

Speeding Up Learning Quantum States Through Group Equivariant Convolutional Quantum Ansätze

Han Zheng^{1,2,*}, Zimu Li^{2,†}, Junyu Liu^{3,4,5,‡}, Sergii Strelchuk^{2,§} and Risi Kondor^{1,6,7,¶}

¹Department of Statistics, The University of Chicago, Chicago, Illinois 60637, USA

²DAMTP, Center for Mathematical Sciences, University of Cambridge, Cambridge CB30WA, United Kingdom


³Pritzker School of Molecular Engineering, The University of Chicago, Chicago, Illinois 60637, USA

⁴Chicago Quantum Exchange, Chicago, Illinois 60637, USA

⁵Kadanoff Center for Theoretical Physics, The University of Chicago, Chicago, Illinois 60637, USA

⁶Department of Computer Science, The University of Chicago, Chicago, Illinois 60637, USA

⁷Flatiron Institute, New York City, New York 10010, USA

 (Received 18 January 2022; revised 3 October 2022; accepted 13 March 2023; published 15 May 2023)

We develop a theoretical framework for S_n -equivariant convolutional quantum circuits with $SU(d)$ symmetry, building on and significantly generalizing Jordan's permutational quantum computing formalism based on Schur-Weyl duality connecting both $SU(d)$ and S_n actions on qudits. In particular, we utilize the Okounkov-Vershik approach to prove Harrow's statement on the equivalence between $SU(d)$ and S_n irrep bases and to establish the S_n -equivariant convolutional quantum alternating ansätze (S_n -CQA) using Young-Jucys-Murphy elements. We prove that S_n -CQA is able to generate any unitary in any given S_n irrep sector, which may serve as a universal model for a wide array of quantum machine-learning problems with the presence of $SU(d)$ symmetry. Our method provides another way to prove the universality of the quantum approximate optimization algorithm and verifies that four-local $SU(d)$ -symmetric unitaries are sufficient to build generic $SU(d)$ -symmetric quantum circuits up to relative phase factors. We present numerical simulations to showcase the effectiveness of the ansätze to find the ground-state energy of the J_1 - J_2 antiferromagnetic Heisenberg model on the rectangular and kagome lattices. Our work provides the first application of the celebrated Okounkov-Vershik S_n representation theory to quantum physics and machine learning, from which to propose quantum variational ansätze that strongly suggests to be classically intractable tailored towards a specific optimization problem.

DOI: [10.1103/PRXQuantum.4.020327](https://doi.org/10.1103/PRXQuantum.4.020327)

I. INTRODUCTION

The combination of new ideas from machine learning and recent developments in quantum computing has led to an impressive array of new applications. [1–13]. Prominent examples of this interplay are the variational quantum eigensolver (VQE) and quantum approximate optimization algorithm (QAOA) [14–16], which are considered among the most promising quantum machine-learning approaches

in the noisy intermediate-scale quantum (NISQ) [17] era. VQEs and QAOA have shown tremendous promise in quantum simulation and quantum optimization [18–23].

One of the most important neural-network architectures in classical machine learning are convolutional neural networks (CNNs) [24–28]. In recent years, CNNs have also found applications in condensed-matter physics and quantum computing. For instance, Ref. [29] proposes a quantum convolutional neural network with $\log N$ parameters to solve topological symmetry-protected phases in quantum many-body systems, where N is the system size. One of the key properties of classical CNNs is equivariance, which roughly states that if the input to the neural network is shifted, then its activations translate accordingly. There have been several attempts to introduce theoretically sound analogs of convolution and equivariance to quantum circuits, but they have generally been somewhat heuristic. The major difficulty is that the translation invariance of CNNs lacks a mathematically rigorous quantum counterpart due to the discrete spectrum of spin-based

*hanz98@uchicago.edu

†lizm@mail.sustech.edu.cn

‡junyuliu@uchicago.edu

§ss870@cam.ac.uk

¶risi@cs.uchicago.edu

Published by the American Physical Society under the terms of the *Creative Commons Attribution 4.0 International* license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

quantum circuits. For example, Ref. [29] uses the quasilocal unitary operators to act vertically across all qubits.

In quantum systems there is a discrete set of translations corresponding to permuting the qudits as well as a continuous notion of translation corresponding to spatial rotations by elements of $SU(d)$. Combining these two is the realm of so-called permutational quantum computing (PQC) [30]. Therefore, a natural starting point for realizing convolutional neural networks in quantum circuits is to look for *permutation equivariance*. In one of our related works [31], we argued that the natural form of equivariance in quantum circuits is permutation equivariance and we introduce a theoretical framework to incorporate group-theoretical CNNs into the quantum circuits, building on and generalizing the PQC framework to what we call PQC+ [31].

In this paper, we further explore PQC+ and its significance for machine-learning applications. Roughly speaking, the PQC+ machine consists of unitary time evolutions of k -local $SU(d)$ -symmetric Hamiltonian. As a feature, *Schur-Weyl duality* between the aforementioned S_n and $SU(d)$ actions on qudit systems appears naturally and is used throughout the paper. Most importantly, it indicates that any $SU(d)$ -symmetric quantum circuits can be expressed in S_n irreducible representations (irreps). Exploiting the power of S_n representation theory in quantum circuits towards NISQ applications is thus the central theme of the paper. The representation theory of S_n has been found to be a powerful tool in various permutation-equivariant learning tasks, e.g., learning set-valued functions [32] and learning on graphs [33,34]. Most applications of permutation-equivariant neural networks work with a subset of representations of S_n . In contrast, in physical and chemical models where the Hamiltonian exhibits global $SU(d)$ symmetry, such as the Heisenberg model, it is necessary to consider *all* the S_n irreps (a detailed explanation of this significant insight can be found in Sec. V). However, even the best classical fast Fourier transforms (FFTs) over the symmetric group S_n require at least $\mathcal{O}(n!n^2)$ operations [35,36], which dashes any hope of calculating the Fourier coefficients even for relatively small n . Indeed, despite increasing realization of the importance of enforcing $SU(2)$ symmetry, none of the neural-network quantum state (NQS) ansätze are able to respect $SU(2)$ symmetry for all $SU(2)$ irreps, due to the superpolynomial growth of the multiplicities of irreps and the superexponential cost to compute Fourier coefficients over S_n . Finding variational ansätze respecting continuous rotation symmetry is desirable because it not only helps to gain important physical insights about the system but also leads to more efficient simulation algorithms [37,38].

Motivated by the class of problems with a global $SU(d)$ symmetry, in Sec. III, we construct what we call the variational S_n -equivariant convolutional quantum alternating ansätze (S_n -CQA), which are products of

alternating exponentials of certain Hamiltonians admitting $SU(d)$ symmetry. This is a concrete example of the PQC+ framework and may also be thought as a special case of QAOA with $SU(d)$ symmetry. Using the *Okounkov-Vershik approach* [39] to S_n representation theory as well as other classical results from the theory of Lie group and Lie algebra [40,41] we prove that S_n -CQA generates any unitary matrix in each given S_n -irrep block decomposed from an n -qudit system, hence it acts as a *restricted universal* variational model for problems that possess global $SU(d)$ symmetry (Theorem 1). Consequently, it can be applied to a wide array of machine learning and optimization tasks that exhibit global $SU(d)$ symmetry or require explicit computation of high-dimensional S_n irreps, presenting a quantum superexponential speedup. Our proof techniques are of independent interest and we provide two more applications. It is shown in Refs. [42,43] that QAOA ansätze generated by simple local Hamiltonians are universal in the common sense. Forgetting the imposed symmetry, we use our techniques to derive the universality for a different but related class of QAOA ansätze with a richer set of mixer Hamiltonians (Theorem 2). In addition, we find a *four-local* S_n -CQA model, which is universal to build any $SU(d)$ symmetric quantum circuits up to phase factors (Theorem 3) with awareness of the fact that *two-local* $SU(d)$ -symmetric unitaries cannot fulfill the task when $d \geq 3$ [44,45]. Consequently, when compared with other $SU(d)$ -symmetric ansätze, products of exponentials of SWAPs (eSWAPs) proposed in Ref. [46] admit the restricted universality with $SU(d)$ symmetry only when $d = 2$ and the CQA model is universal in general cases when restricted to any one of S_n irreps.

In Sec. IV we explore more details about Schur-Weyl duality on qudit systems. To be specific, talking about S_n or $SU(d)$ irrep blocks in a qudit system requires using the *Schur basis*, instead of the computational basis. The Schur basis can be constructed by either $SU(d)$ Clebsch-Gordan decomposition [30,47,48] or by the S_n branching rule [49–51], which yields two ways to label the basis elements by either $SU(d)$ Casimir operators or by the so-called *YJM elements* used in the Okounkov-Vershik approach. We rigorously demonstrate the equivalence of these labeling schemes (Theorem 4), first conjectured in Harrow’s thesis [47], (see also discussions by Bacon, Chuang, and Harrow [52] and Krovi [51]). As a result, we find a state-initialization method, using constant-depth qudit circuits, to produce linear combinations of Schur-basis vectors, which may be preferred in NISQ devices rather than implementing a quantum Schur transform (QST). We show that the measurements taken for variationally updating parameters in S_n -CQA can be efficiently calculated on the Schur basis, while a similar conclusion is unlikely to be drawn classically.

In the numerical part, Sec. V, we illustrate the potential of this framework by applying it to the problem of

finding the ground-state energy of J_1 - J_2 antiferromagnetic Heisenberg magnets, a gapless system with no known sign structure or analytical solution in quantum many-body theory. We compare our model with classical and quantum algorithms like Refs. [38,46,53]. We emphasize the consequence of the failure of the *Marshall-Lieb-Mattis theorem* [54–56] in the frustrated region with which classical neural networks struggle to discover the sign structure to an admissible accuracy due to violation of global SU(2) symmetry [57,58]. We include numerical simulation to show the effectiveness of the S_n -CQA ansätze in finding the ground state with frustration using only $\mathcal{O}(pn^2)$ parameters for p alternating layers. Noisy simulations are also provided to show the robustness of S_n -CQA.

Our theoretical results about SU(d) symmetry can be reversed to exhibit S_n permutation symmetry. We define SU(d)-CQA on SU(d)-irrep blocks and leave it to future research work to explore its theoretical and experimental potential. All statements and theorems discussed in the main text are proved in full detail in the Appendix.

II. BACKGROUND ON REPRESENTATION THEORY OF THE SYMMETRIC GROUP

In this section we define some of the mathematical concepts and notations used in the rest of the paper. Further details can be found in Refs. [49,50,59].

Let V be a d -dimensional complex Hilbert space with orthonormal basis $\{e_1, \dots, e_d\}$. The tensor-product space $V^{\otimes n}$ admits two natural representations: the *tensor-product representation* $\pi_{\text{SU}(d)}$ of SU(d) acting as

$$\pi_{\text{SU}(d)}(g)(e_{i_1} \otimes \dots \otimes e_{i_n}) := g \cdot e_{i_1} \otimes \dots \otimes g \cdot e_{i_n},$$

where $g \cdot e_{i_k}$ is the fundamental representation of SU(d), and the *permutation representation* π_{S_n} of S_n acting as

$$\pi_{S_n}(\sigma)(e_{i_1} \otimes \dots \otimes e_{i_n}) := e_{i_{\sigma^{-1}(1)}} \otimes \dots \otimes e_{i_{\sigma^{-1}(n)}}.$$

We treat $V^{\otimes n}$ as the Hilbert space of an n -qudit system. The so-called Schur-Weyl duality reveals how the above two representations are related.

Schur-Weyl duality is widely used in quantum computing [60], quantum information theory [47], and high-energy physics [61]. In particular, in quantum chromodynamics it was used to decompose the n -fold tensor product of SU(3) representations. In that context, standard Young tableaux are referred to as Weyl tableaux and labeled by the three isospin numbers (u, d, s) . The underlying Young diagrams containing three rows $\lambda = (\lambda_1, \lambda_2, \lambda_3)$ are used to denote an SU(d) irreducible representation (irrep). There is another way using Young diagrams $\lambda' = \lambda_1 - \lambda_3, \lambda_2 - \lambda_3$ labeled by Dynkin integers via the highest weight vectors. In short, there are two conventions in the literature to denote SU(d) irreps. On the other hand, S_n irreps can also

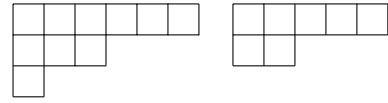


FIG. 1. Young diagrams $\lambda = (6, 3, 1)$ and $\lambda' = (5, 2)$ represent the same SU(3)-irrep of highest weight (3, 2). However, in the context of Schur-Weyl duality, λ corresponds to an S_{10} irrep while λ' gives an S_7 irrep.

be denoted by Young diagrams [59]. Schur-Weyl duality says that irreps of SU(d) and S_n are dual in the following sense and denoted by the same Young diagrams *with n boxes and at most d rows*.

