

Quantum Simulation with Sum-of-Squares Spectral Amplification

Robbie King^{1,2,*} Guang Hao Low¹ Ryan Babbush¹ Rolando D. Somma^{1,†} and Nicholas C. Rubin^{1,‡}

¹Google Quantum AI, Venice, California 90291, USA

²Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, California 91125, United States



(Received 5 May 2025; revised 18 September 2025; accepted 12 December 2025; published 17 March 2026)

We present sum-of-squares spectral amplification (SOSSA), a framework for improving quantum simulation relevant to low-energy problems. We show how SOSSA can be applied to problems like energy and phase estimation and provide fast quantum algorithms for these problems that significantly improve over prior art. To illustrate the power of SOSSA in applications, we consider the Sachdev-Ye-Kitaev model, a representative strongly correlated system, and demonstrate asymptotic speedups over generic simulation methods by a factor of the square root of the system size. Our results reinforce those observed in [G. H. Low *et al.*, [arXiv:2502.15882](#)], where SOSSA was used to achieve state-of-the-art gate costs for phase estimation of real-world quantum chemistry systems.

DOI: [10.1103/m3fj-m4rm](#)

Introduction—Simulating quantum many-body systems is one of the most heralded and valuable applications of quantum computing. Although efficient quantum algorithms exist for numerous quantum simulation problems [1–6], further improvements are essential to fully realize robust large-scale quantum simulation when accounting for the considerable constant factor overhead associated with fault-tolerant quantum computation [7–9]. Quantum algorithm improvements are expected to arise from exploiting specific structures of the simulation instance, thus mitigating worst-case computational costs. Examples of such are exploiting symmetries in chemistry problems or locality in lattice models [10,11]. Here, we introduce a strategy that exploits the low-energy properties of quantum states to improve quantum simulation algorithms. Our framework relies on two distinct algorithmic ideas: sum-of-squares (SOS) representations of Hamiltonians and spectral amplification (SA).

The low-energy setting is one of the most promising areas in quantum simulation. This setting is relevant in the study of matter at low temperatures, including quantum phase transitions and the computation of ground-state properties. In this setting, some of the most important simulation tasks are (i) estimating the energy of a quantum

system with respect to certain quantum state [1]; (ii) estimating the ground state energy of a quantum system by phase estimation [3]; and (iii) simulating the time evolution of a quantum state under a Hamiltonian [4]. We will show how SOSSA significantly improves the gate complexity of generic methods for all these problems. To this end, SOSSA combines two key ideas: first, it produces a suitable SOS representation of the given Hamiltonian H shifted by a constant β such that $H + \beta\mathbb{1}$ has a small ground-state energy, and second, it uses SA to essentially compute the square root of $H + \beta\mathbb{1}$ [12–14].

During the SOS step, $H + \beta\mathbb{1}$ is processed classically and represented as a sum of positive terms. This can modify properties of H that have an impact on the complexity of the simulation algorithm, such as an ℓ_1 norm that depends on its presentation, or the ground state energy. Ideally, the SOS representation is such that $-\beta$ is a tight lower bound on the ground state energy and that the square roots of the positive terms are not too difficult to simulate. This step relates to the well-known optimization task of finding the best SOS lower bound on the ground state energy, which can be solved efficiently in classical preprocessing using semidefinite programming. SOS is used in the context of approximation algorithms [15,16], lower bounds on ground-state energies in quantum chemistry [17,18], and characterizing quantum correlations [19]. These methods often appear in the pseudomoment picture, which is dual to the SOS optimization that we consider here.

During the SA step we produce a different Hamiltonian, the “square root” of $H + \beta\mathbb{1}$, whose eigenvalues are the square roots of those of $H + \beta\mathbb{1}$. The low-lying eigenvalues like the ground state energy are “amplified” and, for example, this can reduce the resources needed for phase estimation since the precision requirements are now less

*Contact author: robbieking@google.com

†Contact author: rsomma@google.com

‡Contact author: nickrubin@google.com

stringent. Similarly, access to the square root of $H + \beta\mathbb{I}$ allows us to reduce the problem of energy estimation to one of amplitude estimation with worse accuracy. While the SOS step might introduce additional overheads from the complexity of the terms, the SA step has the potential to reduce it, and the method is useful if the overall combination still provides an improvement in gate complexities. Notably, this occurs in interesting systems. SA was used to achieve a quadratic speedup in the context of adiabatic quantum computation [13] and was more recently used to obtain improved quantum algorithms for simulating time evolution and phase estimation on low-energy states [12,14,20].

