalues with short-depth quantum circuit on early fault-tolerant quantum computers". Quantum **7**, 1136 (2023).

[17] Haoya Li, Hongkang Ni, and Lexing Ying. "Adaptive low-depth quantum algorithms for robust multiple-phase estimation". Phys. Rev. A **108**, 062408 (2023).
[18] Zhiyan Ding, Haoya Li, Lin Lin, HongKang Ni, Lexing Ying, and Ruizhe Zhang. "Quantum Multiple Eigenvalue Gaussian filtered Search: an efficient and versatile quantum phase estimation method". Quantum **8**, 1487 (2024).
[19] Itai Arad, Tomotaka Kuwahara, and Zeph Landau. "Connecting global and local energy distributions in quantum spin models on a lattice". J. Stat. Mech. Theor. Exp. **2016**, 033301 (2016).
[20] Andrew M Childs and Yuan Su. "Nearly optimal lattice simulation by product formulas". Phys. Rev. Lett. **123**, 050503 (2019).
[21] Haitham Hassanieh, Piotr Indyk, Dina Katabi, and Eric Price. "Nearly optimal sparse Fourier transform". In Proceedings of the forty-fourth annual ACM symposium on Theory of computing. Pages 563–578. (2012).
[22] Wenjing Liao and Albert Fannjiang. "MUSIC for single-snapshot spectral estimation: Stability and super-resolution". Appl. Comput. Harmon. Anal. **40**, 33–67 (2016).
[23] Zhao Song, Baocheng Sun, Omri Weinstein, and Ruizhe Zhang. "Quartic samples suffice for Fourier interpolation". In 2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS). Pages 1414–1425. IEEE (2023).
[24] Zhiyan Ding, Ethan N Epperly, Lin Lin, and Ruizhe Zhang. "The ESPRIT algorithm under high noise: Optimal error scaling and noisy super-resolution". In 2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS). Pages 2344–2366. IEEE (2024).
[25] William T Cochran, James W Cooley, David L Favin, Howard D Helms, Reginald A Kaenel, William W Lang, George C Maling, David E Nelson, Charles M Rader, and Peter D Welch. "What is the fast Fourier transform?". Proceedings of the IEEE **55**, 1664–1674 (1967).
[26] Anna C Gilbert, Shan Muthukrishnan, and Martin Strauss. "Improved time bounds for near-optimal sparse Fourier representations". In Manos Papadakis, Andrew F. Laine, and Michael A. Unser, editors, Wavelets XI. Volume 5914, page 59141A. International Society for Optics and PhotonicsSPIE (2005).
[27] Piotr Indyk, Michael Kapralov, and Eric Price. "(Nearly) sample-optimal sparse Fourier transform". In Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms. Pages 480–499. SIAM (2014).
[28] Brendon L Higgins, Dominic W Berry, Stephen D Bartlett, Morgan W Mitchell, Howard M Wiseman, and Geoff J Pryde. "Demonstrating Heisenberg-limited unambiguous phase estimation without adaptive measurements". New J. Phys. **11**, 073023 (2009).
[29] Shelby Kimmel, Guang Hao Low, and Theodore J Yoder. "Robust calibration of a universal single-qubit gate set via robust phase estimation". Phys. Rev. A **92**, 062315 (2015).
[30] Federico Belliardo and Vittorio Giovannetti. "Achieving Heisenberg scaling with maximally entangled states: An analytic upper bound for the attainable root-mean-square error". Phys. Rev. A**102** (2020).
[31] T Tony Cai and Lie Wang. "Orthogonal matching pursuit for sparse signal recovery with noise". IEEE Transactions on Information theory **57**, 4680–4688 (2011).
[32] Emmanuel J Candès and Terence Tao. "Near-optimal signal recovery from random projections: Universal encoding strategies?". IEEE transactions on information theory **52**, 5406–5425 (2006).

[33] Emmanuel J Candès. "The restricted isometry property and its implications for compressed sensing". Comptes rendus. Mathematique **346**, 589–592 (2008).
[34] Emmanuel J Candès, Justin Romberg, and Terence Tao. "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information". IEEE Transactions on information theory **52**, 489–509 (2006).
[35] David Gross, Yi-Kai Liu, Steven T Flammia, Stephen Becker, and Jens Eisert. "Quantum state tomography via compressed sensing". Phys. Rev. Lett. **105**, 150401 (2010).
[36] Easwar Magesan, Alexandre Cooper, and Paola Cappellaro. "Compressing measurements in quantum dynamic parameter estimation". Phys. Rev. A **88**, 062109 (2013).
[37] Aaron Smith, Riofrío Carlos, Brielle Evelyn Anderson, Hector Sosa Martinez, Ivan H Deutsch, and Poul Jessen. "Quantum state tomography by continuous measurement and compressed sensing". Phys. Rev. A **87**, 030102 (2013).
[38] Amir Kalev, Robert L Kosut, and Ivan H Deutsch. "Quantum tomography protocols with positivity are compressed sensing protocols". Npj Quantum Inf. **1**, 15018 (2015).
[39] Gongguo Tang, Badri Narayan Bhaskar, Parikshit Shah, and Benjamin Recht. "Compressed sensing off the grid". IEEE transactions on information theory **59**, 7465–7490 (2013).
[40] Juditsky Anatoli, Kilinc Karzan Fatma, and Nermirovski Arkadi. "Randomized first order algorithms with applications to ℓ_1-minimization". Math. Program. **142**, 269–310 (2013).
[41] `https://github.com/CYI1995/QEEP/tree/main/Paper_QPE`.
[42] Mark Rudelson and Roman Vershynin. "On sparse reconstruction from Fourier and Gaussian measurements". Comm. Pure Appl. Math. **61**, 1025–1045 (2008).

A Standard results in compressed sensing

Given a vector $v = [v_1, v_2, \cdots, v_N]^\top$, its p-norm is defined as

$$\|v\|_p \equiv \left(\sum_{n=1}^{N} |v_n|^p\right)^{1/p}. \tag{39}$$

Note that $\|v\|_\infty = \max_n |v_n|$ and $\|v\|_0$ is the sparsity of v. In the following paragraph, we use k to label the indices of entries in the frequency domain, and use n to label the indices of entries in the time domain. Denote the set of integers from 1 to N as $[N]$. In regular compressed sensing, we deal with a time domain discrete signal y^0 in the form of

$$y_n^0 = \sum_{f\in\mathcal{F}} p_f e^{-\mathrm{i}2\pi f n} \quad \forall n \in [N], \tag{40}$$

where $p_f > 0, \sum_f p_f = 1, f \in [0, 1)$, and $\mathcal{F}$ is the set of frequencies. In the context of compressed sensing, sparsity means $|\mathcal{F}| = \mathcal{O}(\log N)$ [32]. The time domain signal can thus be written as an N-dimensional vector

$$y^0 = [y_1^0, y_2^0, \cdots, y_N^0]^\top. \tag{41}$$

Define the Fourier matrix by $F_{kn} \equiv e^{-\mathrm{i}2\pi kn/N}$ with $k, n \in [N]$. If all $f \in \mathcal{F}$ are on-grid, the frequency domain signal x can be written in the form of a real vector:

$$x = \frac{1}{N} F^\dagger y^0 = \sum_{f\in\mathcal{F}} p_f \delta_{Nf}. \tag{42}$$

The purpose of compressed sensing is to recover x from noisy signal samples. The algorithm is accomplished in the following sequence. Choose a sampling ratio $r \in (0, 1)$, and assign each integer n in $[N]$ a random variable 1_n such that

$$\Pr\{1_n = 1\} = r, \quad \Pr\{1_n = 0\} = 1 - r. \tag{43}$$

Draw one sample from each 1_n, and denote the set of integers with sampled value 1 as the sample set $\mathcal{T}$. Given $\mathcal{T}$, we define the projection operator $\mathcal{P}_\mathcal{T}$ by

$$(\mathcal{P}_\mathcal{T})_{t_1,t_2} = 1_{t_1\in\mathcal{T}} \cdot \delta_{t_1,t_2} \quad t_1, t_2 \in [N], \tag{44}$$

and $F_\mathcal{T} \equiv \mathcal{P}_\mathcal{T} F, y_\mathcal{T}^0 \equiv \mathcal{P}_\mathcal{T} y^0$.