Theorem (Schur-Weyl duality): *The action of SU(d) and S_n on $V^{\otimes n}$ jointly decompose the space into irreducible representations of both groups in the form*

$$V^{\otimes n} = \bigoplus_{\lambda} W_{\lambda} \otimes S^{\lambda},$$

where W_{λ} and S^{λ} denote irreps of SU(d), respectively, S_n , and λ ranges over all Young diagrams of size n with at most d rows. Consequently,

$$\pi_{\text{SU}(d)} \cong \bigoplus_{\mu} W_{\mu} \otimes 1_{m_{\text{SU}(d),\mu}}, \quad \pi_{S_n} \cong \bigoplus_{\lambda} 1_{m_{S_n,\lambda}} \otimes S^{\lambda},$$

where $m_{\text{SU}(d),\mu} = \dim S^{\mu}$ and $m_{S_n,\lambda} = \dim W_{\lambda}$.

One can easily verify that $\pi_{\text{SU}(d)}$ and π_{S_n} commute (further properties are described in the Appendix). Consider the *symmetric group algebra* $\mathbb{C}[S_n]$ consisting of all formal finite sums $f = \sum_i c_i \sigma_i$. Its representation is then $\tilde{\pi}_{S_n}(f) = \sum_i c_i \pi_{S_n}(\sigma_i)$. When there is no ambiguity, we denote by U_{σ} or simply σ the representations $\pi_{S_n}(\sigma)$.

Working from the perspective of Schur-Weyl duality requires, at least theoretically, using the *Schur basis* rather than the computational basis. A conventional way to build such a basis is conducting sequential coupling and Clebsch-Gordan decompositions of SU(d) representations [30,47,48], which transform common matrix representations of SU(d) into irrep matrix blocks like in Fig. 2(a). Since our focus are ansätze, operators, and quantum circuits with SU(d) symmetry, which commute with $\pi_{\text{SU}(d)}$, and since Schur-Weyl duality and the double commutant theorem (see Appendix) say that they must be established from the group algebra $\mathbb{C}[S_n]$, we need to explore S_n irreps blocks as in Fig. 2(b) however. We are going to introduce a method to decompose permutation matrices in this picture and explain basic notions in S_n representation theory and the Okounkov-Vershik approach, since they are essential to understand the theoretical results in this paper.

We first consider the so-called *permutation module* M^{μ} . In the case of qubits, permutation modules correspond

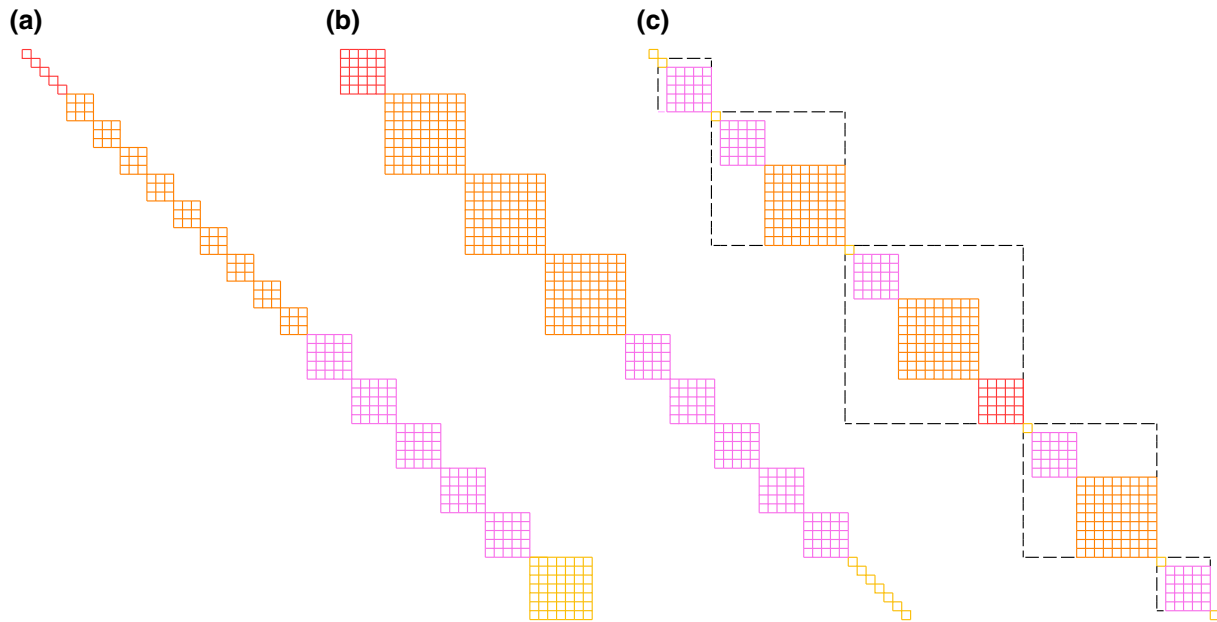


FIG. 2. (a) Decomposition of $(\mathbb{C}^2)^6$ with respect to $SU(2)$ action while (b) is for S_6 by Schur-Weyl duality. (c) Arrangement of (b), which respects both permutation modules and S_6 irreps.

to sets of Schur-basis elements having different readout on total spin components. Fortunately there is an accessible way to understand M^μ in the tensor-product space $V^{\otimes n}$. To make things simpler, consider the $d = 2$ case of $SU(2) - S_n$ duality on $(\mathbb{C}^2)^{\otimes n}$. Only two-row Young diagrams $\lambda = (\lambda_1, \lambda_2)$ appear in this duality and the half of difference $\frac{1}{2}(\lambda_1 - \lambda_2)$ between the lengths of the two rows gives the total spin of the $SU(2)$ -irrep W_λ . The permutation module M^μ is isomorphic with the linear span of all computational basis vectors with z -spin components equal to $\frac{1}{2}(\mu_1 - \mu_2)$.

Note that, $(\mathbb{C}^2)^{\otimes n} = \bigoplus_\mu M^\mu$ and each M^μ is invariant under S_n permutation. Furthermore, M^μ can be further decomposed into S_n irreps. For two-row Young diagrams [$SU(2) - S_n$ duality], the decomposition is easy: $M^\mu = \bigoplus_{\lambda \geq \mu} S^\lambda$ where λ, μ have the same size n and we use the dominance order $\lambda \geq \mu$ if $\lambda_1 \geq \mu_1$. In summary, we have

$$(\mathbb{C}^2)^{\otimes n} = \bigoplus_\mu M^\mu = \bigoplus_\mu \bigoplus_{\lambda \geq \mu} S^\lambda \cong \bigoplus_\lambda 1_{m_{S_n, \lambda}} \otimes S^\lambda.$$

Isomorphic copies of S^λ come from different permutation modules. The largest permutation module contains all distinct S_n irreps in $(\mathbb{C}^2)^{\otimes n}$ [see e.g., Fig. 2(c)]. For general Young diagrams, decompositions of M^μ would have nontrivial multiplicities [50,59].

Each S^λ can be decomposed further with respect to $S_{n-1} \subset S_n$ as $S^\lambda = \bigoplus_\rho S^{\lambda, \rho}$, where $S^{\lambda, \rho}$ denotes an S_{n-1} irrep of Young diagram ρ (with $n - 1$ boxes) contained in the S_n irrep S^λ . The so-called *branching rule* guarantees that the decomposition is multiplicity-free, i.e., each

distinct S_{n-1} -irrep $S^{\lambda, \rho}$ appears only once in the decomposition. The so-called Bratteli diagrams in Fig. 3 show how different irreps are decomposed. Continuing the decomposition process for S_{n-2}, \dots, S_1 , the original space S^λ will be written as a direct sum of one-dimensional subspaces (S_1 irreps are one dimensional):

$$S^\lambda = \bigoplus_{S_{n-1}, \rho} S^{\lambda, \rho} = \dots = \bigoplus_{S_{n-1}, \rho} \dots \bigoplus_{S_1, \tau} S^{\lambda, \rho, \sigma, \dots, \tau}.$$

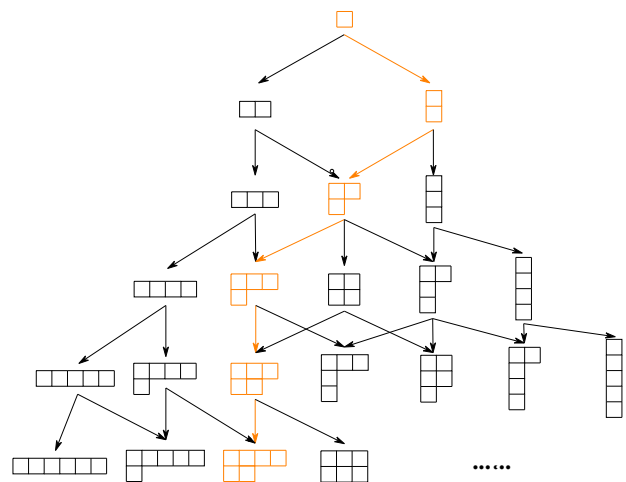


FIG. 3. Bratteli diagram for S_6 . Upper Young diagrams connecting by arrows to lower ones arise from the decomposition. Orange arrows form a path. Note that diagrams with more than two rows cannot appear in $SU(2) - S_6$ duality.

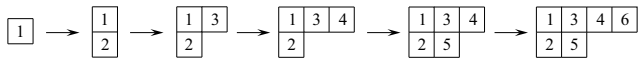


FIG. 4. Standard Young tableau defined by the path in Fig. 3.

Each one-dimensional subspace $S^{\lambda\rho\sigma\cdots\tau}$ can be represented by a nonzero vector in it. Normalizing them, we obtain an orthonormal basis $\{|v_T\rangle\}$ of S^λ called the *Gelfand-Tsetlin basis (GZ)* or *Young-Yamanouchi basis*. Indices $\lambda, \rho, \sigma, \dots, \tau$ form a path in the Bratteli diagram (see Fig. 3) and can be used to define a standard Young tableau T (Fig. 4). Young-basis vectors are in one-to-one correspondence with standard Young tableaux [59]. The branching rule is also discussed in $SU(d)$ representation theory and some authors refer to the $SU(d)$ -irrep basis as the GZ basis if it is constructed in a similar manner. A more comprehensive discussion about building Schur basis by $SU(d)$ Clebsch-Gordon decomposition and by S_n branching rule is postponed in Sec. IV.

Let us introduce the central concept used in our work. The Schur basis, GZ basis, and Young basis elements are in fact synonyms of each other. In what follows we use the name Young basis elements.

Definition 1: For $1 < k \leq n$, the *Young-Jucys-Murphy element*, or *YJM element* for short, is defined as a sum of transpositions $X_k = (1, k) + (2, k) + \cdots + (k-1, k) \in \mathbb{C}[S_n]$. We set $X_1 = 0$ as a convention.

As the name indicates, this concept was developed by Young [62], Jucys [63], and Murphy [64]. Okounkov and Vershik showed that YJM elements generate the *Gelfand-Tsetlin subalgebra* $GZ_n \subset \mathbb{C}[S_n]$ [39], which is a maximal commutative subalgebra consisting of all centers $Z[S_k]$ of $\mathbb{C}[S_k]$ for $k = 1, \dots, n$. Another striking fact is that all YJM elements are strictly diagonal (indeed, representation of GZ_n consists of all diagonal matrices) in a Young basis $\{|v_T\rangle\}$ whose eigenvalues can be read out directly from the standard Young tableaux T . To be precise, let λ be a Young diagram. Since this is a two-dimensional diagram, we can naturally assign integer coordinates to its boxes. The *content* of each of its boxes is determined by the x coordinate minus y coordinate. Suppose T is a standard Young tableau of λ . Arranging all contents with respect to T , we obtain the *content vector* α_T . For instance, the Young diagram $\lambda = (4, 2)$ has contents $0, 1, 2, -1$. The specific standard tableau T in Fig. 4 has content vector $\alpha_T = (0, -1, 1, 2, 0, 3)$. Let $|v_T\rangle$ denote a Young-basis vector corresponding to T . Measuring by YJM elements, we have $X_1 |v_T\rangle = 0$, $X_2 |v_T\rangle = -|v_T\rangle$, $X_3 |v_T\rangle = |v_T\rangle$ and so forth. Each Young-basis vector is determined uniquely by its content vector (see Refs. [39,50] for more details).

III. S_n CONVOLUTIONAL QUANTUM ALTERNATING ANSÄTZE

Consider the YJM elements $\{X_1, \dots, X_n\}$ introduced in Sec. II, which generate the maximally commuting subalgebra GZ_n . The use of YJM elements allows us to design the following *mixer Hamiltonian*:

$$H_M = \sum_{i_1, \dots, i_N} \beta_{i_1 \dots i_N} X_{i_1} \cdots X_{i_N}. \quad (1)$$

This YJM Hamiltonian is still strictly diagonal under the Young basis. As there is an efficient quantum Schur transform U_{Sch} (QST), which transforms the computational basis to Schur basis even for qudits with gate complexity $\text{Poly}(n, \log d, \log(1/\epsilon))$ [47,51,52]. It is reasonable to assume that we can initialize any Young-basis element $|\Psi_{\text{init}}\rangle$ from S_n irreps via QST. Moreover, given a problem Hamiltonian H_P with $SU(d)$ symmetry, it can be written as $H_P = \tilde{\pi}(f)$ for some $f \in \mathbb{C}[S_n]$ by the double commutant theorem and Wedderburn theorem (see Sec. II and Appendix). Inspired by QAOA (Pauli X is diagonal with respect to $|+\rangle^{\otimes n}$) [14], we thus propose the following ansatz:

$$\cdots \exp(-iH_M) \exp(-i\gamma H_P) \exp(-iH_M) \\ \times \exp(-i\gamma' H_P) \cdots$$

To summarize, we define a family of mixer operators H_M parameterized by β_{i_1, \dots, i_N} , which is diagonal under the Young basis and naturally preserves each S_n irrep determined by the initialized states. Following Refs. [65,66] and one of our results [31], which interpret quantum circuits as natural Fourier spaces, we call this family of ansätze S_n -CQA.