This Letter introduces the SOSSA framework, generalizing prior work beyond quantum chemistry to any finite-dimensional Hamiltonian. We use this framework to construct improved quantum algorithms for expectation and ground-state energy estimation in the low-energy regime. Our methods improve upon prior art by removing the need for an upper bound on the energy to be estimated, applying to arbitrary block-encoded operators, and succeeding even when the initial state has an imperfect overlap with the ground state. We provide tight lower bounds showing these algorithms are query optimal and also present a low-depth version for near-term applications. Finally, we apply these algorithms to the Sachdev-Ye-Kitaev (SYK) model, demonstrating an asymptotic advantage in system size and highlighting the framework's broad applicability to strongly correlated quantum systems. Our results are summarized in Table I.

SOSSA—The goal of SOSSA is to improve the gate complexity of simulation tasks in the low-energy sector. To this end, SOSSA uses SA for reducing query complexities first. Quantum signal processing and the related quantum singular value transform provide the modern machinery for

quantum simulation tasks, relying on access to a block-encoding of a Hamiltonian H [22–24]. This is a unitary acting on an enlarged space that contains the matrix H/λ in one of its blocks, where $\lambda \geq \|H\|$ is needed for normalization. In quantum simulation algorithms we are often interested in both, the query and gate complexities. We define the query complexity to be the number of times the block encodings are used (including their inverses), and the gate complexity to be the total number of two-qubit gates to implement the algorithm, which includes the number of gates to implement the block encodings in applications.

A standard approach to construct the block encoding uses an efficient presentation of H as a linear combination of unitaries (LCU) [6,25]. Suppose a Hamiltonian H acts on a system of N qubits and is presented as a sum of terms as

$$H = \sum_{j=0}^{R-1} g_j \sigma_j, \quad (1)$$

where $g_j \in \mathbb{R}$ are coefficients and σ_j are (tensor) products of N Pauli matrices (i.e., Pauli strings). The query complexity of various quantum simulation tasks through this LCU scales linearly in the so-called ℓ_1 norm of the Hamiltonian:

$$\lambda_{\text{LCU}} := \sum_{j=0}^{R-1} |g_j|. \quad (2)$$

The LCU column of Table I presents some known query complexity results.

SA allows one to get around the linear cost in λ_{LCU} in simulation tasks involving states of low energy. SA can only be applied to Hamiltonians that are positive semi-definite. One way to achieve SA is to produce a square root

TABLE I. (a) Simulation tasks on Hamiltonian H to precision ϵ using LCU, termwise SA, and SOSSA. Presented are the query complexities assuming access to the Hamiltonian only via the block encodings H/λ_{LCU} , $H_{\text{SA}}/\sqrt{\lambda_{\text{LCU}}}$, and $H_{\text{SOSSA}}/\sqrt{\lambda_{\text{SOS}}}$, respectively. The gate complexities can be determined from the gate complexities of each block-encoding, which can be different, and the additional arbitrary gates in the algorithm. We show (i) estimation of the energy $E = \langle \psi | H | \psi \rangle$, (ii) phase estimation of the ground state energy E with initial state $|\psi\rangle$ and ground-state $|\psi_0\rangle$ satisfying $p = |\langle \psi | \psi_0 \rangle|^2 > 0$, and (iii) time evolution to implement e^{-iHt} on a state of energy at most E . We assume $-\lambda_{\text{LCU}} \leq E \leq -\lambda_{\text{LCU}} + \Delta_{\text{LCU}}$ for termwise SA and $-\beta \leq E \leq -\beta + \Delta_{\text{SOS}}$ for SOSSA. The lower bound on the ground state energy $-\beta$ is obtained in the SOS step and implies $\Delta_{\text{SOS}} \ll \Delta_{\text{LCU}}$. We also provide adaptive algorithms with improved complexities that do not require knowledge of Δ_{LCU} or Δ_{SOS} . (b) Normalization factors for the SYK model demonstrate an asymptotic speedup in system size N using an appropriate SOS. The gate complexities for all block encodings is similar in this example.