With these notations, the compressed sensing subroutine is to solve the following optimization problem

$$\min_{s\in\mathbb{R}^N} \|s\|_1, \quad \text{s.t.} \quad F_\mathcal{T} s = y_\mathcal{T}^0, \tag{45}$$

which can be rewritten as a linear programming problem. When $|\mathcal{T}| \approx Nr = \mathcal{O}(\mathrm{poly}\log N)$ and the frequency support $\mathcal{F}$ is sparse in the sense that $|\mathcal{F}| = \mathcal{O}(\mathrm{poly}\log N)$, the optimal solution $s^\#$ is equal to the frequency domain signal x with high probability. Rigorous statements can be found in [32].

Provided that the signal has extra noise $y_n = y_n^0 + z_n$, then the signal can be approximately recovered by the convex relaxation algorithm [33]:

$$\min_{s\in\mathbb{R}^N} \|s\|_1, \quad \text{s.t.} \quad \|F_\mathcal{T} s - y_\mathcal{T}\|_2 \le \sqrt{|\mathcal{T}|}\sigma, \tag{46}$$

where σ is the expected mean-square-root noise. The differences between the solution to Eq. (46) and x depend on σ. The subroutine itself is a convex quadratic programming problem.

The robustness of compressed sensing solution can be analyzed through the restricted isometry property (RIP) of random Fourier matrices. If a matrix M satisfies $(1-\eta)\|x\|_2^2 \leq \|Mx\|_2^2 \leq (1+\eta)\|x\|_2^2$ for all $x \in \mathcal{X}$, then we say that the matrix M satisfies η-RIP over set $\mathcal{X}$. We present a few standard results of compressed sensing in the following.

Theorem 3. *[42] Suppose $\mathcal{T}$ is generated by sampling from $[N]$ with ratio*

$$r = \mathcal{O}\left(S\log^2(S)\eta^{-2}\text{poly}\log(N)N^{-1}\right). \tag{47}$$

Then $F_{\mathcal{T}}/\sqrt{|\mathcal{T}|}$ satisfies η-RIP over the set $\{x : x \in \mathbb{C}^N, \|x\|_0 \leq S\}$ with probability

$$1 - \mathcal{O}\left(\frac{\text{poly}(Se^{\eta^{-1}})}{\text{poly}(N)}\right). \tag{48}$$

Proof. Follow Theorem 3.11 of [42]. The theorem can be proved by setting $s = \mathcal{O}(\log N)$ and $\varepsilon = \mathcal{O}(\eta/\log N)$. The meaning of the notations can be found in the same paper. □

Given a vector $v \in \mathbb{R}^N$, we rearrange the entries of the vector by their absolute values in the non-increasing order to obtain $(v_1^{\downarrow}, v_2^{\downarrow}, \cdots, v_N^{\downarrow})$. Define v_{res} as the vector generated from v by removing $v_1^{\downarrow}, v_2^{\downarrow}, \ldots, v_S^{\downarrow}$. The accuracy of Eq. (46) can be estimated by the following theorem.

Theorem 4. *[33] Suppose $M \in \mathbb{C}^{N'\times N}$ satisfies η-RIP over the set $\{x : x \in \mathbb{R}^N, \|x\|_0 \leq 2S\}$, $0 < \eta < \sqrt{2}-1$, and $x_1, x_2 \in \mathbb{R}^N$. If*

$$\|x_1\|_1 \leq \|x_2\|_1, \quad \|M(x_1 - x_2)\|_2 \leq \sigma, \tag{49}$$

then

$$\|x_1 - x_2\|_2 \leq C_1\sigma + C_2\frac{\|x_{2,\text{res}}\|_1}{\sqrt{S}}, \tag{50}$$

where

$$C_1 \equiv \frac{2 + 2(\sqrt{2}-1)\eta}{1-(\sqrt{2}+1)\eta}, \quad C_2 \equiv \frac{4\sqrt{1+\eta}}{1-(\sqrt{2}+1)\eta}. \tag{51}$$

B Proof of Theorem 2

The input signal is

$$y = y_{\text{on}}^0 + y_{\text{off}}^0 + z, \tag{52}$$

where z is the uncertainty from the Hadamard tests that satisfies $\|z\|_\infty \leq \sigma_{\text{H}}$. Recall that s_ν is the solution of

$$\min_{s\in\mathbb{R}^N} \|s\|_1 \quad \text{s.t.} \quad \|F_{\nu,\mathcal{T}}s - y_{\mathcal{T}}\|_2 \leq \sqrt{|\mathcal{T}|}\sigma. \tag{53}$$

Let ν_1 be the parameter in $\mathcal{V}$ that is closest to $\mathfrak{u}$, and ν_* be the one with the smallest $\|s_{\nu_*}\|_1$. Therefore,

$$\begin{aligned}
\|F_{\nu_1,\mathcal{T}}(s_{\nu_*}-s_{\nu_1})\|_2 &\le \|F_{\nu_1,\mathcal{T}}s_{\nu_*}-y_{\mathcal{T}}\|_2+\|F_{\nu_1,\mathcal{T}}s_{\nu_1}-y_{\mathcal{T}}\|_2\\
&\le \|F_{\nu_1,\mathcal{T}}s_{\nu_*}-y_{\mathcal{T}}\|_2+\sqrt{|\mathcal{T}|}\sigma\\
&\le \|F_{\nu_1,\mathcal{T}}s_{\nu_*}-F_{\nu_*,\mathcal{T}}s_{\nu_*}\|_2+\|F_{\nu_*,\mathcal{T}}s_{\nu_*}-y_{\mathcal{T}}\|_2+\sqrt{|\mathcal{T}|}\sigma\\
&\le \|F_{\nu_1,\mathcal{T}}s_{\nu_*}-F_{\nu_*,\mathcal{T}}s_{\nu_*}\|_2+2\sqrt{|\mathcal{T}|}\sigma, \qquad (54)\\
\|F_{\nu_*,\mathcal{T}}s_{\nu_*}-F_{\nu_1,\mathcal{T}}s_{\nu_*}\|_2^2 &= \sum_{t\in\mathcal{T}}\left|\sum_{k\in[N]}s_{\nu_*,k}\left(e^{-\mathrm{i}2\pi(k+\nu_*)t/N}-e^{-\mathrm{i}2\pi(k+\nu_1)t/N}\right)\right|^2\\
&= \sum_{t\in\mathcal{T}}\left|\sum_{k\in[N]}s_{\nu_*,k}e^{-\mathrm{i}2\pi kt/N}\right|^2 4\sin^2\left[\frac{\pi(\nu_*-\nu_1)t}{N}\right]\\
&= \sum_{t\in\mathcal{T}}|(Fs_{\nu_*})_t|^2 4\sin^2\left[\frac{\pi(\nu_*-\nu_1)t}{N}\right]\\
&\le 4\pi^2|\mathcal{T}||\nu_*-\nu_1|^2 \quad \text{with high probability.} \qquad (55)
\end{aligned}$$

The last inequality holds with high probability because $|\mathcal{T}|=\mathcal{O}(\mathrm{poly}\log N)$.

Combining Eqs. (77) and (75) in Lemma 5, if we set

$$\sigma_{\mathrm{test}}=c\sqrt{\frac{3}{2}}\left[\left(C_3^2+\frac{8\pi^2S}{3\log^2 N}\right)^{1/2}+C_0\right]\sigma, \qquad (56)$$

where c is a small constant larger than 1, then with high probability, we have

$$|\nu_*-\mathfrak{u}|\le C[x_{\mathrm{on}}]^{-1}(\sqrt{2}\sigma_{\mathrm{test}}+C_0\sigma)=\mathcal{O}(\sigma). \qquad (57)$$

The condition $\|x_{\mathrm{on}}\|_2\gg\|x_{\mathrm{on}}\|_1/\sqrt{N}$ ensures that $C[x_{\mathrm{on}}]=\mathcal{O}(1)$. Therefore,

$$\|F_{\nu_1,\mathcal{T}}(s_{\nu_*}-s_{\nu_1})\|_2\le 2\sqrt{|\mathcal{T}|}\sigma+2\pi|\nu_1-\mathfrak{u}|+2\pi|\nu_*-\mathfrak{u}|:=\sqrt{|\mathcal{T}|}C_4\sigma_0. \qquad (58)$$

The condition on J in Eq. (29) gives us

$$|\nu_1-\mathfrak{u}|\le J^{-1}\le\frac{\sqrt{S}\sigma_0}{\log N}. \qquad (59)$$

The conditions in Eq. (29) implies that $\sqrt{S}/\log N=\mathcal{O}(1), C_0=\mathcal{O}(1), C_3=\mathcal{O}(1)$, which further implies $C_4=\mathcal{O}(1)$.