Deviating from the original purpose of using QAOA to solve constraint satisfaction problems, our S_n -CQA are applied to problems with global $SU(d)$ symmetry or permutation equivariance, as long as the problem Hamiltonian H_P can be efficiently simulated in the circuit model. Indeed, most practical examples involve only two- or three-local spin interactions, such as the Heisenberg model studied in Sec. V, and thus by Theorem 2 of Ref. [31] can be efficiently simulated. On the other hand, the k -dependent constant in Corollary 1 from Ref. [31] has exponential scaling. So in practice, we would like to have three- or at most four-local terms for the YJM-Hamiltonian simulation in ansätze. We focus on the following mixer Hamiltonian consisting of only first- and second-order products of YJM elements:

$$H_M = \sum_{k \leq l} \beta_{kl} X_k X_l, \quad (2)$$

whose evolution can be efficiently simulated in $\mathcal{O}(n^4 \log(n^4/\epsilon)/\log \log(n^4/\epsilon))$. Below, we prove that with a mixture as in Eq. (2), the S_n -CQA ansätze are able to approximate any unitary from every S_n -irrep block [Fig. 2(b)]. This can be seen as a restricted version of universal quantum computation to S_n irreps. Since the 4-local S_n -CQA is an all-you-need approximation algorithm within PQP+, it is strongly suggestive that the PQP+ class proposed in Ref. [31] contains circuits that can approximate matrix elements of all the S_n Fourier coefficients, for a polynomial number of alternating layers p . Moreover, the four-local S_n -CQA is also the universal approximator for solutions of the problem with global $SU(d)$ symmetry, such as the Heisenberg models, due to its nature as variational ansätze.

A. Restricted universality of S_n -CQA ansätze in S_n irreps

We now present the main theoretical result of this paper: the S_n -CQA ansätze approximate any unitary in any S_n irrep decomposed from the system of qudits. This restricts universal quantum computation on $U(d^n)$ to S_n -irrep blocks [Fig. 2(b)] because our ansätze preserve $SU(d)$ symmetry. This is of interest for three reasons: (a) the density result indicates that S_n -CQA ansätze is a universal approximator in PQP+ proposed in Ref. [31] and it is the theoretical guarantee of our numerical simulations. (b) The result is valid for qudits under $SU(d) - S_n$ duality and we show the advantage of working with S_n as there is no need to deal with complicated $SU(d)$ symmetry, generators in the proof. (c) When changed from the Young basis to the computational basis, i.e., forgetting the $SU(d)$ symmetry, our results form a new proof to the universality of a broad class of QAOA ansätze. (d) It is shown in a recent work [45] that $SU(d)$ -invariant and symmetric quantum circuits with $d \geq 3$ cannot be generated by two-local $SU(d)$ -invariant unitaries. With the focus on locality, we verify that S_n -CQA ansätze can be built by four-local $SU(d)$ -invariant unitaries and four-locality is enough to generate any $SU(d)$ -invariant quantum circuit up to phase factors.

Mathematically, we aim to show that the subgroup generated by S_n -CQA ansätze is equal to the unitary group $U(S^\lambda)$ restricted to S^λ decomposed from $V^{\otimes n}$. However, arguing directly on the level of the Lie group is complicated. Instead, we prove that the generated Lie algebra is isomorphic with the unitary algebra $\mathfrak{u}(S^\lambda)$ restricted to S^λ . Then combining with some classical results from the theory of Lie group [40,41] and the Okounkov-Vershik approach to S_n -representation theory [39], we complete the proof. We outline our results here and put all the proof details into the Appendix.

Our first step is motivated by a classical result from the theory of Lie algebra: any semisimple Lie algebra can be generated by only two elements [40]. Finding these

elements would be tricky and encoding them by a quantum circuit would even be infeasible, so we adopt a different routine and solve these problems gradually. We first work on the complex general linear algebra $\mathfrak{gl}(d, \mathbb{C})$ which is not semisimple, but facilitates our proof. To begin with, it is easy to find its Cartan subalgebra—the collection $\mathfrak{d}(d)$ of all diagonal matrices. Let M be a matrix with nonzero off-diagonal elements c_{ij} . It can be thought of as a perturbation from $\mathfrak{d}(d)$. We want to know how large the subalgebra generated by M and $\mathfrak{d}(d)$ would be. More precisely, as the following.

Lemma 1: *Let $E_{ij} \in \mathfrak{gl}(d, \mathbb{C})$ be the matrix unit with entry 1 at (i, j) and 0 elsewhere. Given any matrix M , let $\mathcal{I} \subset \{1, \dots, d\} \times \{1, \dots, d\}$ be the index set corresponding to nonzero off-diagonal entries c_{ij} of M . Then the Lie subalgebra generated by $\mathfrak{d}(d)$ and M contains*

$$\mathfrak{d}(d) \oplus \left(\bigoplus_{(i,j) \in \mathcal{I}} R_{ij} \right),$$

where R_{ij} is the one-dimensional root space spanned by E_{ij} .

Intuitively speaking, GZ_n defined in Sec. II corresponds precisely to the Cartan subalgebra of $\mathfrak{gl}(\dim S^\lambda, \mathbb{C})$. In proving Lemma 1, we are required to use all basis elements of GZ_n rather than the n YJM generators [39]. Thus we need to employ all high-order products $X_{i_1} \cdots X_{i_n}$ of YJM elements (as $\dim S^\lambda$ increases exponentially for a large number of qudits Fig. 9) and that pose the first problem for a practical ansatz design, which requires k -local $\mathbb{C}[S_n]$ Hamiltonian in order to be efficiently simulated by quantum circuits [31]. This problem is solved in Lemma 2.6 in the Appendix with the help of the Okounkov-Vershik theorem [39,50]. We prove that the collection $\{X_i, X_k X_l\}$ of first- and second-order YJM-elements, while in general cannot form a basis for GZ_n , are enough to establish Lemma 1. As a reminder, merely taking the original YJM elements X_i is not sufficient and we provide counterexamples in the Appendix. As X_i are two local, this result also provides some insights on the fact that two-local $SU(d)$ -invariant unitaries cannot generate all quantum circuits with $SU(d)$ symmetry in the general case [44,45,67].

As another ingredient of S_n -CQA ansätze, the problem Hamiltonians H_P of interest are complicated in general and hard to diagonalize classically. It also forms the other part (the matrix M) in generating the Lie algebra in Lemma 1. For the purpose of easy implementation, we show in Lemma 2 that H_P needs only to be *path connected* or *irreducible* in the language of graph theory. A Hamiltonian is of this kind if its associated index graph \mathcal{G}_{H_P} is connected. For example, the Pauli X and Y are path connected while Z is not. We further prove in Lemma 3.1 in the Appendix that the two-local Hamiltonian $H_S = \sum_{i=1}^{n-1} (i, i+1)$ defined by

all adjacent transpositions $(i, i + 1) \in S_n$ is path connected. We discuss path connectedness further after Theorem 1 as well as in Sec. V. It is also seen in the famous Perron-Frobenius theorem and applied to graph theory.

Lemma 2: *Let H_P be a path-connected Hamiltonian. Then the generated Lie algebra $\langle \partial(d), H_P \rangle = \mathfrak{gl}(d, \mathbb{C})$. Consider $\mathfrak{d}_{\mathbb{R}}(d)$ consisting of all real-valued diagonal matrices. Generated over \mathbb{R} , $\langle i\mathfrak{d}_{\mathbb{R}}(d), iH_P \rangle_{\mathbb{R}} = \mathfrak{u}(d)$.*

Since YJM elements as well as their high-order products have real diagonal entries under Young basis, we concretize $\partial(d)$ by $\{X_i, X_k X_l\}$. With all these preparations, we consider the subgroup H defined in pure algebraic sense by alternating exponentials of $iX_k X_l$ and iH_P where H_P is path connected. To verify that H is a Lie group (with smooth structures [68]), we apply another classical theorem due to Yamabe [41, 69] and conclude with the following.

Theorem 1: *Restricted to any S^λ with isomorphic copies decomposed from $V^{\otimes n}$, the subgroup generated by $X_k X_l$ with any path-connected Hamiltonian H_P equals $U(S^\lambda)$. Then a S_n -CQA ansatz is written as*

$$\begin{aligned} & \cdots \exp \left(-i \sum_{k,l} \beta_{kl} X_k X_l \right) \exp(-i\gamma H_P) \\ & \exp \left(-i \sum_{k,l} \beta'_{kl} X_k X_l \right) \exp(-i\gamma' H_P) \cdots, \quad (3) \end{aligned}$$

where we redefine X_1 as I with which any first-order YJM element X_i can be written as $X_i X_1$.

Consider the case when H_P is not path connected. That is, H_P is block diagonal (after a possible re-cording of basis elements) in S^λ . It is straightforward to check that Theorem 1 still holds within each sub-block of H_P . Suppose our task is to find the lowest eigenstate $|v_0\rangle$ of H_P within S^λ . There is generally no prior knowledge about which sub-block v_0 is in. The brute-force way to find the minimum is by taking a collection of initial states from each of these sub-blocks and applying the theorem repeatedly. One way to do this is by implementing the efficient QST, which gives us access to all Young-basis elements. The state initialization proposed in Sec. IV B with constant depth may take a hit (forcing the depth of the circuit to increase) if the problem Hamiltonian is not path connected.

B. Universality of QAQA

The first proof of the universality of the QAQA ansätze was given in Ref. [42], where the authors considered the problem Hamiltonian of the first-order and second-order nearest-neighbor interaction. References [43] subsequently

generalized the result to broader families of ansätze defined by sets of graphs and hypergraphs. We now describe a new proof based on the techniques developed in this paper that covers a novel, broader family of QAQA ansätze. More precisely, we change to the computational basis $\{|e_i\rangle\}_{i=1}^{2^n}$ in which all tensor products $\tilde{Z}_{r_1 \dots r_s} := Z_{r_1} \otimes \cdots \otimes Z_{r_s}$ of Pauli basis can span any diagonal matrix. In the language of Lie algebra, Z_i generates the Cartan subalgebra $\mathfrak{d}(2^n)$ of $\mathfrak{gl}(2^n, \mathbb{C})$ [comparing with the case of $\text{GZ}_n = \mathfrak{d}(S^\lambda)$ under Young basis]. Let H_X be the uniform summation of Pauli- X operators (we do not write its explicit form to avoid any confusion with the notation of YJM elements). The Hamiltonian H_X is path connected under $\{|e_i\rangle\}$. To restrict the use of high-order Pauli Z operators analogously as we did for YJM elements, we prove in Lemma 5.1 that the Hamiltonian H_Z composed by $\{Z_i, Z_k Z_l\}$ are enough to establish Lemmas 1 and 2 with H_X in the present setting. Unlike the H_Z used in Ref. [42], which contains only nearest-neighbor terms $Z_j Z_{j+1}$, we take all second-order products $Z_k Z_l$ in our proof. The resulting Hamiltonian H_Z is still simple though and the proof works for both an odd and even number of qubits [43]. Moreover, replacing H_X by any other path-connected Hamiltonian, e.g., an unfrustrated Heisenberg Hamiltonian with boundary condition [56], still guarantees the universality, and this fact enables one to experiment with a wide range of mixer Hamiltonians. In summary,

Theorem 2: *Let H_X be any path-connected Hamiltonian on computational basis, the group generated by the QAQA-ansatz with $H_X, H_Z = \sum \beta_{kl} Z_k Z_l$ equals $U(2^n)$, i.e., it is universal.*

C. Four-locality of $SU(d)$ -symmetric quantum circuits

A well-known result in Ref. [70] states that any quantum circuit can be generated by two-local unitaries for qubits as well as for qudits. It has been shown in a recent work [45] that this statement fails to hold when we impose the $SU(d)$ symmetry on the qudit system with $d \geq 3$. Let \mathcal{V}_k denote the subgroup generated by k -local $SU(d)$ -invariant unitaries, so $\mathcal{V}_2 \neq \mathcal{V}_n$, where \mathcal{V}_n stands for all the irrep blocks from Fig. 2(b). On the other hand, we use $U(S^\lambda)$ in Theorem 1, which specifies one (with equivalent copies) of them in searching ground state of Heisenberg Hamiltonian in Sec. V.