		LCU	Termwise SA	SOSSA
			$\lambda = \lambda_{\text{LCU}}$	$\lambda = \lambda_{\text{SOS}}$
		$\lambda = \lambda_{\text{LCU}}$	$\Delta = \Delta_{\text{LCU}}$	$\Delta = \Delta_{\text{SOS}}$
(a)	Energy estimation	$\Theta(\lambda/\epsilon)$ [1]	$\Theta(\sqrt{\Delta\lambda}/\epsilon)$ (See Theorem 5 in Supplemental Material [21])	
	Phase estimation	$\mathcal{O}((\lambda/(\sqrt{p}\epsilon)) \log(1/p))$ [7]	$\mathcal{O}((\sqrt{\Delta\lambda}/(\sqrt{p}\epsilon)) \log(1/p))$ (See Theorem 7 in Supplemental Material [21])	
	Time evolution	$\Theta[\lambda t + \log(1/\epsilon)]$ [2,22]	$\Theta(\sqrt{\Delta\lambda}t + \sqrt{\lambda/\Delta} \log(1/\epsilon))$ [12]	
(b)	SYK model	$\lambda_{\text{LCU}} \sim N^2$	$\sqrt{\Delta_{\text{LCU}}\lambda_{\text{LCU}}} \sim N^2$	$\sqrt{\Delta_{\text{SOS}}\lambda_{\text{SOS}}} \sim N^{\frac{3}{2}}$

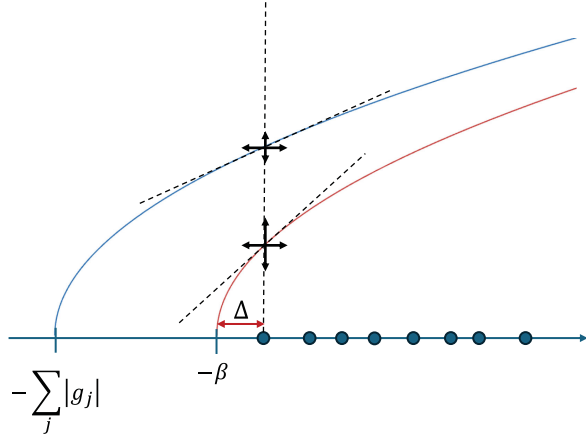


FIG. 1. In SOSSA, we shift the Hamiltonian H by a constant β and express $H + \beta\mathbb{1}$ as a sum of squares $\sum_j B_j^\dagger B_j$. This allows us to find a rectangular operator H_{SOSSA} which acts as the square root of $H + \beta\mathbb{1}$. The dots in the figure represent the eigenvalues of H . By constructing the square root of a positive semidefinite Hamiltonian, we are able to amplify the small eigenvalues due to the divergent behavior of the derivative of $\sqrt{x}$ near zero. An estimate of an eigenvalue E to precision ϵ can be obtained through an estimate of $\sqrt{E}$ to precision $\mathcal{O}(\epsilon/\sqrt{E})$, which becomes coarser as E decreases, thereby reducing the resources needed for phase estimation.

of the Hamiltonian, so that the low-energy spectrum is amplified. Intuitively, this amplification arises because the square root function is steep near zero, i.e. $\sqrt{x} \gg x$ for $x \ll 1$, where x denotes a rescaled eigenvalue. See Fig. 1 for an illustration. For phase estimation, a precise estimate of a small eigenvalue of the original Hamiltonian can then be obtained by squaring a coarser estimate of an eigenvalue of its square root. For time evolution, the amplified low-energy spectrum allows for a more efficient approximation to the evolution operator using lower degree polynomials in quantum signal processing. In both cases, SA can lead to a lower query complexity than in the worst case. See Sec. II of Supplemental Material [21] for results on SA.

The easiest way to apply SA to a Hamiltonian presented as in Eq. (1) is to add a shift to each Pauli string σ_j so that it becomes positive semidefinite; for example, $1 \pm \sigma_j \geq 0$, where $\mathbb{1}$ is the identity matrix. We call this approach “termwise SA” but also anticipate that this approach is not generally effective. During this preprocessing, the shift produces

$$H + \lambda_{\text{LCU}}\mathbb{1} = 2 \sum_{j=0}^{R-1} |g_j| \Pi_j, \quad (3)$$

where $\Pi_j := \frac{1}{2}[\mathbb{1} + \text{sign}(g_j)\sigma_j]$ are simple orthogonal projectors since the Pauli strings σ_j have eigenvalues ± 1 . We can define the spectral amplified operator for this specific example to be

$$H_{\text{SA}} = \sqrt{2} \sum_{j=0}^{R-1} |g_j|^{\frac{1}{2}} |j\rangle \otimes \Pi_j = \sqrt{2} \begin{pmatrix} |g_1|^{\frac{1}{2}} \Pi_1 \\ \vdots \\ |g_R|^{\frac{1}{2}} \Pi_R \end{pmatrix}. \quad (4)$$