Consider the grid decomposition of y_{on}^0 with respect to ν:

$$x_\nu^{\mathrm{R}}\equiv\mathrm{Re}(F_\nu^{-1}y_{\mathrm{on}}^0),\quad x_\nu^{\mathrm{I}}\equiv\mathrm{Im}(F_\nu^{-1}y_{\mathrm{on}}^0), \qquad (60)$$

Note that $x_{\mathfrak{u}}^{\mathrm{R}}=x_{\mathrm{on}}, x_{\mathfrak{u}}^{\mathrm{I}}=0$. By virtue of Theorem 4, we have

$$\|s_{\nu_*}-s_{\nu_1}\|_2\le C_1C_4\sigma+C_2\frac{\|s_{\nu_1,\mathrm{res}}\|_1}{\sqrt{S}}, \qquad (61)$$

where we have used the fact that $C_1, C_2, C_4=\mathcal{O}(1)$. The condition on $p_{\min}$ ensures that the dominant part of s_{ν_1} is the same as the dominant part of x_{on}. By virtue of Lemma 4,

we have $\|s_{\nu_1}\|_1 \leq \|x^{\mathrm{R}}_{\nu_1}\|_1$. Therefore,

$$\begin{aligned}\|s_{\nu_1,\mathrm{res}}\|_1 &\leq \sum_{n\in\mathcal{D}} \left(|x^R_{\nu_1}| - |s_{\nu_1,n}|\right) + \sum_{n\in\mathcal{D}_{\mathrm{res}}} |x^{\mathrm{R}}_{\nu_1,n}| \\ &\leq \|x^{\mathrm{R}}_{\nu_1} - s_{\nu_1}\|_1 + \pi^2|\nu_1 - \mathfrak{u}|\log N \quad \text{(Lemma 3)} \\ &\leq \sqrt{S}(C_1 + C_2\pi)\sigma_{\mathrm{off}} + \sqrt{S}\pi^2\sigma_0 = \mathcal{O}(\sqrt{S}\sigma_0). \end{aligned} \tag{62}$$

Hence $\|s_{\nu_*} - s_{\nu_1}\|_2 = \mathcal{O}(\sigma_0)$. According to Lemma 4, we have $\|s_{\nu_1} - x_{\mathrm{on}}\|_2 \leq C_3\sigma = \mathcal{O}(\sigma_0)$. Therefore,

$$\|s_{\nu_*} - x_{\mathrm{on}}\|_2 \leq \|s_{\nu_*} - s_{\nu_1}\|_2 + \|s_{\nu_1} - x_{\mathrm{on}}\|_2 = \mathcal{O}(\sigma_0). \tag{63}$$

Finally, the accuracy on E_0 can be confirmed by combining Lemma 2 and the above 2-norm error bound.

B.1 Auxiliary lemmas

The following lemmas are critical for the proof of Theorem 2. Henceforth, we assume s_ν is the solution of

$$\min_{s\in\mathbb{R}^N} \|s\|_1, \quad \text{s.t.} \quad \|F_{\mathcal{T}}s - y_{\mathcal{T}}\|_2 \leq \sqrt{|\mathcal{T}|}\sigma. \tag{64}$$

Consider a signal $x = \sum_{n\in\mathcal{D}'} q_n\delta_n$ that follows the description of Lemma 2. Suppose $p_{\min} \gg S\sigma_{\mathrm{off}}$. Then $\mathcal{D}$ is the indices corresponding to the S largest entries of x, and the remaining part is $\mathcal{D}_{\mathrm{res}}$, and

$$x_{\mathrm{dom}} = \sum_{n\in\mathcal{D}} q_n\delta_n, \quad x_{\mathrm{res}} = \sum_{n\in\mathcal{D}_{\mathrm{res}}} q_n\delta_n, \quad |\mathcal{D}| = S, \quad \|x_{\mathrm{res}}\|_1 = \mathcal{O}(\sigma_{\mathrm{off}}). \tag{65}$$

Lemma 3. *Suppose $y = Fx$ is a length-N ($N \geq 100$) on-grid signal and x satisfies the condition in Eq. (65) and $\|x\|_1 \leq 1$. Let*

$$C[x] \equiv \left[(4 + 2\pi^2/N)\|x\|_2^2 - 2\pi^2\|x\|_1^2/N\right]^{1/2}. \tag{66}$$

If $y = F_\nu(x^{\mathrm{R}}_\nu + \mathrm{i}x^{\mathrm{I}}_\nu)$, then

$$C[x]|\nu| \leq \|x^{\mathrm{I}}_\nu\|_2 \leq \frac{2\pi}{\sqrt{3}}|\nu|, \tag{67}$$

$$\|x^{\mathrm{R}}_\nu - x\|_2 \leq \frac{2\pi}{\sqrt{3}}|\nu|, \tag{68}$$

$$\sum_{n\in\mathcal{D}_{\mathrm{res}}} |x^{\mathrm{R}}_{\nu,n}| \leq \pi^2|\nu|\log N. \tag{69}$$

The proof of Lemma 3 is presented in Appendix C.2, which is a direct corollary of Lemma 10 and Lemma 11.

In Appendix B.4, we prove the following lemma.

Lemma 4 (A good ν generates a good solution)**.** *Suppose y^0 is a length-N signal with optimal grid decomposition $y^0 = F_{\mathfrak{u}}(x_{\mathrm{on}} + \mathrm{i}x_{\mathrm{off}}), \sigma_{\mathrm{off}} = \|F_{\mathfrak{u}}x_{\mathrm{off}}\|_\infty$; the target signal is $y = y^0 + z$ with $\|z\|_\infty \leq \sigma_{\mathrm{H}}$; let*

$$C_0 = (\sigma_{\mathrm{off}} + \sigma_{\mathrm{H}})/\sigma, \quad C_3 = 2C_1 + C_2\pi + \frac{2\pi}{\log N}\sqrt{\frac{S}{3}}; \tag{70}$$

$\mathcal{T}$ is an integer set generated by sampling from $[N]$ with ratio

$$r = \mathcal{O}\left(S\log^2(S)\mathrm{poly}\log(N)N^{-1}\right), \tag{71}$$

and s_ν is the solution of Eq. (53). If

$$|\nu| \le \frac{\sqrt{S}\sigma}{\log N}, \quad C_0 \le 1 - \frac{4\pi}{\log N}\sqrt{\frac{S}{3}}, \tag{72}$$

then and with probability at least $1 - 1/\mathrm{poly}(N)$, we have

$$\|s_\nu\|_1 \le \|x_\nu^{\mathrm{R}}\|_1, \quad \|s_\nu - x_{\mathrm{on}}\|_2 \le C_3\sigma. \tag{73}$$

In Appendix B.5, we prove the following lemma.

Lemma 5 (A bad ν generates a bad solution)**.** *Suppose y^0 is a length-N signal with optimal grid decomposition $y^0 = F_{\mathfrak{u}}(x_{\mathrm{on}} + \mathrm{i}x_{\mathrm{off}}), \sigma_{\mathrm{off}} = \|F_{\mathfrak{u}}x_{\mathrm{off}}\|_\infty$ and satisfies Eq. (24); the target signal is $y = y^0 + z$ with $\|z\|_\infty \le \sigma_{\mathrm{H}}$; $\mathcal{T}$ is an integer set generated by sampling from $[N]$ with ratio*

$$r = \mathcal{O}\left(S\log^2(S)\mathrm{poly}\log(N)N^{-1}\right), \tag{74}$$

and s_ν is the solution of Eq. (53); $\mathcal{I}$ is an integer set generated by sampling from $[N]$ with equal probability for L times; let σ_{test} be a constant with lower bound

$$\sigma_{\mathrm{test}} \ge \sqrt{\frac{3}{2}}\left[\left(C_3^2 + \frac{8\pi^2 S}{3\log^2 N}\right)^{1/2} + C_0\right]\sigma. \tag{75}$$

If $|\nu| \le \sqrt{S}\sigma/\log N$, then with probability at least $(1 - e^{-L/2})(1 - 1/\mathrm{poly}(N))$,

$$\sum_{t\in\mathcal{I}} |y_t - (F_\nu s_\nu)_t|^2 \le L\sigma_{\mathrm{test}}^2, \tag{76}$$

If instead

$$|\nu| > C[x_{\mathrm{on}}]^{-1}(\sqrt{2}\sigma_{\mathrm{test}} + C_0\sigma), \tag{77}$$

then with probability at least $(1 - e^{-L/2})(1 - 1/\mathrm{poly}(N))$, we have

$$\sum_{t\in\mathcal{I}} |y_t - (F_\nu s_\nu)_t|^2 > L\sigma_{\mathrm{test}}^2. \tag{78}$$

The list of constants used in the proof are as follows.