Counting all inequivalent S^λ is an interesting problem of its own, especially when studying the subgroups $\mathcal{V}_k \subset U(d^n)$ induced by symmetry, but it would cause a *phase factor problem*: one may not be able to manipulate relative phase factors of unitaries generated in inequivalent S^λ arbitrarily. We could simply ignore these phase factors as they make no difference in measurements respecting the symmetry. Then we consider $\mathcal{S}\mathcal{V}_k \subset \mathcal{V}_k$ restricted to $SU(S^\lambda)$ for all S^λ decomposed from $V^{\otimes n}$. It is shown in

Refs. [44,67,71,72] that $\mathcal{SV}_2 = \mathcal{SV}_n$ when $d = 2$. However, Refs. [44,67] prove in a pure math flavor by Brauer algebra from representation theory that the statement fails when $d \geq 3$ and Ref. [45] shows by constructing a counterexample based on the qudit-fermion correspondence that \mathcal{V}_2 is not even a 2-design. That is, the distribution of unitaries generated by random two-local unitaries cannot converge to the Haar measure of \mathcal{V}_n . With results about CQA developed above, we prove the following theorem.

Theorem 3: *Ignoring phase factors, $SU(d)$ -invariant quantum circuits can be generated by four-local $SU(d)$ -invariant unitaries for any $d \geq 2$. Using group-theoretical notation, $\mathcal{SV}_4 = \mathcal{SV}_n \subset \text{CQA}$.*

We sketch the proof strategy and leave the details in the Appendix: we define the subgroup generated by CQA, still denoted by CQA for simplicity, by a two-local path-connected Hamiltonian $H_S = \sum_{i=1}^{n-1} (i, i+1)$ mentioned in Sec. III A. Since second-order YJM elements are at most four-local, one can intuitively conclude that $\mathcal{SV}_4 = \mathcal{SV}_n$. Moreover, in contrast to Theorem 1, which addresses the universality restricted to one fixed S^λ , we now consider all inequivalent S^λ from $V^{\otimes n}$ and our method handles only the problem when ignoring phase factors. Thus we claim that $\mathcal{SV}_4 = \mathcal{SV}_n$ and they are all included in CQA because CQA contains generators with nontrivial phases: e.g., $e^{i\theta I}$. We discuss more details about the phase-factor problem by S_n representation theory and show that CQA is a compact subgroup of $\mathcal{V}_4 \subsetneq \mathcal{V}_n$ generally in the Appendix.

IV. CORRESPONDENCE BETWEEN SPIN LABELS AND CONTENT VECTORS

As introduced in Sec. II, Young-basis vectors are labeled by content vectors via YJM elements. A similar phenomenon is also seen for the $SU(d)$ irrep basis vectors constructed by Clebsch-Gordon decompositions [30,47,48] under which the space decomposes as Fig. 2(a): they are labeled by $d - 1$ Casimir operators [73]. We now turn to the question whether these two labeling schemes are equivalent in a certain sense explained in the following. This was conjectured to be true in Ref. [47] and surfaced again in Ref. [51] when the author introduced an efficient quantum Schur transform (QST). An affirmative answer to this conjecture is crucial in this work for three reasons: (a) The Young basis is algebraic. Thus, the gate action drawing from the group algebra $\mathbb{C}[S_n]$ is basis independent. In particular, it can be implemented directly in the computational basis without computing the Fourier coefficients—this is a key observation that underpins the superexponential quantum speedup. (b) This identification allows us to apply both classical tools from $SU(d)$ representation theory as well as the Okounkov-Vershik approach to Schur basis no matter how it is established. As an example, we show in

Sec. IV B an efficient algorithm to generate Schur-basis states required for optimization and learning tasks. (c) A detailed examination on Schur basis enables us convert all the previous results about S_n -CQA to, what we call, $SU(d)$ -CQA with S_n symmetry.

For two-row Young diagrams, this conjecture was shown to be correct in Ref. [74], where the author studied the question by $\frac{1}{2}$ -spin eigenfunctions instead of YJM elements. The general case for $SU(d) - S_n$ duality still holds and can be proven in a surprisingly easy way using YJM elements and the Okounkov-Vershik approach. We present details in the Appendix.

Lemma 3: *Under $SU(d) - S_n$ duality, sequentially coupled Casimir operators commute with YJM elements.*

As a brief illustration of this result, let us consider the *sequentially coupled total spin basis* $|j_1, \dots, j_n; m\rangle$ of $SU(2)$. The *spin component* m and *spin labels* j_k are determined by spin operator S_z^n as the summation of all half Pauli-Z matrices $\frac{1}{2}Z_i$ at each i site and sequential coupled Casimir operators $J_k^2 = (S_x^k)^2 + (S_y^k)^2 + (S_z^k)^2$, respectively, [we abuse our language for simplicity as true eigenvalues of J_k^2 are $j_k(j_k + 1)$]. Since they commute with YJM elements, $J_k^2 X_i |j_1, \dots, j_n, m\rangle = X_i J_k^2 |j_1, \dots, j_n, m\rangle$. It is well known from linear algebra that commutative operators can be simultaneously diagonalized and we elaborate this fact with the following theorem.

Theorem 4: *YJM elements are strictly diagonal under the $SU(d)$ irrep basis built by sequential Clebsch-Gordon decompositions. Conversely, sequentially coupled Casimir operators are strictly diagonal under the Young basis decomposed by branching rule.*

We now illustrate by examples the correspondence between spin labels and content vectors for the simplest $SU(2) - S_n$ duality, then we go to the general case. Let $|j_1, \dots, j_n; m\rangle$ be any $SU(2)$ -irrep basis vector. Theorem 4 says that it is also a Young basis element, thus we can talk about its eigenvalues (content vector) $(\alpha_T(1), \dots, \alpha_T(n))$ with respect to the YJM elements (recall that T denotes a denotes a standard Young tableau, or equivalently the corresponding GZ path from the Bratteli diagram like that from Figs. 3 and 4). An *equivalence* between two labeling schemes means the spin label $J = \{j_1, \dots, j_n\}$ uniquely determines the content vector $\alpha_T = (\alpha_T(1), \dots, \alpha_T(n))$ and vice versa.

SU(2) case: for brevity, let us denote $SU(2)$ -irrep basis vectors by $|J; m\rangle$. It is possible to find two basis elements $|J; m\rangle, |J'; m\rangle$ with the same spin component m but different spin labels. This is due to the fact that $(\mathbb{C}^2)^{\otimes n}$ would decompose into copies of isomorphic $SU(2)$ irreps [Fig. 2(a)]. On the other hand, a Young basis element is then denoted by $|\alpha_T; \mu\rangle$ where μ , as explained in Sec. II,

comes with choosing the permutation module M^μ . It is also possible to find two basis elements $|\alpha_T; \mu\rangle, |\alpha_T; \mu'\rangle$ with the same content vector α_T but from different permutation modules M^μ [Fig. 2(c)]. Let us forget the problem of copies or multiplicities for a while and only focus on the correspondence between J and α_T . We then come back to discuss this in Sec. IV A. Let $|\alpha_T; \mu\rangle$ be a Young-basis element such that α_T equals $(0, -1, 1, 2, 0, 3)$ in Fig. 4. It is also an $SU(2)$ -irrep basis vector. Acted on by J_1^2 , the first spin label is definitely $j_1 = \frac{1}{2}$. To measure the second spin label, let us apply Schur-Weyl duality to the subset system consisting of only the first two qubits. Since $|\alpha_T; \mu\rangle$ is constructed by branching rule, it can be seen as a Young-basis element of S_2 irreps of Young diagram $\lambda = (\lambda_1, \lambda_2) = (1, 1)$ (read off from the first two elements of α_T). Schur-Weyl duality says that $|\alpha_T; \mu\rangle$ should stay in the $SU(2)$ irrep denoted by the same Young diagram, hence $j_2 = \frac{1}{2}(\lambda_1 - \lambda_2) = 0$. Inductively, $j_3 = \frac{1}{2}$ and we obtain $J = (\frac{1}{2}, 0, \frac{1}{2}, 1, \frac{1}{2}, 1)$. As a brief comment on multiplicities, since the total spin (last spin label) is 1, there are three possible choices of z -spin components $m = 1, 0, -1$. Correspondingly, μ from $|\alpha_T; \mu\rangle$ can be three different permutation modules [Fig. 2(c)]. The mechanism for reading off content vectors from spin labels is similar.

$SU(3)$ case: note that the pattern of constructing $SU(2)$ spin labels is simply the familiar branching rule seen in $SU(2)$ irreps [48,75]. We now instantiate with $d = 3$ to show how this pattern generalizes. Let $\mathbf{0}, \mathbf{3}, \bar{\mathbf{3}}, \mathbf{8}$ denote the trivial, the fundamental, the conjugate and the adjoint representations of $SU(3)$. Then we consider the following coupling scheme:

$$\mathbf{3} \otimes \mathbf{3} = \bar{\mathbf{3}} \oplus \mathbf{6} \Rightarrow \bar{\mathbf{3}} \otimes \mathbf{3} = \mathbf{0} \oplus \mathbf{8} \Rightarrow \mathbf{0} \otimes \mathbf{3} = \mathbf{3},$$

where we couple four qudits in which we take the GZ path corresponding to $\alpha_T = (0, -1, -2, 1)$ and ended up with the Young diagram $\lambda = (2, 1, 1)$.

The group $SU(3)$ has two Casimir operators

$$C_1 = \sum_{i=1}^8 T_i^2, \quad C_2 = \sum_{i,j,k} d_{ijk} T_i T_j T_k, \quad (4)$$

where $T_i = \frac{1}{2}\lambda_i$ are half of the Gell-Mann matrices and d_{ijk} are determined by the anticommutation relation $\{T_i, T_j\} = \frac{1}{3}\delta_{ij} + d_{ijk}T_k$. Thus the k th sequential coupling of these operators, denoted by $(C_1, C_2)_k$, corresponds to the YJM element X_k and they are used to record irreps like $\mathbf{3}, \bar{\mathbf{3}}, \mathbf{0}, \mathbf{3}$ appearing in the above example and yield ‘‘spin labels’’ $(1, 0), (0, 1), (0, 0), (1, 0)$, which are highest weights for $SU(3)$ irreps. For a general $SU(d) - S_n$ duality, each YJM element X_k corresponds to a pair of $d - 1$ Casimirs [73] for the k th sequential coupling. It is therefore more concise to use YJM elements in the general case as sequential

coupling and branching rule decomposition are equivalent in describing Schur basis.

A. More facts about state labeling and CQA with S_n symmetry

With spin label-content vector correspondence, we denote a Schur-basis vector by $|\alpha_T; \mu_S\rangle$, where α_T is its content vector. In the S_n picture, α_T tells us exactly the path to restrict an S_n irrep determined by the Young diagram T to S_{n-1} irrep and so forth. However, there are copies of that S_n irrep decomposed from the entire Hilbert space and μ_S labels the multiplicity. One may wish to distinguish these isomorphic copies by permutation modules like Fig. 2(c). However, as mentioned in Sec. II, when $d \geq 3$ isomorphic copies of S_n irreps can even be found from the same permutation module M^μ [50,59] and hence the superscript μ is no longer enough to identify μ_S .

Interestingly, this problem can be solved in the $SU(d)$ picture, in which α_T tells us exactly the path to couple $SU(d)$ irreps sequentially. Our final destination W_λ is uniquely determined by the path, but we now need to label $|\alpha_T; \mu_S\rangle$ as a state in W_λ . When $d = 2$, μ_S is simply taken as the spin- z component m . When $d \geq 3$, we use classical results by Gelfand and Tsetlin [76] and Biedenharn [77]. We illustrate the main idea for $d = 3$: consider weight diagrams of $SU(3)$ irreps on Fig 5. Each dot in a diagram stands for a basis vector of the irrep [49,78]. With α_T being determined by YJM elements, we need only to identify which dot $|\alpha_T; \mu_S\rangle$ corresponds to the weight diagram of W_λ . Diagrammatically, these dots have planar coordinates, which are rigorously called *weights* measured by the isospin I_3 and hypercharge Y operator of $SU(3)$ [49,78] just like measuring spin components by S_z of $SU(2)$. However, being different from the $SU(2)$ case, some weight vectors (dots) occupy the same positions. It is known as the branching rule for $SU(3)$ that dots with the same horizontal coordinate form irreps of $SU(2) \subset SU(3)$. For instance, two brown dots in Fig. 5 form a spin-1/2 irrep while the other four red dots from the same horizontal line form a spin-3/2 irrep. Thus after measuring weights and positions by I_3, Y , we simply need to apply the $SU(2)$ Casimir operator J^2 to discern dots occupying the same position.

In Ref. [77], authors provide the recipe to find these operators for a general $SU(d)$ group. Roughly speaking, we

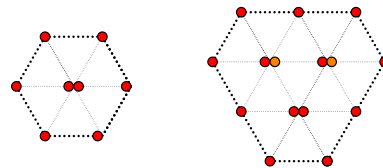


FIG. 5. Typical weight diagrams of $SU(3)$ irreps. Some weight vectors (dots) may occupy the same positions.

first employ $d - 1$ operators μ_i , which span its Cartan subalgebra, to label the weights of a given basis vector. Then we take Casimir operators of $SU(2) \subset \dots \subset SU(d - 1)$ to distinguish dots that occupy the same position. Their eigenvalues, which we record as μ_S , form the so-called *Gelfand-Tsetlin pattern* [76], which is widely used to study $SU(d)$ irreps. To construct a CQA model with S_n symmetry, we replace the YJM elements labeling any S^λ basis states by the following ones labeling any W_λ basis states:

$$\mu_1, \dots, \mu_{d-1}, C_1^{SU(d-1)}, \dots, C_{d-2}^{SU(d-1)}, \dots, C_1^{SU(2)} = J^2.$$

The number of required operators to build a $SU(d)$ -CQA ansatz is $\frac{1}{2}d(d - 1)$ —a constant for fixed d , no matter how many qudits the system contains. However, Casimir operators of $SU(d - 1)$ are supported on $d - 1$ qudits [see Eq. (4) and [78]]. Applying the same proof method from Theorem 1, $SU(d)$ -CQA is thus made of $2(d - 1)$ -local S_n -symmetric unitaries. As a simple example, $SU(2)$ -irrep basis states are uniquely determined by the summation S_z^n of spin operator $\frac{1}{2}Z_i$ on each site i . Without having to employ the Casimir operator, $S_z^n, (S_z^n)^2$ and a problem Hamiltonian H_P already form a $SU(2)$ -CQA model. More precisely, as follows.