This is a rectangular matrix that acts as a square root of $H + \lambda_{\text{LCU}}\mathbb{1}$, since

$$H_{\text{SA}}^\dagger H_{\text{SA}} = H + \lambda_{\text{LCU}}\mathbb{1}. \quad (5)$$

Consider now a simulation problem where the corresponding quantum states are supported in the low-energy subspace. We can introduce a parameter Δ_{LCU} that quantifies the low-energy assumption. In particular, for energy estimation we will assume our states of interest $|\psi\rangle$ to satisfy $E = \langle \psi | H | \psi \rangle \leq -\lambda_{\text{LCU}} + \Delta_{\text{LCU}}$. For phase estimation we will assume $\langle \xi | \psi \rangle = 0$ for every eigenstate $|\xi\rangle$ of H of eigenvalue $\langle \xi | H | \xi \rangle > -\lambda_{\text{LCU}} + \Delta_{\text{LCU}}$. For these problems, by using the block encoding of $H_{\text{SA}}/\sqrt{\lambda_{\text{LCU}}}$ instead of that of H/λ_{LCU} , SA could potentially allow us to achieve an improvement in the query complexity. See the termwise SA results in Table I, proven in Theorems 5 and 7 of Supplemental Material [21]. The net effect is an improved query complexity linear in $\sqrt{\Delta_{\text{LCU}}\lambda_{\text{LCU}}}$ rather than λ_{LCU} . (We can think of $\sqrt{\Delta_{\text{LCU}}\lambda_{\text{LCU}}}$ as the new effective ℓ_1 norm when considering the low-energy subspace.) The potential improvement occurs when $\Delta_{\text{LCU}} \ll \lambda_{\text{LCU}}$, which is possible in specific instances where the Hamiltonian H is close to “frustration-free” [13].

For this example, an improvement in query complexity ultimately gives an improvement in gate complexity, because the corresponding block encodings can be implemented with similar gate costs. This would readily achieve the goal of SOSSA, however, for general and frustrated H , the termwise preprocessing outlined might result in a Δ_{LCU} that is comparable to λ_{LCU} . This severely limits the general applicability of termwise SA, and a different kind of preprocessing is desirable to have a significant reduction in gate costs.

Our strategy is then to use a different representation of the shifted Hamiltonian as an SOS of more general operators B_j :

$$H + \beta\mathbb{1} = \sum_{j=0}^{R-1} B_j^\dagger B_j. \quad (6)$$

Using this representation, which implies $H + \beta\mathbb{1} \geq 0$, gives also a path to applying SA to general Hamiltonians. Indeed, the previous termwise SA is an example of SOS where $\beta = \lambda_{\text{LCU}}$ and $B_j = |g_j|^{\frac{1}{2}} \Pi_j$. However, we can now consider the B_j ’s to be linear combinations of Pauli strings σ_j or more general operators. It is important, however, to constrain the B_j ’s so that they can be efficiently block-encoded; see the next section.

SOSSA then reduces the overall gate complexity by first reducing the query complexity via SA as much as possible. We do this by finding “good” SOS representations that have the following properties. Let

$$H_{\text{SOSSA}} := \sum_{j=0}^{R-1} |j\rangle \otimes B_j = \begin{pmatrix} B_1 \\ \vdots \\ B_R \end{pmatrix} \quad (7)$$

be such that $H_{\text{SOSSA}}^\dagger H_{\text{SOSSA}} = H + \beta \mathbb{1}$. Then, we wish to block encode $H_{\text{SOSSA}}/\sqrt{\lambda_{\text{SOS}}}$ efficiently, for some normalization factor $\lambda_{\text{SOS}} \geq \|H + \beta \mathbb{1}\|$, which is often very different from λ_{LCU} . Additionally, we wish for the lower bound $-\beta$ to be as close to the ground state energy of H as possible. For energy estimation, we will assume $\langle \psi | H | \psi \rangle \leq -\beta + \Delta_{\text{SOS}}$ and for phase estimation the low-energy state is supported on the subspace of energies in $[-\beta, -\beta + \Delta_{\text{SOS}}]$. These parameters give $\sqrt{\Delta_{\text{SOS}} \lambda_{\text{SOS}}}$, which determines the query complexity when using SA with the Hamiltonian $H + \beta \mathbb{1}$; see the query complexity results of SOSSA in Table I. The goal of the good SOS representation is to satisfy $\Delta_{\text{SOS}} \lambda_{\text{SOS}} \ll \Delta_{\text{LCU}} \lambda_{\text{LCU}}$ and hence improve upon the query complexity of termwise SA.