B.2 Proof of Lemma 1

Without loss of generality, we assume $\mathfrak{u} = 0$. After computation, we have

$$\begin{aligned}
y_{\mathrm{off},t} &= \begin{cases} -\mathrm{i}\sum_{f\in\mathcal{F}} p_f \sin(\pi\nu_f) e^{\mathrm{i}\pi\nu_f} e^{-\mathrm{i}2\pi ft} & t \ne 0 \\ 0 & t = 0 \end{cases} \\
y_{\mathrm{on},t} &= \begin{cases} \sum_{f\in\mathcal{F}} p_f \cos(\pi\nu_f) e^{\mathrm{i}\pi\nu_f} e^{-\mathrm{i}2\pi ft} & t \ne 0 \\ 1 & t = 0 \end{cases} \\
x_{\mathrm{on},k} &= \sum_{f\in\mathcal{F}} p_f \cos[\pi_N(k - Nf)] D_N(k - Nf), \\
x_{\mathrm{off},k} &= \sum_{f\in\mathcal{F}} p_f \sin[\pi_N(k - N\nu_f)] D_N(k - Nf).
\end{aligned} \tag{79}$$

Table 2: Constants used in the proof.

Constant	Meaning
C_0	$(\sigma_{\text{off}} + \sigma_{\text{H}})/\sigma$
C_1	Eq. (51)
C_2	Eq. (51)
C_3	$C_1 + C_2\pi + 2\pi\sqrt{S}/(\sqrt{3}\log N)$
π_N	$\pi(1 - N^{-1})$

Let $\mathcal{D}$ denote the integer part of frequencies in $\mathcal{F}_{\text{dom}}$. For each $n \in \mathcal{D}$, define f_n as the frequency that is closest to it, and

$$q_n \equiv p_{f_n} \sin(\pi\nu_{f_n}) e^{\mathrm{i}\pi\nu_{f_n}}. \tag{80}$$

Hence,

$$\sum_{n\in\mathcal{D}} q_n e^{-\mathrm{i}2\pi ft} = \mathrm{i}y_{\text{off},t} - \sum_{f\in\mathcal{F}_{\text{res}}} p_f \sin(\pi\nu_f) e^{\mathrm{i}\pi\nu_f} e^{-\mathrm{i}2\pi ft} = \mathcal{O}(\sigma_{\text{off}}). \tag{81}$$

Consider an approximate Fourier matrix $\tilde{F}$, such that

$$\tilde{F}_{nk} = \begin{cases} e^{-\mathrm{i}2\pi nk/N} & n \neq \mathcal{D}, \\ e^{-\mathrm{i}2\pi f_n t} & n \in \mathcal{D}. \end{cases} \tag{82}$$

Note that $|Nf_n - n'| \geq 1/2 \ \forall n' \notin \mathcal{D}$. Hence,

$$|p_f \sin(\pi\nu_f)| = |q_n| = \mathcal{O}(\sigma_{\text{off}}) \quad \forall n \in \mathcal{D}. \tag{83}$$

B.3 Proof of Lemma 2

Suppose the signal has optimal grid decomposition (without loss of generality, we assume $\mathfrak{u} = 0$)

$$y = F^{-1}(x_{\text{on}} + \mathrm{i}x_{\text{off}}), \quad x_{\text{on}}, x_{\text{off}} \in \mathbb{R}^N. \tag{84}$$

After the computation, we obtain the following.

$$x_{\text{on},n} = \sum_{f\in\mathcal{F}} p_f \cos[\pi_N(n - Nf)] D_N(n - Nf). \tag{85}$$

The optimal on-grid condition $\|y_{\text{off}}\|_\infty = \sigma_{\text{off}}$ requires that $\|x_{\text{off}}\|_2 \leq \sigma_{\text{off}}$. For each $n \in \mathcal{D}$, define f_n as the frequency that is closest to it, hence

$$\begin{aligned} x_{\text{on},n} &= \sum_{f\in\mathcal{F}} p_f \cos[\pi_N(n - Nf)] D_N(n - Nf) \\ &= p_{f_n} \cos[\pi_N \nu_{f_n}] D_N(\nu_{f_n}) + \sum_{f\in\mathcal{F}_{\text{res}}} p_f \cos[\pi_N(n - Nf)] D_N(n - Nf) \\ &\quad + \sum_{m\neq n\in\mathcal{D}} p_{f_m} \cos[\pi_N(m - n - \nu_{f_m})] D_N(m - n - \nu_{f_m}) \end{aligned} \tag{86}$$

Note that when $N \geq 100, |x| \geq 1/2$, we have $\pi\cos(\pi_N x)/2x \in [-1.1, 0.55]$. Therefore,

$$p_{f_n}\cos[\pi_N\nu_{f_n}]D_N(\nu_{f_n}) \geq p_{f_n}\left(1-\frac{1}{2}\pi_N^2\nu_{f_n}^2\right)\left(1-\frac{\pi}{3}\nu_{f_n}^2\right), \tag{87}$$

$$\begin{aligned} p_{f_m}\cos[\pi_N(n-Nf)]D_N(n-Nf) &\geq -\pi p_{f_m}|\nu_{f_m}|\left|\frac{\cos[\pi_N(n-Nf)}{N\sin[\pi(n-Nf)/N]}\right| \\ &\geq -1.1|p_{f_m}\nu_{f_m}|, \end{aligned} \tag{88}$$

$$\sum_{f\in\mathcal{F}_{\rm res}} p_f\cos[\pi_N(n-Nf)]D_N(n-Nf) \geq -\sum_{f\in\mathcal{F}_{\rm res}} p_f = -\mathcal{O}(\sigma_{\rm off}). \tag{89}$$

Recall that $\forall f \in \mathcal{F}_{\rm dom}, |p_f\nu_f| = \mathcal{O}(\sigma_{\rm off})$ (Lemma 1). Hence,

$$x_{{\rm on},n} \geq p_{f_n} - \mathcal{O}(S\sigma_{\rm off}). \tag{90}$$

Using the same argument, we can show that for all $n \neq \mathcal{D}$, $x_{{\rm on},n} \leq \mathcal{O}(S\sigma_{\rm off})$.

B.4 Proof of Lemma 4

To prove the lemma, we first need a sufficient condition for $s = x_\nu^{\rm R}$ to be a feasible solution of Eq. (53), i.e.,

$$\|F_{\nu,\mathcal{T}}x_\nu^{\rm R} - y_\mathcal{T}\|_2 \leq \sqrt{|\mathcal{T}|}\sigma, \tag{91}$$

so that we can bound $\|s_\nu - x_\nu^{\rm R}\|_2$. Note that

$$F_{\nu,\mathcal{T}}x_\nu^{\rm R} - y_\mathcal{T} = -{\rm i}F_{\nu,\mathcal{T}}x_\nu^{\rm I} - y_{{\rm off},\mathcal{T}}^0 - z_\mathcal{T}, \tag{92}$$

where $\|y_{{\rm off},\mathcal{T}}^0\|_2 \leq \sqrt{|\mathcal{T}|}\sigma_{\rm off}, \|z_\mathcal{T}\|_2 \leq \sqrt{|\mathcal{T}|}\sigma_{\rm H}$, and $\|F_{\nu,\mathcal{T}}x_\nu^{\rm I}\|_2$ has the following upper bound.