Corollary 1: *Let H_P be any path-connected Hamiltonian on the $SU(2)$ -irrep basis, the CQA-ansatz generated by $H_P, S_z^n, (S_z^n)^2$ is dense in each $SU(2)$ -irrep block $U(\dim W_\lambda)$.*

B. State preparation for S_n -CQA ansätze

To investigate evaluation of the matrix elements of S_n Fourier coefficients, we are confined to the Young basis, which requires the implementation of quantum Schur transform [47,51,52]. However, for a wide variety of quantum machine learning and optimization tasks, such as determining the ground-state sign structure of frustrated magnets, it is often advantageous to relax the constraints and ask how easy it is to initialize the states that live in any given S_n irrep. An algorithm to initialize a state in the S_n irrep with Young diagram being $(n/2, n/2)$, is given in Ref. [46]. We generalize this result to an arbitrary S_n irrep in general $SU(d) - S_n$ duality. The key is to utilize different permutation modules and multiplicities of S_n irreps as in Fig. 2(c). Similar to the previous subsection, we construct the algorithm inductively: we first consider the $SU(2) - S_n$ duality in which a (λ_1, λ_2) - S_n irrep is dual to a spin- $(\lambda_1 - \lambda_2)/2$ irrep. Let

$$|\Psi_{\text{init}}\rangle = \underbrace{|0 \dots 0\rangle}_{k \text{ many}} \otimes \underbrace{|s\rangle \otimes \dots \otimes |s\rangle}_{[(n-k)/2] \text{ many}}, \quad (5)$$

where $|s\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$ is one of the Bell states and we assume $n - k$ is even. Then we have the following.

Lemma 4: *Let $\mu = \lambda = [(n + k)/2, (n - k)/2]$. The initialized state $|\Psi_{\text{init}}\rangle$ is contained in S^λ and belongs to the permutation module M^μ .*

Proof. Acting by the spin operator $S^z = \sum_i S_i^z$, it is easy to check that the spin component of $|\Psi_{\text{init}}\rangle$ equals $j = k/2$ hence it belongs to M^μ . By Theorem 4 and discussion from the previous subsection, we have the expansion $|\Psi_{\text{init}}\rangle = \sum_{\alpha_T} c_T |\alpha_T; k/2\rangle$. Since by definition $J_+ |\Psi_{\text{init}}\rangle = 0$, the Young diagram underlying each α_T from the summation must be the same and equals λ . ■

We now illustrate by several examples how to expand $|\Psi_{\text{init}}\rangle$ as a linear combination of Young-basis elements: (a) let $|\Psi_{\text{init}}\rangle = |s\rangle^{\otimes(n/2)}$. This is the state used in Ref. [46]. One can check by YJM elements that it is a single Young-basis element. (b) For a more involved case, consider the $(4, 2)$ irrep of S_6 and write

$$\begin{aligned} |\Psi_{\text{init}}\rangle &= |00\rangle \otimes |s\rangle \otimes |s\rangle \\ &= \frac{2}{3} |\alpha_{T_1}; 1\rangle - \frac{\sqrt{2}}{3} |\alpha_{T_2}; 1\rangle - \frac{\sqrt{2}}{3} |\alpha_{T_3}; 1\rangle \\ &\quad + \frac{1}{3} |\alpha_{T_4}; 1\rangle, \end{aligned}$$

where α_{T_i} corresponds to GZ paths in Fig. 6.

The first two boxes in Fig. 6 correspond to trivial irrep of S_2 acting on the subsystem formed by the first two qubits. Indeed, S_2 acts trivially on $|00\rangle$ from $|\Psi_{\text{init}}\rangle$. As it would be more apparent to see how to get α_{T_i} in the $SU(2)$ picture, we use the spin label-content vector duality and trace the path of spin coupling. As the total spin of $|00\rangle$ and $|00\rangle \otimes |s\rangle$ are identical (Lemma 4), there are two ways to add two more boxes: putting the third box on the rhs of the first two and then putting the fourth on the bottom or conversely. Tensoring again with the singlet $|s\rangle$, we retrieve four branching paths in total. Moreover, by the same reason, it is easy to see that reordering tensor products of $|0 \dots 0\rangle$ and $|s\rangle$ in Eq. (5) yields a different expansion of Young-basis elements for the same S_n irreps. Figure 7 illustrates two more cases: $|s\rangle \otimes |0\rangle \otimes |s\rangle \otimes |0\rangle$ and $|s\rangle \otimes |00\rangle \otimes |s\rangle$.

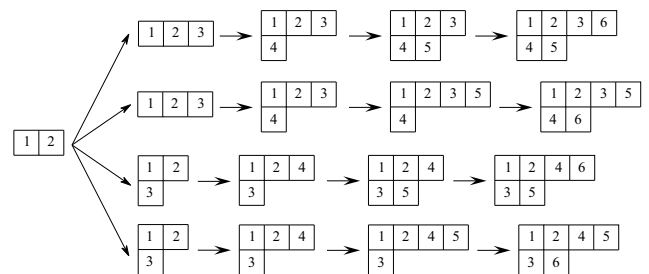


FIG. 6. Decomposing the initial state by content vectors and spin labels.

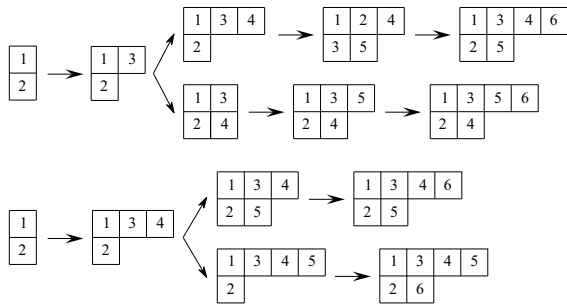


FIG. 7. Reordering tensor products yields different Young-basis expansions.

This method can be generalized to $SU(d) - S_n$ duality. For instance, when $d = 3$ to initialize states for three-row Young diagrams, let us consider the upper, down, strange states u, d, s of 3. Let

$$|\Psi_0\rangle = \underbrace{|u \cdots u\rangle}_{k \text{ many}} \otimes \underbrace{|ud - du\rangle \otimes \cdots \otimes |ud - du\rangle}_{[(n-k)/2] \text{ many}}.$$

This state lies in the $(n+k, n-k, 0)$ irrep. Tensoring with $SU(3)$ singlet $|s\rangle = (uds - usd + dsu - dus + sud - sdu)/\sqrt{6}$, $|\Psi_{\text{init}}\rangle = |\Psi_0\rangle \otimes |s\rangle^{\otimes l}$ is a state from the $(n+k+l, n-k+l, l)$ -irrep. Its expansion can still be tracked by the branching rule as in Figs. 6 and 7. An S_6 -CQA quantum circuit with state initialization described above can be seen in Fig. 8.

C. Quantum superpolynomial speedup

For variational algorithms, typically one would make many measurements with updated parameters $\{\theta_\mu\}$ by

some classical gradient descent scheme:

$$\theta_\mu(t+1) = \theta_\mu(t) - \sum_\nu \eta_\mu(t) A_{\mu\nu}^{-1}(\theta(t)) \frac{\partial}{\partial \theta_\nu} \langle H \rangle_{\theta(t)}, \quad (6)$$

where the learning rate tensor $A_{\mu\nu}(\theta(t))$ is often taken as the identity matrix while $\eta_\mu = \eta$ is the learning rate. The quantity $\langle H \rangle_{\theta(t)}$ is the expectation value to minimize and $(\partial/\partial \theta_\nu) \langle H \rangle_{\theta(t)}$ is the derivative with respect to θ_ν . With an explicitly parameterized unitaries such as in our case, we can utilize the quantum circuits to measure the gradient of the expectation.

Here, we refer taking one measurement at time t as a *query* and *query complexity* as the total time T in order to converge. The query complexity can be analyzed by recent development of the *quantum neural tangent kernel* [79]. Though it would be an interesting case to consider the bound on the query complexity to converge, in this work we focus only on showing that the circuit complexity *per query* can be efficiently simulated on quantum circuits while this is not known in classical regime.

Theorem 5: Let $U_{\text{CQA}}^{(p)}$ denote the CQA ansätze with p alternating layers and let $H \in \mathbb{C}[S_n]$ be a $SU(d)$ -symmetric k -local Hamiltonian with most N terms. Then for any S_n irrep S^λ , the Fourier coefficients:

$$\frac{\partial}{\partial \theta_\mu} \langle \alpha_{T', \mu_S} | U_{\text{CQA}}^{(p)\dagger}(\theta) H U_{\text{CQA}}^{(p)}(\theta) | \alpha_{T, \mu_S} \rangle, \quad (7)$$

where T, T' are standard tableaux of λ (Figs. 3 and 4) and μ_S records the multiplicity of S^λ (Sec. IV A), can be simulated in $O(pN(\theta n^4 + k^2))$ with θ being the largest absolute values of parameters.

The proof is also given in the Appendix. Precisely, we assume that there exists an efficient Schur transform (QST)

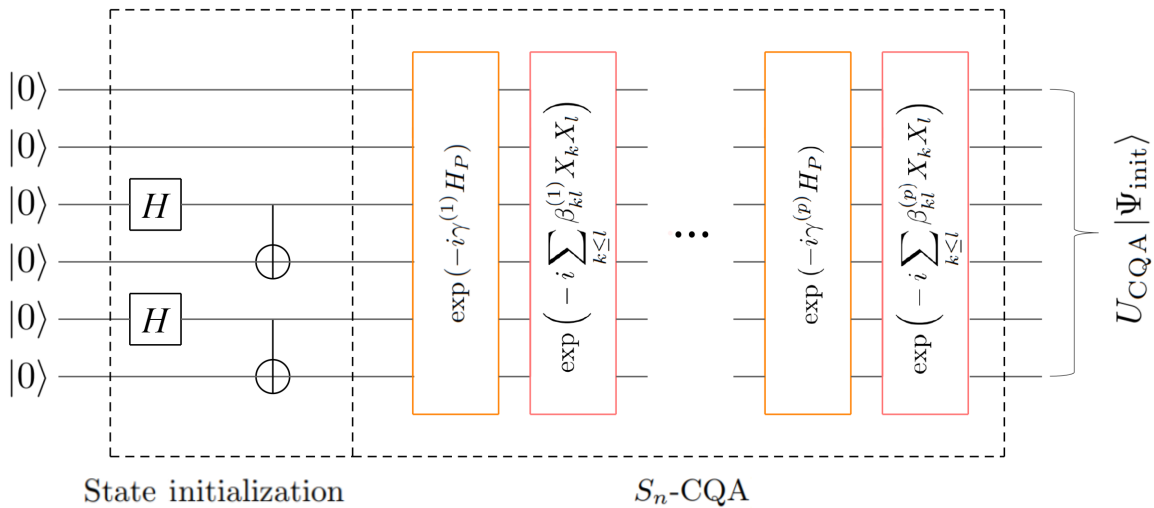


FIG. 8. S_6 -CQA circuit with state initialization in Fig. 6.

[47,51,80] with a polynomial overhead to prepare $|\alpha_T, \mu_S\rangle$. Calculating the Fourier coefficients over S_n is a classically difficult question and the best classical algorithms S_n -FFT requires a factorial complexity [35,36] as S_n has $n!$ group elements and so is the dimension of its regular representation (see the Wedderburn theorem in the Appendix for more details). Therefore, comparing with the complexity of S_n -FFT, there is a superexponential quantum speed per query. However as a caveat, the entire Hilbert space of n qudits only scales exponentially with n and S_n irreps decomposed from the system by Schur-Weyl duality also scales exponentially. Therefore, it would be more reasonable and cautious to refer to a *superpolynomial* quantum speedup for S_n -CQA.

Except comparing with S_n -FFT, recent work from Refs. [10,81] proposes the notion of dequantization to compare the efficiencies of classical and quantum algorithms. Roughly speaking, with well-prepared quantum initial states, quantum algorithms can always be exponentially faster than the best counterpart classical algorithms. Assume classical algorithms also have efficient access to input. If the output can now be evaluated with at most polynomially larger query complexity than the quantum analogy, it is said to be dequantized with no genuine quantum speedup. In our case, let us assume our initial states—Schur-basis elements $|\alpha_T, \mu_S\rangle$ or their linear combinations can be efficiently accessed with classical methods. Even though, dequantization still unlikely happens. Except conducting S_n -FFT, matrix representations of $\sigma \in S_n$ can also be efficiently sampled [48,75], but the method works exclusively for a single group element. To sample $U_{\text{CQA}}^{(p)}(\theta) |\alpha_T, \mu_S\rangle$ processed after S_n -CQA from Eq. (3), the time evolution of CQA Hamiltonians is expanded and approximated by at least superpolynomially many S_n group elements (see the Appendix for more details) and hence is still thought to be classically intractable.