After obtaining improved query complexities from the SOS representation, we wish to determine the resulting gate complexities, and we need to account for the gate cost of implementing the block encoding of $H_{\text{SOSSA}}/\sqrt{\lambda_{\text{SOS}}}$. Unfortunately, this gate cost might be higher than that of implementing the block encoding of $H_{\text{SA}}/\sqrt{\lambda_{\text{SA}}}$, since the operators B_j are more general and possibly more difficult to simulate. Nevertheless, we find that these two competing effects—the improvement in query complexity versus the increase in the gate complexity of each query—can still give a significant improvement overall, as demonstrated by the examples we studied.

Hence, to obtain improved gate complexities, SOSSA uses SA in combination with good SOS representations to reduce the query complexity as much as possible, even when the gate cost per query can increase.

SOS optimization for SOSSA—To achieve greater SA, we would like Δ_{SOS} to be as small as possible, so we would like $-\beta$ to be a tight lower bound on the ground state energy. In this section, we show that optimizing the lower bound $-\beta$ can be achieved efficiently in classical preprocessing using semidefinite programming. Optimization of $-\beta$ alone does not directly optimize the total gate cost but can instead improve the query complexity. However, it can be a useful starting point as demonstrated in Ref. [20]. We will see in “Estimation of unknown ground state energies” that this optimization is sufficient to achieve asymptotic speedups in the total gate cost for the SYK model.

We can start by selecting an ansatz for the B_j ’s as low-degree polynomials of some natural operator set $\mathcal{B}$, the “SOS algebra,” of size L . For example, given the

Hamiltonian on N qubits, we can choose a constant $k \geq 1$ and let $\mathcal{B}$ be the set of $L = \mathcal{O}(N^k)$ monomials, each being a product of k Pauli operators. Then, $B_j(\vec{b}_j) \in \mathcal{B}$ denotes a polynomial of Pauli strings of weight or degree k , where $\vec{b}_j \in \mathbb{C}^L$ are the coefficients of the polynomial. For a fermionic Hamiltonian, we could let $\mathcal{B}$ be the set of degree- k products of fermionic creation and annihilation operators instead, or equivalently in the Majorana operators. Often, we may not want to include all degree- k terms in $\mathcal{B}$, but only a subset like nearest-neighbor products when considering a system in a lattice.

Having selected the set $\mathcal{B}$, we can optimize the lower bound on the ground state energy while constraining the coefficients $\vec{b}_j$ so that Eq. (6) is satisfied. This optimization can be cast as the following program:

$$\min_{\vec{b}_j \in \mathbb{C}^L} \beta \text{ s.t. } H + \beta \mathbb{1} = \sum_{j=0}^{R-1} B_j(\vec{b}_j)^\dagger B_j(\vec{b}_j),$$

which can be directly translated into an SDP whose dimension is polynomial in the operator basis size L and whose linear constraints are derived from the algebra of the polynomials; see Sec. III-A of Supplemental Material [21], which includes Refs. [26–41]. Mathematical problems of this form can be solved in polynomial time with respect to the number of variables and constraints [42–44]. The resulting number of terms R is related to the rank of the primal variable of the SDP and is also polynomial in L .

The dual problem of this SDP is the pseudomoment problem [16,45] where it is well understood that increasing the complexity of B_j by, for example, allowing a larger set of monomials in $\mathcal{B}$, and thus the size of the SDP, $-\beta$ can be made arbitrarily close to the true ground-state energy, implying a smaller energy gap Δ_{SOS} .

Estimation of unknown ground state energies—A universal quantum computational task is estimating with confidence $1 - q$ the ground state energy E to accuracy ϵ , given a trial state with finite overlap $\sqrt{p}$ with the ground state [36]. In the general setting, the initial uncertainty of E is the trivial range $[-\lambda_{\text{LCU}}, 0]$, and it is not obvious that this can be accomplished with $\sqrt{\Delta_{\text{LCU}}}$ (or $\sqrt{\Delta_{\text{SOS}}}$) scaling. The difficulty is that Δ_{LCU} is unknown *a priori*, hence by Fig. 1 we do not know what target accuracy to select in phase estimation. Nevertheless, we show that Δ_{LCU} can be learned on the fly by an adaptive algorithm in two steps.