Lemma 6 (Concentration of $\|F_{\nu,\mathcal{T}}x_\nu^{\rm I}\|_2$)**.** *Suppose $\mathcal{T}$ is an integer set in $[N]$ generated by sampling ratio $r = \mathcal{O}(N^{-1}S^2\,{\rm poly}\log N)$. Then*

$$\|F_{\nu,\mathcal{T}}x_\nu^{\rm I}\|_2 \leq 4\pi\sqrt{\frac{|\mathcal{T}|}{3}}|\nu| \tag{93}$$

with probability at least $1 - 1/{\rm poly}(N)$.

Proof. Let $V_t \equiv |(F_\nu x_\nu^{\rm I})_t|^2$. Given a sampling ratio r, we introduce the following random variables:

$$\Pr[\hat{X}_t = V_t] = r, \quad \Pr[\hat{X}_t = 0] = 1-r; \quad \hat{R} = \sum_{t=0}^{N-1}\hat{X}_t. \tag{94}$$

Note that $\hat{R} = \|F_{\nu,\mathcal{T}}x_\nu^{\rm I}\|_2^2$, whose expectation value is

$$\mathbb{E}[\hat{R}] = r\|F_\nu x_\nu^{\rm I}\|_2^2 = Nr\|x_\nu^{\rm I}\|_2^2. \tag{95}$$

Bernstein's inequality states that

$$\Pr\left(\hat{R} \geq 2\mathbb{E}[\hat{R}]\right) \leq \exp\left(-\frac{\frac{1}{2}\mathbb{E}[\hat{R}]^2}{\sum_{t=0}^{N-1}E[\hat{X}_t^2] + \frac{1}{3}\max_t V_t\cdot\mathbb{E}[\hat{R}]}\right). \tag{96}$$

We need an upper bound for $\sum_{t=0}^{N-1}E[\hat{X}_t^2]$. Note that

$$\frac{1}{r}\sum_{t=0}^{N-1}\mathbb{E}[\hat{X}_t^2] = \sum_{t=0}^{N-1}V_t^2 = \sum_{t=0}^{N-1}\left|(F_\nu x_\nu^{\rm I})_t\right|^4. \tag{97}$$

After computation, we obtain

$$
\begin{aligned}
(F_\nu x_\nu^{\mathrm{I}})_t &= \sum_{k=0}^{N-1} e^{-\mathrm{i}2\pi(k+\nu)t/N} x_{\nu,k}^{\mathrm{I}} = e^{-\mathrm{i}2\pi\mathfrak{u}t/N} e^{-\mathrm{i}\pi(\nu-\mathfrak{u})} \sin[\pi(\nu-\mathfrak{u})] y_{\mathrm{on},t} \\
V_t &= \left|(F_\nu x_\nu^{\mathrm{I}})_t\right|^2 = \sin^2[\pi(\nu-\mathfrak{u})] \cdot |y_{\mathrm{on},t}|^2 \le \pi^2 |\nu-\mathfrak{u}|^2, \\
\sum_{t=0}^{N-1} V_t^2 &= \sum_{t=0}^{N-1} \left|(F_\nu x_\nu^{\mathrm{I}})_t\right|^4 \le \pi^4 |\nu-\mathfrak{u}|^4 \sum_{t=0}^{N-1} |y_{\mathrm{on},\mathrm{t}}|^4 \le \pi^4 N |\nu-\mathfrak{u}|^4.
\end{aligned} \tag{98}
$$

In conjunction with Eq. (67), we have

$$
\Pr\left(\sum_{t=0}^{N-1} \hat{X}_t \ge 2\mathbb{E}[\hat{R}]\right) \le \exp\left[-\frac{-\frac{1}{2}NrC_4^4}{\pi^4 + \frac{1}{3}\pi^2 C_4^2}\right]. \tag{99}
$$

Since $Nr = \mathcal{O}(S^2 \operatorname{poly}\log N)$, we can conclude that

$$
\begin{aligned}
\|F_{\nu,\mathcal{T}} x_\nu^{\mathrm{I}}\|_2 &\le \sqrt{\frac{2\mathbb{E}[\mathcal{T}]}{N}} \|F_\nu x_\nu^{\mathrm{I}}\|_2 \le 2\sqrt{\frac{|\mathcal{T}|}{N}} \|F_\nu x_\nu^{\mathrm{I}}\|_2 \\
&\le 2\sqrt{|\mathcal{T}|} \|x_\nu^{\mathrm{I}}\|_2 \le 4\pi \sqrt{\frac{|\mathcal{T}|}{3}} |\nu|
\end{aligned} \tag{100}
$$

with probability at least $1 - 1/\mathrm{poly}(N)$. □

Proof of Lemma 4. The upper bound of C_0 in Eq. (72) and Lemma 6 ensure that x_ν^{R} is feasible with high probability, which implies $\|s_\nu\|_1 \le \|x_\nu^{\mathrm{R}}\|_1$. Meanwhile,

$$
\|F_{\nu,\mathcal{T}}(s_\nu - x_\nu^{\mathrm{R}})\|_2 \le \|F_{\nu,\mathcal{T}} s_\nu - y_{\mathcal{T}}\|_2 + \|F_{\nu,\mathcal{T}} x_\nu^{\mathrm{R}} - y_{\mathcal{T}}\|_2 \le 2\sqrt{|\mathcal{T}|}\sigma. \tag{101}
$$

In conjunction with Theorems 3, 4, Lemma 3 and the condition that $|\nu| \le \sqrt{S}\sigma/\log N$, we obtain

$$
\begin{aligned}
\|s_\nu - x_\nu^{\mathrm{R}}\|_2 &\le 2C_1\sigma + C_2 \|x_{\nu,\mathrm{res}}^{\mathrm{R}}\|_1/\sqrt{S} \quad \text{(Theorems 3, 4)} \\
&\le 2C_1\sigma + C_2\pi|\nu|\log N/\sqrt{S} \quad \text{(Lemma 3)} \\
&\le (2C_1 + C_2\pi)\sigma, \qquad (102) \\
\|x_\nu^{\mathrm{R}} - x_{\mathrm{on}}\|_2 &\le \frac{2\pi}{\sqrt{3}}|\nu| \qquad \text{(Lemma 3)} \\
&\le \frac{2\pi}{\sqrt{3}} \frac{\sqrt{S}}{\log N}\sigma. \qquad (103)
\end{aligned}
$$

The lemma is proved by triangle inequality from here.

□

B.5 Proof of Lemma 5

The Chernoff-Hoeffding's inequality states that

Lemma 7. *Given a complex vector $b \in \mathbb{C}^N$ with $\|b\|_\infty \le 1$. Randomly sample integers in $[N]$ for L times and denote the result as $\mathcal{I}$. Then*

$$
\Pr\left(\sum_{i\in\mathcal{I}} |b_i|^2 \le \frac{L}{2N}\|b\|_2^2\right) < \exp\left(-\frac{L}{2}\right), \quad \Pr\left(\sum_{i\in\mathcal{I}} |b_i|^2 \ge \frac{3L}{2N}\|b\|_2^2\right) < \exp\left(-\frac{L}{2}\right). \tag{104}
$$

Proof. Introduce i.i.d. random variables $\hat{B}_1, \hat{B}_2, \ldots, \hat{B}_L$ satisfying fora all $\ell \in [1, L]$,

$$\Pr\left(\hat{B}_l = |b_i|^2\right) = \frac{1}{N}, \quad i \in [N]. \tag{105}$$

The distribution of $\sum_{i\in\mathcal{I}} |b_i|^2$ is identical to that of $\sum_{l=1}^{L} \hat{B}_l$. The Chernoff-Hoeffding's inequality states that

$$\Pr\left(\sum_{l=1}^{L} \hat{B}_l \leq \frac{L}{2}\mathbb{E}[\hat{B}]\right) < \exp\left(-\frac{L}{2\max_i |b_i|^2}\right) \leq \exp\left(-\frac{L}{2}\right), \tag{106}$$

where

$$\mathbb{E}[\hat{B}] = \frac{1}{N}\sum_{i=1}^{N} |b_i|^2 = \frac{1}{N}\|b\|_2^2. \tag{107}$$

Hence, the first concentration inequality in Eq. (104) is confirmed. The other one can be proved in the same manner.

□

Now we can prove Lemma 5.