V. $\mathbb{C}[S_n]$ SYMMETRIES OF THE J_1 - J_2 HEISENBERG HAMILTONIAN

The spin-1/2 J_1 - J_2 Heisenberg model is defined by the Hamiltonian:

$$\hat{H}_p = J_1 \sum_{\langle ij \rangle} \hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_j + J_2 \sum_{\langle\langle ij \rangle\rangle} \hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_j, \quad (8)$$

where $\hat{\mathbf{S}}_i = (\hat{S}_i^x, \hat{S}_i^y, \hat{S}_i^z)$ represents the spin operators at site i of the concerned lattice. The symbols $\langle \dots \rangle$ and $\langle\langle \dots \rangle\rangle$ indicate pairs of nearest- and next-nearest-neighbor sites, respectively. The J_1 - J_2 model has been the subject of intense research over its speculated novel spin-liquid phases at frustrated region [82]. The unfrustrated regime ($J_2 = 0$ or $J_1 = 0$) for the antiferromagnetic Heisenberg model is characterized by the bipartite lattices, for which the sign structures of the respective ground states are

analytically given by the *Marshall-Lieb-Mattis theorem* [54–56]. As an important result, ground states of unfrustrated bipartite models are proven to live in the S_n irrep corresponding to the Young diagram $\lambda = (n/2, n/2)$. By Schur-Weyl duality, this subspace is often referred to as the direct sum of $SU(2)$ -invariant subspaces with total spin $J = 0$ in the context of physics [cf. Figs. 2(a) and 2(b)]. With this fact, algorithms like Ref. [38] has been designed to enforce $SU(2)$ symmetry at $J = 0$ and solve Heisenberg models without frustration.

The system is known to be highly frustrated when J_1 and J_2 are comparable $J_2/J_1 \approx 0.5$ [83] and near the region of two phase transitions from Neel ordering to the quantum paramagnetic phase and from quantum paramagnetic to colinear phase, where no exact solution is known. Moreover, little is known about the intermediate quantum paramagnetic phase—recent evidence of deconfined quantum criticality [84,85] sparked further interest in studying these regimes. Gaining physical insights in the intermediate quantum paramagnetic phase requires solving the problem of the ground-state sign structure the system approaches the phase transition. Recently, there were a number of numerical attempts to address the existence of the $U(1)$ gapless spin-liquid phase, using recently the tensor networks [86], restricted Boltzmann machine (RBM) [87], convolutional neural network (CNN) [57,58,88], and graphical neural network (GNN) [89]—all yielding partial progress. As a significant difference from the unfrustrated case, the Marshall-Lieb-Mattis theorem does not hold generally and there is no guarantee that the ground state still lives at $J = 0$ or equivalently $\lambda = (n/2, n/2)$, which urges us to preserve the global $SU(2)$ symmetry, which further gives us access to search in all inequivalent S_n irreps decomposed from the system by Schur-Weyl duality.

Global $SU(2)$ symmetry and challenges in NQS ansätze

Taking advantage of the global $SU(2)$ symmetry, we address this problem in a different way: we recast the Hamiltonian in Eq. (8) by the following identity:

$$\pi(\langle\langle ij \rangle\rangle) = 2\hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_j + \frac{1}{2}I, \quad (9)$$

with $\hat{\mathbf{S}}_i$ being further expanded as the half of standard Pauli operators $\{X, Y, Z\}$. Equation (9) was first discovered by Heisenberg himself [90,91] (an elementary proof can be found in the Appendix) and more recently noted by Ref. [46] in analyzing the ground-state property of the 1D Heisenberg chain. As designed by products of exponentials of SWAPs (eSWAPs), the method proposed in Ref. [46] truly preserves the global $SU(2)$ symmetry. As a brief comparison with S_n -CQA, eSWAP ansätze are universal in relevant sectors given by the $SU(2)$ symmetry. However, this property no longer holds for qudits with

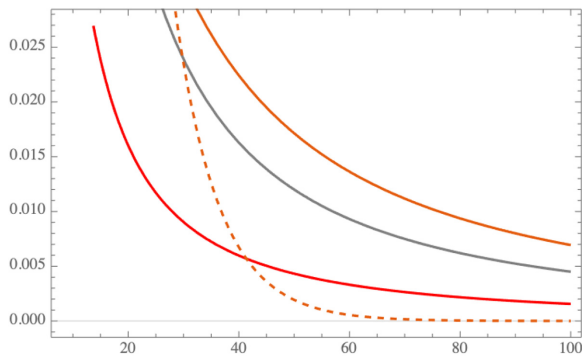


FIG. 9. Scaling properties of small total spin irreps dimension, respectively. The graph shows the scale: $2^n / \dim S^k$ with the partitions $\lambda_1 = (n/2, n/2)$ (red), $\lambda_2 = (n/2 + 1, n/2 - 1)$ (gray), $\lambda_3 = (n/2 + 2, n/2 - 2)$ (orange). The orange dashed line is $\exp(-n/8)$ the exponential decay, since the plot starts at $n = 8$.

$d \geq 3$ (Sec. III C). As eSWAPs are noncommutative operators in general, there are various ways to place them in a quantum circuit. A more suitable perspective to describe the ansatz might be sampling them as two-local $SU(2)$ random circuits [45]. On the other hand, the S_n -CQA ansatz is designed by alternating exponentials of the problem and mixer Hamiltonians H_p, H_M just like the framework of QAOA. Similar to QAOA, S_n -CQA at large p corresponds to a form of adiabatic evolution with global $SU(d)$ symmetry, which could hint a theoretically guaranteed performance as p is large (see Sec. VII in the Appendix).

An immediate consequence of using Eq. (9) is that the resulting Heisenberg Hamiltonian can be expressed in the Young basis where every S_n irrep is indexed by the total spin label j . Mapping to this basis can be done using the constant-depth circuit state initialization in Sec. IV B. Using our $\mathbb{C}[S_n]$ variational ansatz leads to a more efficient algorithm by polynomially reducing the space. In the

NISQ application, especially between 10 to 50 qubits, we have much better scaling, see Fig. 9.

Numerous efforts in applying NQS variational architecture to represent the complicated sign structure in the frustrated regime essentially use the energy as the only criterion for assessing its accuracy. This would result in the optimized low-energy variational states in frustrated regime still obeying the Marshall sign rules even though the true ground state is likely to deviate from it significantly [58], or breaks the $SU(2)$ symmetry [57]. The preservation of spatial symmetry has been the core topic of discussion in the literature, with proposed C_4 equivariant CNN. However, on the two-dimensional model Heisenberg model, the spatial symmetry consideration can only reduce the search space redundancy by a constant factor, thus scaling very poorly at even intermediate n . By reinforcing $SU(2)$ symmetry, we achieve a polynomial reduction of Hilbert space and ensure the result to be physically reasonable, hence offering a second criterion to assess the variational ansätze.

The number of qubits scaling linearly with the number of qubits naturally circumvent the issue of having generalization property, a crucial property for the NQS ansätze to function [92]. In fact, in a related work of us [31], we showed that, making use of the representation theory of the symmetric group, this leads to the superexponential quantum speedup. To this end, it is unlikely that any classically trained ansätze are capable to reinforce the global $SU(2)$ symmetry of the system.

A. Numerical simulation

We provide numerical simulations to showcase the effectiveness of the S_n -CQA ansätze, using a JAX automatic differentiation framework [93]. The implementation of the S_n -CQA ansätze utilizes the classical Fourier-space

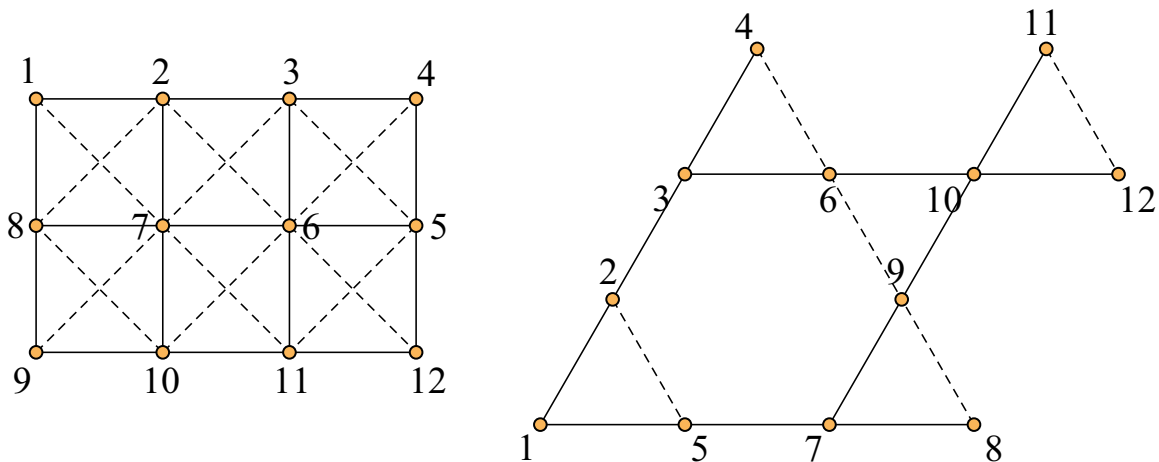


FIG. 10. Square and kagome lattices.

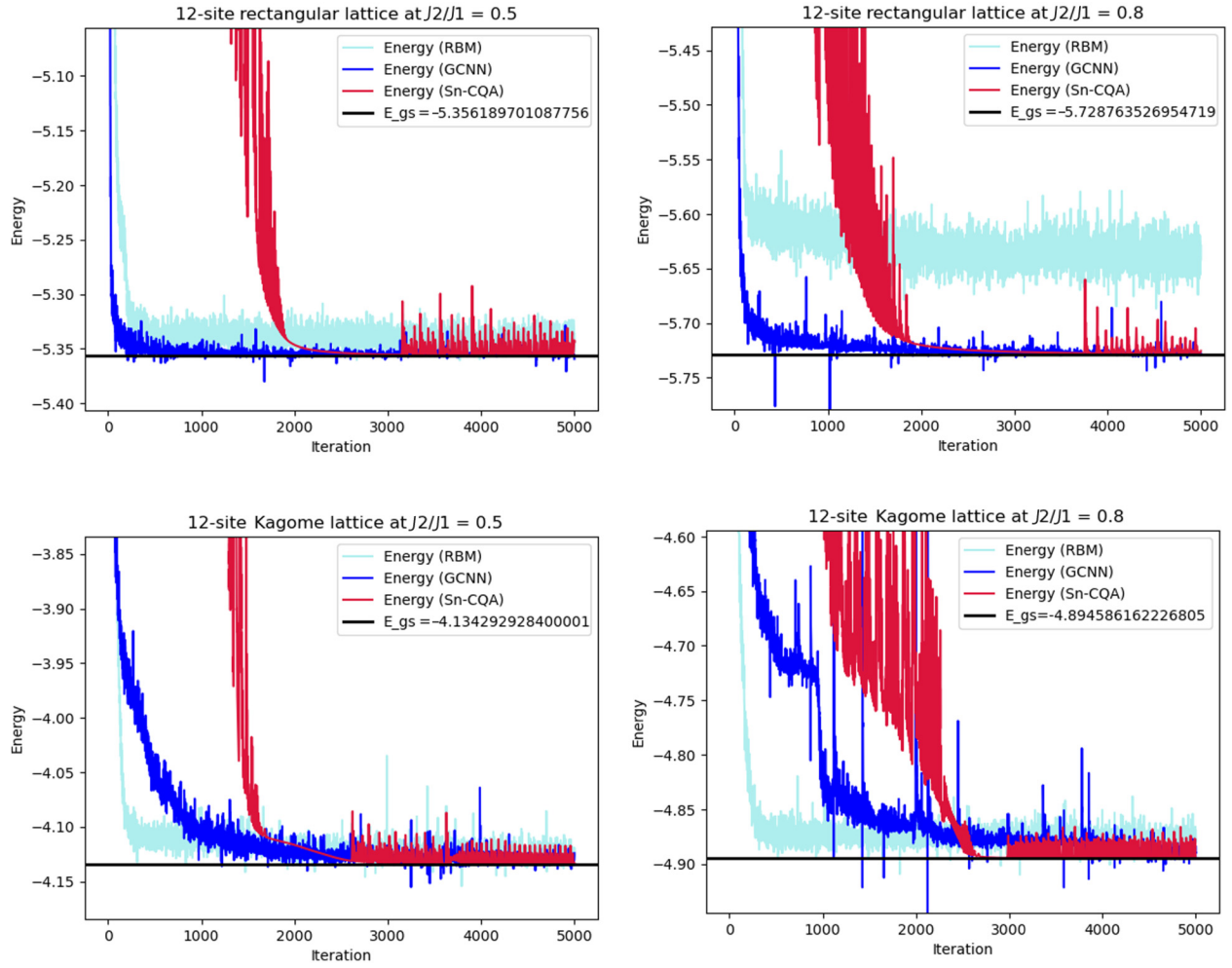


FIG. 11. For the 3×4 lattice, in either case, the S_n -CQA ansätze are able to converge to the ground state at least 10^{-4} precision with explicitly reinforced $SU(2)$ symmetry. This can be seen from the expectation of Sn-CQA never falls below the exact ground state, while nonsymmetry respecting algorithms inevitably do. The numerical results are subject to room for further development, for instance, with better gradient-descent algorithms such as to utilize the Hessian, since we have only 200–500 learnable parameters to optimize. Therefore, we expect the performance and convergence rate of S_n -CQA ansätze to further increase with perhaps more refined tuning.