First, we choose a coarse phase estimation target accuracy of $\epsilon' = \mathcal{O}(\sqrt{\epsilon/\lambda_{\text{LCU}}})$. To avoid suboptimal logarithmic overheads from filtering small eigenenergies and amplifying the overlap [36], this is accomplished by a careful modification of the fuzzy binary search [7]. We show in Sec. II-B of Supplemental Material [21] that this either returns with confidence $1 - (q/2)$ an estimate $\lambda + \hat{E}$ correct to constant multiplicative error or returns success by

concluding that $E \in [-\lambda_{\text{LCU}}, -\lambda_{\text{LCU}} + \epsilon]$. This uses $\mathcal{O}[\sqrt{(\lambda_{\text{LCU}}/p\epsilon)} \log(1/p) \log(1/q)]$ queries, and the $\mathcal{O}(1/\sqrt{\epsilon})$ scaling generalizes the novel super-Heisenberg results previously known for expectation estimation [33].

Second, $\lambda_{\text{LCU}} + \hat{E}$ is also an estimate $\hat{\Delta}_{\text{LCU}} = \Theta(\Delta_{\text{LCU}})$ correct to constant multiplicative accuracy. Hence we choose a fine phase estimation target accuracy $\epsilon' = \mathcal{O}(\epsilon/\sqrt{\hat{\Delta}_{\text{LCU}}\lambda_{\text{LCU}}})$. Again by fuzzy binary search, this achieves the desired accuracy ϵ with confidence $1 - q/2$ using $\mathcal{O}[(\sqrt{\Delta_{\text{LCU}}\lambda_{\text{LCU}}}/\sqrt{p\epsilon}) \log(1/p) \log(1/q)]$ queries. By a union bound, the overall algorithm fails with probability at most q , and strictly generalizes state of art in energy estimation to the regime of spectrum amplification.

Optimality of SOSSA—SOSSA uses spectral amplification to obtain runtime improvements. While we have discussed upper bounds on the runtimes of several quantum simulation tasks, in a model where we assume block-encoding access to the relevant operators, it is indeed possible to show that the complexities in Table I are also optimal. The following (informal) result establishes a minimum query cost for quantum phase estimation in this setting.

Theorem (Lower bound on quantum phase estimation using SA, informal)—Let H be a Hamiltonian that satisfies $H + \beta\mathbb{1} = H_{\text{SA}}^\dagger H_{\text{SA}}$ for some (known) $\beta \in \mathbb{R}$ and assume block-encoding access to $H_{\text{SA}}/\sqrt{\lambda}$ for some $\lambda > 0$. Then, performing quantum phase estimation on H with precision ϵ requires $\Omega(\sqrt{\lambda\Delta}/\epsilon)$ queries to the block encoding, where $\Delta \geq 0$ is the lowest energy of $H + \beta\mathbb{1}$.

The formal statement and proof of this result is in Sec. II-C of Supplemental Material [21]. Essentially we construct a reduction from the search problem to a quantum phase estimation problem, and the lower bounds in [12] can be used to establish the result. While SOSSA is not formulated in the query model, it builds upon SA and the lower bound provides evidence that further optimizations in the asymptotic complexities will be challenging in general.

Application to SYK—We analyze the performance of SOSSA on the SYK model, which describes a fermionic Hamiltonian containing all degree-4 Majorana terms with random Gaussian coefficients:

$$H_{\text{SYK}} = \frac{1}{\sqrt{\binom{N}{4}}} \sum_{a,b,c,d} g_{abcd} \gamma_a \gamma_b \gamma_c \gamma_d, \quad (8)$$

$$g_{abcd} \sim \mathcal{N}(0, 1) \text{ i.i.d.} \quad (9)$$

(Here $\{\gamma_1, \dots, \gamma_N\}$ are Majorana operators.) We will observe an asymptotic speedup in phase estimation of the ground-state energy and energy estimation over the traditional approach of LCU block encoding. The speedup reduces the gate complexity by a factor $\sqrt{N}$.