Proof. Recall that $y = y_{\text{on}}^0 + y_{\text{off}}^0 + z$, $y_{\text{on}}^0 = F_\nu(x_\nu^{\text{R}} + \mathrm{i}x_\nu^{\text{I}})$. Therefore,

$$\begin{aligned}
\|y - F_\nu s_\nu\|_2 &= \|F_\nu(x_\nu^{\text{R}} - s_\nu + \mathrm{i}x_\nu^{\text{I}}) + y_{\text{off}}^0 + z\|_2 \qquad (108)\\
&\leq \|F_\nu(x_\nu^{\text{R}} - s_\nu)\|_2 + \|F_\nu x_\nu^{\text{I}}\|_2 + \|y_{\text{off}}\|_2 + \|z\|_2\\
&= \sqrt{N}\sqrt{\|x_\nu^{\text{R}} - s_\nu\|_2^2 + \|x_\nu^{\text{I}}\|_2^2} + \|y_{\text{off}}\|_2 + \|z\|_2\\
&\leq \sqrt{N}\left(\|x_{\text{on}} - s_\nu\|_2^2 + \|x_\nu^{\text{R}} - x_{\text{on}}\|_2^2 + \|x_\nu^{\text{I}}\|_2^2\right)^{1/2} + \sqrt{N}\sigma_{\text{off}} + \sqrt{N}\sigma_{\text{H}}\\
&\leq \sqrt{N}\left(\|x_{\text{on}} - s_\nu\|_2^2 + \frac{8\pi^2}{3}\nu^2\right)^{1/2} + \sqrt{N}\sigma_{\text{off}} + \sqrt{N}\sigma_{\text{H}} \quad \text{(Lemma 3)}\\
&\leq \sqrt{N}\left(C_3^2\sigma^2 + 8\pi^2\nu^2/3\right)^{1/2} + \sqrt{N}\left(\sigma_{\text{off}} + \sigma_{\text{H}}\right) \quad \text{(Lemma 4)}. \qquad (109)
\end{aligned}$$

According to Lemma 7, with probability at least $1 - \exp(-L/2)$, we have

$$\sum_{t\in\mathcal{I}} |y_t - (F_\nu s_\nu)_t|^2 \leq \frac{3L}{2}\left[\left(C_3^2\sigma^2 + 8\pi^2\nu^2/3\right)^{1/2} + \sigma_{\text{off}} + \sigma_{\text{H}}\right]^2. \tag{110}$$

Equation (76) is confirmed when we take the upper bound $|\nu| = \sqrt{S}\sigma/\log N$.

By virtue of Lemma 7, with probability at least $1 - \exp(-L/2)$, we have

$$\sum_{t\in\mathcal{I}} |y_t - (F_\nu s_\nu)_t|^2 \geq \frac{L}{2N}\|y - F_\nu s_\nu\|_2^2. \tag{111}$$

The 2-norm has the following lower bound.

$$\begin{aligned}
\|y - F_\nu s_\nu\|_2 &\geq \|F_\nu(x_\nu^{\text{R}} + \mathrm{i}x_\nu^{\text{I}} - s_\nu) + y_{\text{off}} + z\|_2\\
&\geq \|F_\nu(x_\nu^{\text{R}} + \mathrm{i}x_\nu^{\text{I}} - s_\nu)\|_2 - \|y_{\text{off}}\|_2 - \|z\|_2\\
&\geq \sqrt{N}\|x_\nu^{\text{I}}\|_2 - \sqrt{N}\sigma_{\text{off}} - \sqrt{N}\sigma_{\text{H}}\\
&\geq \sqrt{N}(C[x_{\text{on}}]|\nu| - C_0\sigma) \qquad \text{(Lemma 3)} \qquad (112)
\end{aligned}$$

Hence,

$$\sum_{t\in\mathcal{I}} |y_t - (F_\nu s_\nu)_t|^2 \geq \frac{L}{2}(C[x_{\text{on}}]|\nu| - C_0\sigma)^2. \tag{113}$$

This completes the proof of Eq. (78). □

C Proof of technical lemmas

C.1 Properties of the Dirichlet kernel

The Dirichlet kernel is defined as

$$D_N(\nu) \equiv \frac{1}{N}\sum_{m=0}^{N-1} e_\nu^{-N+1+2m}, \quad e_\nu \equiv e^{\mathrm{i}\pi\nu/N}. \tag{114}$$

In a more concise form, it equals

$$D_N(\nu) = \begin{cases} 1, & \nu = 0 \\ \frac{\sin(\pi\nu)}{N\sin(\pi\nu/N)}, & \nu \neq 0. \end{cases} \tag{115}$$

We start with a few estimations for the Dirichlet kernel.

Lemma 8. *The Dirichlet kernel satisfies*

$$1 - D_N(\nu)^2 \leq \frac{\pi\nu^2}{3}, \tag{116}$$

$$|D_N(n+\nu)| \leq \frac{\pi|\nu|}{2|n+\nu|}, \tag{117}$$

$$\sum_{n=0}^{N-1} D_N(n+\nu)D_N(n+\nu+l) = \delta_{l,0}. \tag{118}$$

Proof. If $1 - D_N(\nu) \leq c\nu^2$, then

$$1 - D_N^2(\nu) \leq 1 - (1-c\nu^2)^2 \leq 2c\nu^2. \tag{119}$$

Therefore, we work on $1 - D_N(\nu)$ instead. Suppose N is odd. Then

$$\begin{aligned} 1 - D_N(\nu) &= \frac{1}{N}\sum_{m=0}^{N-1}(1 - e_\nu^{-N+1+2m}), \\ &= \frac{1}{N}\sum_{m'=1}^{\frac{N-1}{2}}(2 - e_\nu^{2m'} - e_\nu^{-2m'}) = \frac{1}{N}\sum_{m'=1}^{\frac{N-1}{2}} 4\sin^2\left(\frac{m'\pi\nu}{N}\right) \\ &\leq \frac{4}{N}\sum_{m'=1}^{\frac{N-1}{2}} \frac{\pi^2(m')^2\nu^2}{N^2} = \frac{4\pi^2}{N^3}\frac{1}{6}\frac{N-1}{2}\frac{N+1}{2}N\nu^2 \leq \frac{\pi^2}{6}\nu^2. \end{aligned} \tag{120}$$

Hence, $1 - D_N(\nu)^2 \leq \frac{\pi^2}{3}\nu^2$. The even N situation is similar. Equation (116) is confirmed.

Equation (117) is confirmed by

$$|D(n+\nu)| = \left|\frac{\sin[\pi(n+\nu)]}{N\sin[\pi(n+\nu)/N]}\right| \leq \frac{\pi|\nu|}{2|n+\nu|}. \tag{121}$$

If $l = 0$, then

$$D_N(n+\nu) = \frac{1}{N}\sum_{m=0}^{N-1} e_{n+\nu}^{-N+1+2m}, \quad e_{n+\nu} \equiv e^{\mathrm{i}\pi(n+\nu)/N}, \tag{122}$$

$$\begin{aligned} D_N(n+\nu)^2 &= \frac{1}{N^2}\sum_{m,m'=0}^{N-1} e_{n+\nu}^{-2N+2+2m+2m'} \\ &= \frac{1}{N} + \frac{1}{N^2}\sum_{m+m'\neq N-1} e_{n+\nu}^{-2N+2+2m+2m'}. \end{aligned} \tag{123}$$

Note that for any $k \neq 0$,

$$\sum_{n=0}^{N-1} e_{n+\nu}^{k} = \sum_{n=0}^{N-1} e^{\mathrm{i}\pi(n+\nu)k/N} = e^{\mathrm{i}k\nu\pi/N} \sum_{n=0}^{N-1} e^{\mathrm{i}\pi nk/N} = 0. \tag{124}$$

Hence,

$$\sum_{n=0}^{N-1} D_N(n+\nu)^2 = 1. \tag{125}$$

If $l \neq 0$, then

$$\begin{aligned}\sum_{n=0}^{N-1} D_N(n+\nu)D_N(n+\nu+l) &= \frac{1}{N^2} \sum_{m_1,m_2=0}^{N-1} N\delta_{m_1+m_2+1-N} \exp\left[\mathrm{i}\frac{\pi l}{N}(2m_2 - N + 1)\right] \\ &= \frac{1}{N} \sum_{m=0}^{N-1} \exp\left[\mathrm{i}\frac{\pi l}{N}(2m - N + 1)\right] = 0. \end{aligned} \tag{126}$$

Equation (118) is confirmed.