activation by working in the S_n irreducible representation subspace where the ground-state energy lies. This would impact the stability of the numerical simulations, which imply the best-suited models are with 8–16 spins. This bottleneck in computational resource, as shown in Sec. III, presents no issue for a potential larger-scale implementation on quantum computers. The benchmarked examples with RBM and group-equivariant convolutional neural network (GCNN) [53] are drawn from NetKet [94] tutorial [95], which form the baseline comparison. Note that we implement no explicit global $SU(2)$ or $U(1)$ symmetry for these benchmark algorithms. For numerical simulation of S_n -CQA, we perform random initialization of the parameters. We find that the random initialization already returns the energy, which is within roughly 10^{-2} precision within ED ground-state energy and nonoscillating descents around the ED ground-state energy comparing with that of

GCNN and RBM. This is likely due to the fact that we use the S_n -Fourier space activation with real-valued trial wave functions with explicit $SU(2)$ symmetry. We record the optimized energy for the S_n -CQA ansätze every five iterations, and we set the number of alternating layers $p = 4$ for the 3×4 lattice and $p = 6$ for the 12-spin kagome lattice. In the implementation, we shift the Hamiltonian to $\tilde{H}(\lambda) = H(\lambda) + m 1_{d_\lambda}$ to ensure $\tilde{H}(\lambda)$ is positive semidefiniteness in the S_n irrep specificity by the partition $\lambda = (\lambda_1, \lambda_2)$ with the total spin label $j = (\lambda_1 - \lambda_2)/2$ (Sec. II), where m is the total number of transpositions. We take only the real (normalized) part of the wave function $\text{Re}(\psi) = \psi + \psi^*$. This can be seen as a postprocessing step for the realization of S_n -CQA on a quantum computer. We use the Nesterov-accelerated Adam [96] for the S_n -CQA optimization with hyperparameters: $\text{betas} = [0.99, 0.999]$. We also utilize NetKet’s exact diagonalization (ED) result

for the comparison with the exact ground-state energy. The ground state is additionally calculated in the Young (Schur) basis by diagonalizing the Heisenberg Hamiltonian $H(\lambda)$ in the irrep λ where the ground state lives. It is worth mentioning that our optimized ground state is strictly real valued and has explicitly $SU(2)$ symmetry, offering the missing yet essential physical interpretation. We provide code and Jupiter notebook in open access on Github in python. The numerical simulations are run in the CPU platform with the 9th-Gen 1.4-GHz Intel Core i5 processors.

3 × 4 rectangular lattice

In frustrated region of $J_2/J_1 = 0.5, J_2/J_1 = 0.8$, we find that the ground state lives entirely in the total spin 0 irrep, corresponding to the partition $\lambda = (4, 4)$. In the case of $J_2 = 0.5$, we report that the S_n -CQA ansätze are able to smoothly converge to the ground state, with error to the exact ground-state energy $9.1049e^{-5}$. For $J_2 = 0.8$, the S_n -CQA returns $5.0587e^{-4}$ precision to the ED ground-state energy. We notice that the S_n -CQA seem always to converge to the ground state with reasonable good accuracy without the issue of trapping in local minima, regardless of initialization (random initialization from Gaussian is used). The learning rate used here is 0.01. For the GCNN layers in both J_2 values we set the feature dimensions of hidden layers (8, 8, 8, 8) and 1024 samples with the learning rate set for 0.02. For the RBM model, we fix the learning rate 0.02 with 1024 samples.

12-spin kagome lattice

We find by comparing with the ED result that the ground state of 12-spin kagome lattice lives in the total spin-2 irrep, corresponding to partition [8, 4] in $J_2 = 0$, which suggests it to be fivefold degenerate. For both frustration levels $J_2/J_1 = 0.5$ and $J_2/J_1 = 0.8$, the ground state lives in total spin-0 irrep, which appear to be nondegenerate. We aim to learn the ground state for the 12-spin kagome lattice at $J_2/J_1 = 0.5$ and $J_2/J_1 = 0.8$. In the case of $J_2/J_1 = 0.5$, the optimized ground-state energy by the S_n -CQA ansätze at the end of iteration returns $1.5721e^{-4}$ precision to the ED result. In the case $J_2/J_1 = 0.8$, we have the final optimized energy $6.2065e^{-5}$ precision to the ED ground-state energy. The learning rate is set for 0.01 for the $J_2/J_1 = 0.5$ and 0.8. We set the GCNN in both frustration points of feature dims (8, 8, 8, 8) with 1024 samples. The learning rate in both frustrations is set to be 0.02. The RBM implementation uses 1024 samples with a learning rate of 0.02 for both cases.

VI. DISCUSSION

In this paper, we introduce a framework to design non-Abelian group-equivariant quantum variational ansätze as an example of PQC+ extended from permutational

quantum computation. The restricted universality of the S_n -CQA ansätze makes it applicable to a wide array of practical problems, which would explicitly encode permutation equivariant structure or exhibit global $SU(d)$ symmetry. Our proof techniques can be used to show the universality of QAOA and verify the four locality of generic $SU(d)$ symmetric quantum circuits. Moreover, we illustrate the remarkable efficacy of our approach by finding the ground state of the Heisenberg antiferromagnet J_1 - J_2 spins in a 3×4 rectangular lattice and 12-spin kagome lattice in highly frustrated regimes near the speculated phase transition boundaries. We provide strong numerical evidence that our S_n -CQA can approximate the ground state with a high degree of precision, and strictly respecting $SU(2)$ symmetry. This opens up new avenues for using representation theory and quantum computing in solving quantum many-body problems.

A. Open problems

We conclude with several interesting open problems: (a) We would like to find out the computational power of PQC+. In particular, it is interesting to investigate whether quantum circuits can (in polynomial time) approximate matrix elements of any S_n Fourier coefficients. A natural starting place is perhaps based on the restricted universality of S_n -CQA ansätze in each S_n irrep by asking if a polynomial bounded number of alternating layers p are able to approximate any matrix element of S_n Fourier coefficients. Or we may further lose the condition by asking if a polynomial bounded number of alternating layers p would form an approximate k design for subgroups $U(S^\lambda)$ restricted from $U(V^{\otimes n})$ when imposing the global $SU(d)$ symmetry. A detailed study of this question will shed some light on the nature and scope of the prospective quantum advantage. (b) In the Appendix we show that S_n -CQA ansätze at large p can simulate certain quantum adiabatic evolution with random path-dependent coupling strengths. It would be important to investigate whether the path-dependent coupling strength parameters β_{kl} lead to potential amplitude amplification of the spectral gap in the adiabatic path. In particular, one might need to address the physical dynamics of the random path-dependent coupling strengths. (c) More generally, the quantum speedup we demonstrate here is inherently connected to the PQC+. Are there other quantum speedups within this framework? In particular, (b) suggests a possible route related to quantum annealing. Another possible route may have to do with measurement-based quantum advantage. For instance, see Ref. [97]. Therefore, one might want to ask if our S_n -CQA ansätze have other sources of quantum exponential speedup. (d) Another open direction would be to benchmark the performance of the S_n -CQA ansätze in various Heisenberg models and to implement the S_n -CQA ansätze on a quantum device.

CODE AVAILABILITY

The codes for the numerical simulation can be found at <https://github.com/hanzheng98/Sn-CQA>. The C++ implementation of S_n operations can be found at <https://github.com/risi-kondor/Snob2>. Data availability is upon request by emailing hanz98@uchicago.edu.

ACKNOWLEDGMENT

H.Z. and L.Z. are contributed equally. We thank Alexander Bogatskiy, Giuseppe Carleo, Jiequn Han, Antonio Mezzacapo, Horace Pan, Christopher Roth, Hy Truong Son, Miles Stoudenmire, Kristan Temme, Erik H. Thiede, Chihchan Tien, Zilong Zhang, Wenda Zhou, and Pei Zeng for their useful discussions. We especially thank Liang Jiang and Kanav Setia for suggestions during the preparation of the manuscript. J.L. is supported in part by International Business Machines (IBM) Quantum through the Chicago Quantum Exchange, and the Pritzker School of Molecular Engineering at the University of Chicago through AFOSR MURI (FA9550-21-1-0209). S.S. acknowledges support from the Royal Society University Research Fellowship.

-
- [1] A. W. Harrow, A. Hassidim, and S. Lloyd, Quantum Algorithm for Linear Systems of Equations, *Phys. Rev. Lett.* **103**, 150502 (2009).
 - [2] N. Wiebe, D. Braun, and S. Lloyd, Quantum Algorithm for Data Fitting, *Phys. Rev. Lett.* **109**, 050505 (2012).
 - [3] S. Lloyd, M. Mohseni, and P. Rebentrost, Quantum principal component analysis, *Nat. Phys.* **10**, 631 (2014).
 - [4] P. Wittek, *Quantum Machine Learning: What Quantum Computing Means to Data Mining* (Academic Press, 2014), <https://www.amazon.com/Quantum-Machine-Learning-Computing-Mining/dp/0128100400>.
 - [5] N. Wiebe, A. Kapoor, and K. M. Svore, Quantum deep learning, e-prints [ArXiv:1412.3489](https://arxiv.org/abs/1412.3489) (2014).
 - [6] P. Rebentrost, M. Mohseni, and S. Lloyd, Quantum Support Vector Machine for Big Data Classification, *Phys. Rev. Lett.* **113**, 130503 (2014).
 - [7] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, *Nature* **549**, 195 (2017).
 - [8] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, *Nat. Commun.* **9**, 1 (2018).
 - [9] M. Schuld and N. Killoran, Quantum Machine Learning in Feature Hilbert Spaces, *Phys. Rev. Lett.* **122**, 040504 (2019).
 - [10] E. Tang, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing* (2019), p. 217, <https://dl.acm.org/doi/10.1145/3313276.3316310>.
 - [11] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Supervised learning with quantum-enhanced feature spaces, *Nature* **567**, 209 (2019).
 - [12] Y. Liu, S. Arunachalam, and K. Temme, A rigorous and robust quantum speed-up in supervised machine learning, *Nat. Phys.* **17**, 1013 (2021).
 - [13] J. Liu, F. Tacchino, J. R. Glick, L. Jiang, and A. Mezzacapo, Representation learning via quantum neural tangent kernels, [ArXiv:2111.04225](https://arxiv.org/abs/2111.04225) (2021).
 - [14] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, [ArXiv:1411.4028](https://arxiv.org/abs/1411.4028) (2014).
 - [15] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, The theory of variational hybrid quantum-classical algorithms, *New J. Phys.* **18**, 023023 (2016).
 - [16] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, and L. Cincio, *et al.*, Variational quantum algorithms, *Nat. Rev. Phys.* **3**, 625 (2021).
 - [17] J. Preskill, Quantum computing in the NISQ era and beyond, *Quantum* **2**, 79 (2018).
 - [18] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O’Brien, A variational eigenvalue solver on a photonic quantum processor, *Nat. Commun.* **5**, 1 (2014).
 - [19] S. Hadfield, Z. Wang, B. O’Gorman, E. Rieffel, D. Venturelli, and R. Biswas, From the quantum approximate optimization algorithm to a quantum alternating operator ansatz, *Algorithms* **12**, 34 (2019).
 - [20] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Quantum computational chemistry, *Rev. Mod. Phys.* **92**, 015003 (2020).
 - [21] X. Yuan, J. Sun, J. Liu, Q. Zhao, and Y. Zhou, Quantum Simulation with Hybrid Tensor Networks, *Phys. Rev. Lett.* **127**, 040501(2021).
 - [22] J. Liu, J. Sun, and X. Yuan, Towards a variational Jordan-Lee-Preskill quantum algorithm, [ArXiv:2109.05547](https://arxiv.org/abs/2109.05547) (2021).
 - [23] E. Farhi, J. Goldstone, S. Gutmann, and L. Zhou, The quantum approximate optimization algorithm and the Sherrington-Kirkpatrick model at infinite size, [ArXiv:1910.08187](https://arxiv.org/abs/1910.08187) (2021).
 - [24] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition, *Proc. IEEE* **86**, 2278 (1998).
 - [25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, Imagenet classification with deep convolutional neural networks, *Adv. Neural Inf. Process. Syst.* **25**, 1097 (2012).
 - [26] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, e-prints [ArXiv:1409.1556](https://arxiv.org/abs/1409.1556) (2014).
 - [27] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015), p. 1, <https://ieeexplore.ieee.org/document/7298594>.
 - [28] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, *Nature* **521**, 436 (2015).
 - [29] I. Cong, S. Choi, and M. D. Lukin, Quantum convolutional neural networks, *Nat. Phys.* **15**, 1273 (2019).
 - [30] S. P. Jordan, Permutational quantum computing, [ArXiv:0906.2508](https://arxiv.org/abs/0906.2508) (2009).
 - [31] H. Zheng, Z. Li, J. Liu, S. Strelchuk, and R. Kondor, On the super-exponential quantum speedup of equivariant quantum machine learning algorithms with $SU(d)$ symmetry, e-prints [ArXiv:2207.07250](https://arxiv.org/abs/2207.07250) (2022).