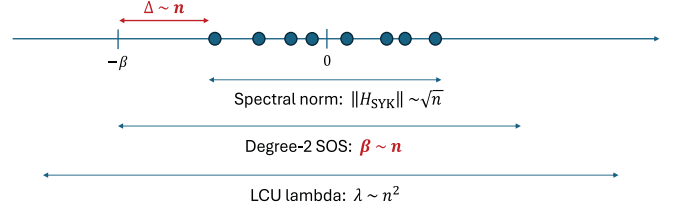


FIG. 2. This figure shows the relative scalings of the spectral norm, λ_{LCU} , and β for the SYK model. The dots in the figure represent the eigenvalues of the SYK model.

For our SOS ansatz, we take the degree-2 polynomials in Majorana operators. The asymptotic gate complexities to construct the necessary block encodings for the LCU and SOSSA approaches is $\mathcal{O}(N^4)$ in all cases, and thus we can focus on comparing the query complexities shown in Table I. We compare to the query complexity of LCU, which scales with $\lambda_{\text{LCU}} \sim N^2$.

The energy gap Δ_{SOS} between the ground energy and the SOS lower bound $-\beta$ is a key property to analyze. In Sec. IV-C of Supplemental Material [21], we prove that the degree-2 Majorana SOS achieves a lower bound scaling as $\beta = \mathcal{O}(N)$ with high probability, which guarantees $\Delta_{\text{SOS}} = \mathcal{O}(N)$. The proof is achieved by using the strong duality between the SOS program and its dual pseudomoment program. We then bound the expectation value of the pseudomoment program using random matrix theory.

Next, we must consider λ_{SOS} . In Sec. IV-C of Supplemental Material [21], we show that we can implement a block encoding of H_{SOSSA} with $\lambda_{\text{SOS}} = \mathcal{O}(N^2)$. Our block-encoding compilation makes essential use of the double factorization technique [38]. Double factorization involves finding for each B_j in the SOS ansatz a Gaussian unitary U_j which diagonalizes the quadratic portion of B_j , reducing the number of terms from $\mathcal{O}(N^2)$ to $\mathcal{O}(N)$.

Overall, we get $\sqrt{\Delta_{\text{SOS}}\lambda_{\text{SOS}}} = \mathcal{O}(N^{3/2})$, which is improved over $\lambda_{\text{LCU}} = \Theta(N^2)$ by a factor of $\sqrt{N}$.

Acknowledgments—We thank Matthew Hastings for the analysis of the performance of degree-2 Majorana SOS on the SYK model, and William Huggins for helpful discussions.

Data availability—No data were created or analyzed in this study.

- [1] E. Knill, G. Ortiz, and R. D. Somma, *Phys. Rev. A* **75**, 012328 (2007).
- [2] G. H. Low and I. L. Chuang, *Phys. Rev. Lett.* **118**, 010501 (2017).
- [3] A. Y. Kitaev, [arXiv:quant-ph/9511026](https://arxiv.org/abs/quant-ph/9511026).
- [4] S. Lloyd, *Science* **273**, 1073 (1996).
- [5] N. Wiebe, D. Berry, P. Høyer, and B. C. Sanders, *J. Phys. A* **43**, 065203 (2010).