□

C.2 Proof of Lemma 3

Define vectors $\boldsymbol{c}_\nu, \boldsymbol{s}_\nu \in \mathbb{R}^N$ by

$$\boldsymbol{c}_{\nu,k} \equiv \cos[\pi_N(k+\nu)]D_N(k+\nu), \quad \boldsymbol{s}_{\nu,k} \equiv \sin[\pi_N(k+\nu)]D_N(k+\nu). \tag{127}$$

Lemma 9. *Suppose $N \geq 100$. Then*

$$\|\boldsymbol{s}_\nu\|_1 \leq |\boldsymbol{s}_{\nu,0}| + \pi^2|\nu|\log N, \tag{128}$$

$$\|\boldsymbol{c}_\nu\|_1 \leq |\boldsymbol{c}_{\nu,0}| + \pi^2|\nu|\log N, \tag{129}$$

$$2|\nu| \leq \|\boldsymbol{s}_\nu\|_2 \leq \frac{2\pi}{\sqrt{3}}|\nu|, \tag{130}$$

$$\|\boldsymbol{c}_\nu - \delta_0\|_2 \leq \frac{2\pi}{\sqrt{3}}|\nu|. \tag{131}$$

Proof. By virtue of Lemma 8, we have

$$|\boldsymbol{s}_{\nu,0}| \leq \pi|\nu|; \quad |\boldsymbol{s}_{\nu,k}| \leq \frac{\pi|\nu|}{2|k+\nu|}, k \neq 0. \tag{132}$$

Suppose N is odd. Then

$$\begin{aligned}\sum_{k=0}^{N-1} |\boldsymbol{s}_{\nu,k}|^2 &\leq |\sin(\pi_N\nu)|^2 D_N(\nu)^2 + \sum_{k\neq 0} D_N(k+\nu)^2 \\ &\leq 1 - D_N(\nu)^2 + |\sin(\pi_N\nu)|^2 \leq \frac{4}{3}\pi^2\nu^2, \end{aligned} \tag{133}$$

$$\begin{aligned}\sum_{k\neq 0} |\boldsymbol{s}_{\nu,k}| &\leq \sum_{k\neq 0} |D_N(k+\nu)| \leq \frac{\pi|\nu|}{2} \sum_{k=1}^{\frac{N-1}{2}} \left(\frac{1}{k+\nu} + \frac{1}{k-\nu}\right) \\ &\leq \pi|\nu| \sum_{k=1}^{\frac{N-1}{2}} \left(\frac{1}{2k-1} + \frac{1}{2k}\right) \leq \pi|\nu| \left(1 + \ln\frac{N-1}{2}\right). \end{aligned} \tag{134}$$

One can verify that when $N \geq 100$,

$$1 + \ln \frac{N-1}{2} \leq \pi \log N. \tag{135}$$

The proof of even N situation is similar. This completes the proof of Eq. (128) and the upper bound of Eq. (130). The proof for Eq. (129) is similar.

Note that

$$\begin{aligned} \sin^2[\pi(k+\nu)] - \sin^2[\pi_N(k+\nu)] &= \frac{1}{2}\left[\cos(2\pi\nu) - \cos(2\pi_N(k+\nu))\right] \\ &= \cos(2\pi\nu)\sin^2\left[\frac{\pi(k+\nu)}{N}\right] + \sin(2\pi\nu)\sin\left[\frac{\pi(k+\nu)}{N}\right]\cos\left[\frac{\pi(k+\nu)}{N}\right]. \end{aligned} \tag{136}$$

Therefore,

$$\begin{aligned} \|\boldsymbol{s}_\nu\|_2^2 = \sin^2(\pi\nu) &+ \cos(2\pi\nu)\sum_{k=0}^{N-1}\sin^2\left[\frac{\pi(k+\nu)}{N}\right]D_N(k+\nu)^2 \\ &- \sin(2\pi\nu)\sum_{k=0}^{N-1}\sin\left[\frac{\pi(k+\nu)}{N}\right]\cos\left[\frac{\pi(k+\nu)}{N}\right]D_N(k+\nu)^2. \end{aligned} \tag{137}$$

Let $e_{k+\nu} \equiv e^{\mathrm{i}\pi(k+\nu)/N}$. After computation, we obtain the following.

$$\begin{aligned} \sum_{k=0}^{N-1}\sin^2\left[\frac{\pi(k+\nu)}{N}\right]D_N(k+\nu)^2 &= \sum_{k=0}^{N-1}\frac{2 - e_{k+\nu}^2 - e_{k+\nu}^{-2}}{4N^2}\sum_{m,m'=0}^{N-1} e_{k+\nu}^{-2N+2+2m+2m'} \\ &= \frac{1}{2} - \frac{1}{4N}(N - 1 + e^{\mathrm{i}2\pi\nu}) - \frac{1}{4N}(N - 1 + e^{-\mathrm{i}2\pi\nu}) \\ &= \frac{1}{N}\sin^2(\pi\nu), \end{aligned} \tag{138}$$

and

$$\begin{aligned} &\sum_{k=0}^{N-1}\sin\left[\frac{\pi(k+\nu)}{N}\right]\cos\left[\frac{\pi(k+\nu)}{N}\right]D_N(k+\nu)^2 \\ &= \sum_{k=0}^{N-1}\frac{e_{k+\nu}^2 - e_{k+\nu}^{-2}}{4N^2\mathrm{i}}\sum_{m,m'=0}^{N-1} e_{k+\nu}^{-2N+2+2m+2m'} \\ &= \frac{1}{4N\mathrm{i}}(N - 1 + e^{\mathrm{i}2\pi\nu}) - \frac{1}{4N\mathrm{i}}(N - 1 + e^{-\mathrm{i}2\pi\nu}) = \frac{1}{2N}\sin(2\pi\nu). \end{aligned} \tag{139}$$

Hence,

$$\begin{aligned} \|s_\nu\|_2^2 &= \sin^2(\pi\nu) + \frac{1}{N}\cos(2\pi\nu)\sin^2(\pi\nu) - \frac{1}{2N}\sin^2(2\pi\nu) \\ &= (1 - 2N^{-1})\sin^2(\pi\nu). \end{aligned} \tag{140}$$

In conjunction with the condition $N \geq 100$, we obtain the lower bound in Eq. (130).

Finally,

$$\begin{aligned} \|\boldsymbol{c}_\nu - \delta_0\|_2^2 &\leq |1 - \boldsymbol{c}_{\nu,0}|^2 + \sum_{k\neq 0}|\boldsymbol{c}_{\nu,k}|^2 \\ &\leq |1 - \cos(\pi_N\nu)\, D_N(\nu)|^2 + 1 - |D_N(\nu)|^2 \\ &\leq 2 - 2|D_N(\nu)| + \sin^2(\pi_N\nu)\, D_N(\nu)^2 \leq \frac{4}{3}\pi^2|\nu|^2. \end{aligned} \tag{141}$$

This completes the proof of Eq. (131).

□

The first part of Lemma 3 is proved in the following lemma.

Lemma 10. *Suppose $y = Fx$ is a length-N ($N \geq 100$) on-grid signal and x satisfies the condition in Eq. (65). Let $y = F_\nu(x_\nu^{\mathrm{R}} + \mathrm{i}x_\nu^{\mathrm{I}})$. Then*

$$\frac{\|x_\nu^{\mathrm{I}}\|_2}{\|x\|_1} \leq \frac{2\pi}{\sqrt{3}}|\nu|, \quad \frac{\|x_\nu^{\mathrm{R}} - x\|_2}{\|x\|_1} \leq \frac{2\pi}{\sqrt{3}}|\nu|, \tag{142}$$

$$\sum_{n\in\mathcal{D}_{\mathrm{res}}} |x_{\nu,n}^{\mathrm{R}}| \leq \|x\|_1 \pi^2 |\nu| \log N. \tag{143}$$

Proof. By definition, we have

$$x_\nu^{\mathrm{R}} = \sum_{n\in\mathcal{D}'} q_n \boldsymbol{c}_\nu^{(n)}, \quad \boldsymbol{c}_{\nu,k}^{(n)} \equiv \cos\left[\pi_N(k-n+\nu)\right] D_N(k-n+\nu), \tag{144}$$

$$x_\nu^{\mathrm{I}} = \sum_{n\in\mathcal{D}'} q_n \boldsymbol{s}_\nu^{(n)}, \quad \boldsymbol{s}_{\nu,k}^{(n)} \equiv \sin\left[\pi_N(k-n+\nu)\right] D_N(k-n+\nu). \tag{145}$$