- [32] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Póczos, R. Salakhutdinov, and A. Smola, Deep sets, *ArXiv:1703.06114* (2017).
- [33] H. Maron, H. Ben-Hamu, N. Shamir, and Y. Lipman, in *International Conference on Learning Representations* (2019), <https://openreview.net/forum?id=Syx72jC9tm>.
- [34] E. H. Thiede, T. S. Hy, and R. Kondor, The general theory of permutation equivariant neural networks and higher order graph variational encoders, *ArXiv:2004.03990* (2020).
- [35] M. Clausen and U. Baum, Fast Fourier transforms for symmetric groups: Theory and implementation, *Math. Comput.* **61**, 833 (1993).
- [36] D. K. Maslen, The efficient computation of Fourier transforms on the symmetric group, *Math. Comput.* **67**, 1121 (1998).
- [37] T. Viejira and J. Nys, Many-body quantum states with exact conservation of non-Abelian and lattice symmetries through variational Monte Carlo, *Phys. Rev. B* **104**, 045123 (2021).
- [38] T. Viejira, C. Casert, J. Nys, W. De Neve, J. Haegeman, J. Ryckebusch, and F. Verstraete, Restricted Boltzmann Machines for Quantum States with Non-Abelian or Anyonic Symmetries, *Phys. Rev. Lett.* **124**, 097201 (2020).
- [39] A. Okounkov and A. Vershik, A new approach to representation theory of symmetric groups, *Sel. Math.* **2**, 581 (1996).
- [40] M. Kuranishi, On Everywhere Dense Imbedding of Free Groups in Lie Groups, *Nagoya Math. J.* **2**, 63 (1951).
- [41] H. Yamabe, On an arcwise connected subgroup of a Lie group, *Osaka Mathematical Journal* **2**, 13 (1950).
- [42] S. Lloyd, Quantum approximate optimization is computationally universal, e-prints *ArXiv:1812.11075* (2018).
- [43] M. E. S. Morales, J. D. Biamonte, and Z. Zimborás, On the universality of the quantum approximate optimization algorithm, *Quantum Inf. Process.* **19**, 1 (2020).
- [44] I. Marin, Quotients infinitésimaux du groupe de tresses, *Ann. Inst. Fourier* **53**, 1323 (2003).
- [45] I. Marvian, H. Liu, and A. Hulse, Qudit circuits with $SU(d)$ symmetry: Locality imposes additional conservation laws, e-prints *ArXiv:2105.12877* (2021).
- [46] K. Seki, T. Shirakawa, and S. Yunoki, Symmetry-adapted variational quantum eigensolver, *Phys. Rev. A* **101**, 052340 (2020).
- [47] A. W. Harrow, Applications of coherent classical communication and the Schur transform to quantum information theory, e-prints *ArXiv:quant-ph/0512255* (2005).
- [48] V. Havlíček and S. Strelchuk, Quantum Schur Sampling Circuits can be Strongly Simulated, *Phys. Rev. Lett.* **121**, 060505 (2018).
- [49] R. Goodman and N. R. Wallach, *Symmetry, Representations, and Invariants* (Springer, New York, 2009).
- [50] T. Ceccherini-Silberstein, F. Scarabotti, and F. Tolli, *Representation Theory of the Symmetric Groups* (Cambridge University Press, 2009), <https://www.cambridge.org/us/academic/subjects/mathematics/algebra/representation-theory-and-harmonic-analysis-wreath-products-finite-groups?format=PB&isbn=9781107627857>.
- [51] H. Krovi, An efficient high dimensional quantum Schur transform, *Quantum* **3**, 122 (2019).
- [52] D. Bacon, I. L. Chuang, and A. W. Harrow, Efficient Quantum Circuits for Schur and Clebsch-Gordan Transforms, *Phys. Rev. Lett.* **97**, eid170502 (2006).
- [53] C. Roth and A. H. MacDonald, Group convolutional neural networks improve quantum state accuracy, *ArXiv:2104.05085* (2021).
- [54] W. Marshall and R. E. Peierls, Antiferromagnetism, *Proc. R. Soc. London, A* **232**, 48 (1955).
- [55] E. Lieb and D. Mattis, Ordering energy levels of interacting spin systems, *J. Math. Phys.* **3**, 749 (1962).
- [56] H. Tasaki, *Physics and Mathematics of Quantum Many-Body Systems* (Springer International Publishing, 2020), <https://www.amazon.com/Physics-Mathematics-Quantum-Many-Body-Graduate-ebook/dp/B088D2SQ26>.
- [57] K. Choo, T. Neupert, and G. Carleo, Two-dimensional frustrated J1J2 model studied with neural network quantum states, *Phys. Rev. B* **100**, 125124 (2019).
- [58] A. Szabó and C. Castelnovo, Neural network wave functions and the sign problem, *Phys. Rev. Res.* **2**, 033075 (2020).
- [59] B. E. Sagan, *The Symmetric Group*, 2nd ed., p. Graduate Texts in Mathematics (203, Springer-Verlag, New York, 2001), Vol. xvi+238, Representations, combinatorial algorithms, and symmetric functions.
- [60] A. M. Childs, A. W. Harrow, and P. Wocjan, in *Annual Symposium on Theoretical Aspects of Computer Science* (Springer, 2007), p. 598, https://link.springer.com/chapter/10.1007/978-3-540-70918-3_51.
- [61] S. Keppeler, Birdtracks for $SU(N)$, *SciPost Phys. Lect. Notes* **3** (2018).
- [62] A. Young, *The Collected Papers of Alfred Young 1873–1940*, edited by G. de Beauregard Robinson (University of Toronto Press, 1977), <https://www.amazon.com/Collected-Papers-Alfred-1873-1940-Heritage/dp/1487572727>.
- [63] A.-A. Jucys, Symmetric polynomials and the center of the symmetric group ring, *Rep. Math. Phys.* **5**, 107 (1974).
- [64] G. Murphy, A new construction of Young’s seminormal representation of the symmetric groups, *J. Algebra* **69**, 287 (1981).
- [65] R. Kondor and S. Trivedi, On the generalization of equivariance and convolution in neural networks to the action of compact groups, *ArXiv:1802.03690* (2018).
- [66] R. Kondor, Z. Lin, and S. Trivedi, Clebsch-Gordan nets: A fully Fourier space spherical convolutional neural network, *ArXiv:1806.09231* (2018).
- [67] I. Marin, L’algèbre de Lie des transpositions, *J. Algebra* **310**, 742 (2007).
- [68] J. Hilgert and K.-H. Neeb, *Structure and Geometry of Lie Groups*, Springer Monographs in Mathematics (Springer, New York, 2012). oCLC: ocn757479578
- [69] M. Goto, On an arcwise connected subgroup of a Lie group, *Proc. Am. Math. Soc.* **20**, 157 (1969).
- [70] A. Y. Vlasov, Clifford algebras and universal sets of quantum gates, *Phys. Rev. A* **63**, 054302 (2001).
- [71] I. Marvian, Restrictions on realizable unitary operations imposed by symmetry and locality, *Nat. Phys.* **18**, 283 (2022).

- [72] I. Marvian, H. Liu, and A. Hulse, Rotationally-invariant circuits: Universality with the exchange interaction and two ancilla qubits, e-prints [ArXiv:2202.01963](https://arxiv.org/abs/2202.01963) (2022).
- [73] L. C. Biedenharn, On the representations of the semisimple lie groups. I. The explicit construction of invariants for the unimodular unitary group in N dimensions, *J. Math. Phys.* **4**, 436 (1963).
- [74] R. Pauncz, *The Symmetric Group in Quantum Chemistry* (CRC Press, 2018), <https://www.routledge.com/The-Symmetric-Group-in-Quantum-Chemistry/Pauncz/p/book/9781315898124>.
- [75] V. Havlíček, S. Strelchuk, and K. Temme, Classical algorithm for quantum $SU(2)$ Schur sampling, *Phys. Rev. A* **99**, 062336 (2019).
- [76] I. M. Gelfand, *Collected Papers. II*, reprint of the 1988 ed., reprint 2015 ed., Springer Collected Works in Mathematics (Springer-Verlag, Berlin Heidelberg, 2015).
- [77] G. E. Baird and L. C. Biedenharn, On the representations of the semisimple Lie groups. II, *J. Math. Phys.* **4**, 1449 (1963).
- [78] W. Greiner and B. Müller, *Quantum Mechanics: Symmetries* (Springer-Verlag, Berlin; New York, 1994).
- [79] J. Liu, K. Najafi, K. Sharma, F. Tacchino, L. Jiang, and A. Mezzacapo, An analytic theory for the dynamics of wide quantum neural networks, Preprint [ArXiv:2203.16711](https://arxiv.org/abs/2203.16711) (2022).
- [80] W. M. Kirby and F. W. Strauch, A practical quantum algorithm for the Schur transform, [ArXiv:1709.07119](https://arxiv.org/abs/1709.07119) (2018).
- [81] E. Tang, Quantum Principal Component Analysis Only Achieves an Exponential Speedup Because of its State Preparation Assumptions, *Phys. Rev. Lett.* **127**, 060503 (2021).
- [82] L. Balents, Spin liquids in frustrated magnets, *Nature* **464**, 199 (2010).
- [83] M. Bukov, M. Schmitt, and M. Dupont, Learning the ground state of a non-stoquastic quantum Hamiltonian in a rugged neural network landscape, *SciPost Phys.* **10**, 147 (2021).
- [84] A. Nahum, J. Chalker, P. Serna, M. Ortuño, and A. Somoza, Deconfined Quantum Criticality, Scaling Violations, and Classical Loop Models, *Phys. Rev. X* **5**, 041048 (2015).
- [85] L. Wang, Z.-C. Gu, F. Verstraete, and X.-G. Wen, Tensor-product state approach to spin-1/2 square J1J2 antiferromagnetic Heisenberg model: Evidence for deconfined quantum criticality, *Phys. Rev. B* **94**, 075143 (2016).
- [86] W.-Y. Liu, S.-S. Gong, Y.-B. Li, D. Poilblanc, W.-Q. Chen, and Z.-C. Gu, Gapless quantum spin liquid and global phase diagram of the spin-1/2 J_1 - J_2 square antiferromagnetic heisenberg model, [ArXiv:2009.01821](https://arxiv.org/abs/2009.01821) (2021).
- [87] Y. Nomura and M. Imada, Dirac-Type Nodal Spin Liquid Revealed by Refined Quantum Many-Body Solver Using Neural-Network Wave Function, Correlation Ratio, and Level Spectroscopy, *Phys. Rev. X* **11**, 031034 (2021).
- [88] X. Liang, W.-Y. Liu, P.-Z. Lin, G.-C. Guo, Y.-S. Zhang, and L. He, Solving frustrated quantum many-particle models with convolutional neural networks, *Phys. Rev. B* **98**, 104426 (2018).
- [89] D. Kochkov, T. Pfaff, A. Sanchez-Gonzalez, P. Battaglia, and B. K. Clark, Learning ground states of quantum hamiltonians with graph networks, [ArXiv:2110.06390](https://arxiv.org/abs/2110.06390) (2021).
- [90] W. Heisenberg, Zur Theorie des Ferromagnetismus, *Z. Phys.* **49**, 619 (1928).
- [91] D. J. Klein and W. A. Seitz, Symmetric-group algebraic variational solutions for Heisenberg models at finite temperature, *Int. J. Quantum Chem.* **41**, 43 (1992).
- [92] T. Westerhout, N. Astrakhantsev, K. S. Tikhonov, M. I. Katsnelson, and A. A. Bagrov, Generalization properties of neural network approximations to frustrated magnet ground states, *Nat. Commun.* **11**, 1593 (2020).
- [93] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, JAX: Composable transformations of Python+NumPy programs, [http://github.com/google/jax](https://github.com/google/jax) (2018).
- [94] G. Carleo, K. Choo, D. Hofmann, J. E. Smith, T. Westerhout, F. Alet, E. J. Davis, S. Efthymiou, I. Glasser, and S.-H. Lin, *et al.*, NetKet: A machine learning toolkit for many-body quantum systems, *SoftwareX* **10**, 100311 (2019).
- [95] NetKet Tutorial <https://www.netket.org/tutorials.html>.
- [96] T. Dozat, Workshop track—iclr 2016—openreview <https://openreview.net/pdf/OM0jvwB8jlp57ZJjtNEZ.pdf>.
- [97] H.-Y. Huang, R. Kueng, and J. Preskill, Information-Theoretic Bounds on Quantum Advantage in Machine Learning, *Phys. Rev. Lett.* **126**, 190505 (2021).