- [6] D. W. Berry, A. M. Childs, R. Cleve, R. Kothari, and R. D. Somma, *Phys. Rev. Lett.* **114**, 090502 (2015).
- [7] D. W. Berry, Y. Tong, T. Khattar, A. White, T. I. Kim, G. H. Low, S. Boixo, Z. Ding, L. Lin, S. Lee, G. K.-L. Chan, R. Babbush, and N. C. Rubin, *PRX Quantum* **6**, 020327 (2025).
- [8] N. C. Rubin, D. W. Berry, F. D. Malone, A. F. White, T. Khattar, A. Eugene DePrince III, S. Siculo, M. Kühn, M. Kaicher, J. Lee, and R. Babbush, *PRX Quantum* **4**, 040303 (2023).
- [9] R. Babbush, J. R. McClean, M. Newman, C. Gidney, S. Boixo, and H. Neven, *PRX Quantum* **2**, 010103 (2021).
- [10] I. Loaiza and A. F. Izmaylov, *J. Chem. Theory Comput.* **19**, 8201 (2023).
- [11] J. Haah, M. B. Hastings, R. Kothari, and G. H. Low, *SIAM J. Comput.* **52**, FOCS18-i (2023).
- [12] A. Zlokapa and R. D. Somma, *Quantum* **8**, 1449 (2024).
- [13] R. D. Somma and S. Boixo, *SIAM J. Comput.* **42**, 593 (2013).
- [14] G. H. Low and I. L. Chuang, *arXiv:1707.05391*.
- [15] M. X. Goemans and D. P. Williamson, *J. ACM* **42**, 1115 (1995).
- [16] S. Pironio, M. Navascués, and A. Acín, *SIAM J. Optim.* **20**, 2157 (2010).
- [17] D. A. Mazziotti, *Acc. Chem. Res.* **39**, 207 (2006).
- [18] M. Nakata, H. Nakatsuji, M. Ehara, M. Fukuda, K. Nakata, and K. Fujisawa, *J. Chem. Phys.* **114**, 8282 (2001).
- [19] M. Navascués, S. Pironio, and A. Acín, *New J. Phys.* **10**, 073013 (2008).
- [20] G. H. Low, R. King, D. W. Berry, Q. Han, A. E. DePrince III, A. White, R. Babbush, R. D. Somma, and N. C. Rubin, *Phys. Rev. X* **15**, 041016 (2025).
- [21] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/m3fj-m4rm> for quantum simulation with sum-of-squares spectral amplification.
- [22] G. H. Low and I. L. Chuang, *Quantum* **3**, 163 (2019).
- [23] A. Gilyén, Y. Su, G. H. Low, and N. Wiebe, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing* (Phoenix AZ USA, 2019), pp. 193–204.
- [24] J. M. Martyn, Z. M. Rossi, A. K. Tan, and I. L. Chuang, *PRX Quantum* **2**, 040203 (2021).
- [25] A. M. Childs and N. Wiebe, *Quantum Inf. Comput.* **12**, 901 (2012).
- [26] D. W. Berry, M. Kieferová, A. Scherer, Y. R. Sanders, G. H. Low, N. Wiebe, C. Gidney, and R. Babbush, *npj Quantum Inf.* **4**, 22 (2018).
- [27] N. S. Mande and R. de Wolf, *arXiv:2305.04908*.
- [28] A. M. Childs, D. Maslov, Y. Nam, N. J. Ross, and Y. Su, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 9456 (2018).
- [29] Z. Jiang, A. Kalev, W. Mruczkiewicz, and H. Neven, *Quantum* **4**, 276 (2020).
- [30] R. Babbush, C. Gidney, D. W. Berry, N. Wiebe, J. McClean, A. Paler, A. Fowler, and H. Neven, *Phys. Rev. X* **8**, 041015 (2018).
- [31] Y. He, M.-X. Luo, E. Zhang, H.-K. Wang, and X.-F. Wang, *Int. J. Theor. Phys.* **56**, 2350 (2017).
- [32] G. Brassard, P. Hoyer, M. Mosca, and A. Tapp, *Contemp. Math.* **305**, 53 (2002).
- [33] S. Simon, M. Degroote, N. Moll, R. Santagati, M. Streif, and N. Wiebe, *arXiv:2402.14791*.
- [34] G. H. Low and Y. Su, *arXiv:2410.18178*.
- [35] Y. Ge, J. Tura, and J. I. Cirac, *arXiv:1712.03193*.
- [36] L. Lin and Y. Tong, *Quantum* **4**, 372 (2020).
- [37] J. Lee, D. W. Berry, C. Gidney, W. J. Huggins, J. R. McClean, N. Wiebe, and R. Babbush, *PRX Quantum* **2**, 030305 (2021).
- [38] V. von Burg, G. H. Low, T. Häner, D. S. Steiger, M. Reiher, M. Roetteler, and M. Troyer, *Phys. Rev. Res.* **3**, 033055 (2021).
- [39] M. B. Hastings and R. O'Donnell, in *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing* (Wien, Austria, 2022), pp. 776–789, [10.1145/3519935.3519960](https://doi.org/10.1145/3519935.3519960).
- [40] B. Baumgartner, *arXiv:1106.6189*.
- [41] N. C. Rubin, R. Babbush, and J. McClean, *New J. Phys.* **20**, 053020 (2018).
- [42] S. P. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, England, 2004).
- [43] S. Burer and R. D. Monteiro, *Math. Program.* **95**, 329 (2003).
- [44] J. Povh, F. Rendl, and A. Wiegele, *Computing* **78**, 277 (2006).
- [45] R. M. Erdahl, *Int. J. Quantum Chem.* **13**, 697 (1978).