Note that $\boldsymbol{s}_\nu^{(n)}, \boldsymbol{c}_\nu^{(n)}$ are simply $\boldsymbol{s}_\nu, \boldsymbol{c}_\nu$ with a permutation in entries. In Lemma 9, we have proved that

$$\|\boldsymbol{s}_\nu\|_2 \leq \frac{2\pi}{\sqrt{3}}|\nu|, \quad \|\boldsymbol{c}_\nu - \delta_n\|_2 \leq \frac{2\pi}{\sqrt{3}}|\nu|. \tag{146}$$

Hence,

$$\|x_\nu^{\mathrm{R}} - x\|_2 \leq \sum_{n\in\mathcal{D}'} q_n \|\boldsymbol{c}_\nu^{(n)} - \delta_n\|_2 \leq \|x\|_1 \frac{2\pi}{\sqrt{3}}|\nu|, \tag{147}$$

$$\|x_\nu^{\mathrm{I}}\|_2 \leq \sum_{n\in\mathcal{D}'} q_n \|\boldsymbol{s}_\nu^{(n)}\|_2 \leq \|x\|_1 \frac{2\pi}{\sqrt{3}}|\nu|. \tag{148}$$

Similarly, in conjunction with the first two inequalities in Lemma 9, we obtain

$$\sum_{n\in\mathcal{D}_{\mathrm{res}}} |x_{\nu,n}^{\mathrm{R}}| \leq \sum_{m\in\mathcal{D}'} q_m \left(\sum_{n\in\mathcal{D}_{\mathrm{res}}} \left| \boldsymbol{c}_{\nu,n}^{(m)} \right| \right) \leq \|x\|_1 \pi^2 |\nu| \log N. \tag{149}$$

□

The most critical part is the lower bound of $\|x_\nu^{\mathrm{I}}\|_2$. We prove it separately in the following lemma.

Lemma 11. *Suppose $y = Fx$ is a length-N ($N \geq 100$) on-grid signal and x satisfies the condition in Eq. (65). Let $y = F_\nu(x_\nu^{\mathrm{R}} + \mathrm{i}x_\nu^{\mathrm{I}})$. Then $\|x_\nu^{\mathrm{I}}\|_2 \geq C[x]|\nu|$ with function C defined in Eq. (66).*

Proof. Denote the on-grid component as $x = \sum_{n\in\mathcal{D}'} q_n \delta_n$, then

$$\|x_\nu^{\mathrm{I}}\|_2^2 = \sum_{n\in\mathcal{D}'} q_n^2 \mathcal{M}_{n,n} + \sum_{n\neq m\in\mathcal{D}'} q_n q_m \mathcal{M}_{n,m}, \tag{150}$$

where

$$\mathcal{M}_{n,m} \equiv \sum_{k=0}^{N-1} \sin\left[\pi_N(k-n+\nu)\right] \sin\left[\pi_N(k-m+\nu)\right] D_N(k-n+\nu) D_N(k-m+\nu). \tag{151}$$

By virtue of Lemma 9, for all n, we have

$$\mathcal{M}_{n,n} = \sum_{k=0}^{N-1} \sin^2\left[\pi_N(k-n+\nu)\right] D_N^2(k-n+\nu) = \|\boldsymbol{s}_\nu\|_2^2 \geq 4\nu^2. \tag{152}$$

By virtue of Lemma 8, for all $n \neq m$, we have

$$\begin{aligned}
\mathcal{M}_{n,m} &= \sum_{k=0}^{N-1} \cos[\pi_N(n-m)] D_N(k-n+\nu) D_N(k-m+\nu) \\
&\quad - \sum_{k=0}^{N-1} \cos\left[\pi_N(2k-n-m+2\nu)\right] D_N(k-n+\nu) D_N(k-m+\nu) \\
&= - \sum_{k=0}^{N-1} \cos\left[\pi_N(2k-n-m+2\nu)\right] D_N(k-n+\nu) D_N(k-m+\nu).
\end{aligned} \tag{153}$$

Let $e_{k-n+\nu} \equiv e^{\mathrm{i}\pi(k-n+\nu)/N}, e_{k-m+\nu} \equiv e^{\mathrm{i}\pi(k-m+\nu)/N}$. Then

$$\begin{aligned}
-\mathcal{M}_{n,m} &= \sum_{k=0}^{N-1} \frac{1}{2}(e_{k-n+\nu}^{N-1} e_{k-m+\nu}^{N-1} + e_{k-n+\nu}^{1-N} e_{k-m+\nu}^{1-N}) \cdot \frac{1}{N^2} \sum_{j=0}^{N-1} \sum_{j'=0}^{N-1} e_{k-n+\nu}^{-N+1+2j} e_{k-m+\nu}^{-N+1+2j'} \\
&= \frac{1}{2N^2} \sum_{j=0}^{N-1} \sum_{j'=0}^{N-1} \sum_{k=0}^{N-1} e_{k-n+\nu}^{2j} e_{k-m+\nu}^{2j'} + \frac{1}{2N^2} \sum_{j=0}^{N-1} \sum_{j'=0}^{N-1} \sum_{k=0}^{N-1} e_{k-n+\nu}^{-2N+2+2j} e_{k-m+\nu}^{-2N+2+2j'} \\
&= \frac{1}{2N^2} \sum_{j=0}^{N-1} \sum_{j'=0}^{N-1} \sum_{k=0}^{N-1} e_{k-n+\nu}^{2j} e_{k-m+\nu}^{2j'} + \frac{1}{2N^2} \sum_{j=0}^{N-1} \sum_{j'=0}^{N-1} \sum_{k=0}^{N-1} e_{n-k-\nu}^{2j} e_{m-k-\nu}^{2j'} \\
&:= M_1 + M_2.
\end{aligned} \tag{154}$$

Note that

$$\frac{1}{N} \sum_{k=0}^{N-1} e_{k-n+\nu}^{2j} e_{k-m+\nu}^{2j'} = \begin{cases} \delta_{j'=0} & j=0 \\ \delta_{j'=N-j} e^{\mathrm{i}2\pi j(m-n)/N} e^{-\mathrm{i}2\pi(m-\nu)} & j \neq 0 \end{cases}; \tag{155}$$

$$\frac{1}{N} \sum_{k=0}^{N-1} e_{n-k-\nu}^{2j} e_{m-k-\nu}^{2j'} = \begin{cases} \delta_{j'=0} & j=0 \\ \delta_{j'=N-j} e^{\mathrm{i}2\pi j(n-m)/N} e^{\mathrm{i}2\pi(m-\nu)} & j \neq 0 \end{cases}. \tag{156}$$

Therefore,

$$\begin{aligned}
M_1 &= \frac{1}{2N} \left[1 + e^{-\mathrm{i}2\pi(m-\nu)} \sum_{j=1}^{N-1} e^{\mathrm{i}2\pi j(m-n)/N} \right] \\
&= \frac{1}{2N} \left[1 - e^{-\mathrm{i}2\pi(m-\nu)} \right] = \frac{1}{2N}(1 - e^{\mathrm{i}2\pi\nu}).
\end{aligned} \tag{157}$$

$$\begin{aligned}
M_2 &= \frac{1}{2N} \left[1 + e^{\mathrm{i}2\pi(m-\nu)} \sum_{j=1}^{N-1} e^{\mathrm{i}2\pi j(n-m)/N} \right] \\
&= \frac{1}{2N} \left[1 - e^{\mathrm{i}2\pi(m-\nu)} \right] = \frac{1}{2N}(1 - e^{-\mathrm{i}2\pi\nu}).
\end{aligned} \tag{158}$$

Hence, $\mathcal{M}_{m,n} = (e^{\mathrm{i}2\pi\nu} + e^{-\mathrm{i}2\pi\nu} - 2)/(2N) = -2\sin^2(\pi\nu)/N$, and

$$\begin{aligned}
\|x_\nu^{\mathrm{I}}\|_2^2 &\geq 4\nu^2 \|x\|_2^2 - 2(\|x\|_1^2 - \|x\|_2^2) \sin^2(\pi\nu)/N \\
&\geq \nu^2 \left[(4 + 2\pi^2/N)\|x\|_2^2 - 2\pi^2 \|x\|_1^2/N \right].
\end{aligned} \tag{159}$$

□