# Almost gauge-invariant states and the ground state of Yang-Mills theory

A. Mariani[ORCID]

*Albert Einstein Center for Fundamental Physics, Institute for Theoretical Physics,*
*University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland*

We consider the problem of the explicit description of the gauge-invariant subspace of pure lattice gauge theories in the Hamiltonian formulation, where the gauge group is either a compact Lie group or a finite group. The latter case is particularly interesting for quantum simulation. A basis of states where configurations are grouped according to their holonomies is shown to have several advantages over other descriptions. Using this basis, we compute some properties of interest for some non-Abelian finite groups on small lattices, and in particular we examine the question of whether a certain ansatz introduced long ago is a good approximation for the ground state.

## I. INTRODUCTION

The Hamiltonian formulation of lattice gauge theories [1] has attracted renewed interest in recent years due to its relevance in the field of quantum simulation, i.e., the possibility of performing numerical calculations on quantum devices. On the theoretical side, quantum simulation requires a formulation of gauge theories with a finite-dimensional Hilbert space. While many proposals have been made, all of which may in principle work, here we choose to focus on the approach where one replaces the gauge group, typically a compact Lie group, with one of its finite subgroups [2–5]. The general strategy was outlined in [2]: the continuum limit is approached by tuning the theory to a point in the phase diagram where the correlation length is large (i.e., a phase transition). Despite the fact that finite group theories only have a first-order transition (i.e., the correlation length becomes large but not infinite), by systematic improvement of the lattice action one may nonetheless enter the scaling regime. This has been demonstrated for the group $S(1080)$, which is the largest crystal-like subgroup of SU(3) and is therefore relevant for QCD [6]. We also note that gauge theories with a finite gauge group are also relevant in condensed matter [7–11] and quantum gravity [12,13].

A peculiarity of gauge theories is that, due to the Gauss law, only a small subset of the states in their Hilbert space are physical. It is thus natural to look for a formulation purely in terms of physical degrees of freedom. Generally speaking, working in the physical subspace provides a memory advantage, in that one needs to consider fewer states; but the Hamiltonian acquires some degree of non-locality which makes operations more expensive. A further advantage of a gauge-invariant formulation is that one does not need to correct errors arising from violations of the Gauss law, which reduces memory and operational requirements. Whether this trade-off is worth making is application- and platform-dependent. For finite groups, an explicit description of the physical subspace in terms of spin networks was recently considered in [14].

In this work we focus on pure gauge theories with finite gauge group, but many of our results are also relevant for compact Lie groups. In Sec. II, we review the Hamiltonian formulation of gauge theories. In Sec. III, based on a formulation introduced in the early days of lattice gauge theory [15] to describe the gauge invariant content of classical observables, we consider a set of "holonomy states" which solve the Gauss law with the exception of one global constraint which is treated separately. We discuss how these states provide several advantages over the spin network formulation. Working in this basis may be useful for Hamiltonian simulations on current classical hardware as well as on current quantum devices with few qubits. We also review the issue of whether the Wilson loop observables are sufficient to describe the gauge-invariant states. Further, the issues of topology and center symmetry in this basis are discussed. Finally, in Sec. IV we perform some numerical simulations to demonstrate an application of the holonomy basis. In particular, we discuss whether a conjectured ansatz for the ground state wave function of Yang-Mills theory proposed in [16–19] provides a good approximation to the exact ground state for some non-Abelian finite groups.

## II. HAMILTONIAN FORMULATION

In this section, we briefly review the Hamiltonian formulation of pure lattice gauge theories, where the gauge

group is either a compact Lie group or a finite group. We also use this section to set the notation for the rest of this work.

In the lattice Hamiltonian formulation [1], time is continuous while space is given by a finite set of points, most commonly a (periodic) cubic or hypercubic lattice. In this work, however, we consider more generally the situation where space is discretized into an arbitrary finite graph (with vertices and links) which includes as a special case any type of lattice discretization. In fact, as we will see shortly, in order to define the Hilbert space and the Gauss law (which are the focus of this work) a graph structure is sufficient. Note that this excludes pathological cases such as dangling links connected to only one site. We will always assume that the graph discretizing space is connected. On the other hand, to define the magnetic Hamiltonian one requires the notion of a minimal cycle (i.e., the plaquette) for which a graph structure is not sufficient; in general, one requires the notion of a cell complex. We will not consider this here; whenever discussion of the Hamiltonian is required, we always specialize to a hypercubic lattice.

The Hilbert space is most easily described in the "group element basis". Let $G$ be the gauge group, which is either a compact Lie group or a finite group. A classical configuration is given by an assignment of a group element to each link. In the quantum theory, states can also be found in a superposition of various group elements. Therefore we introduce a set of basis states $|g\rangle$ on each link indexed by group elements $g \in G$. These are orthonormal, i.e., $\langle g|h\rangle = \delta_{gh}$. The Hilbert space on each link is given by the group algebra $\mathbb{C}[G]$, i.e., the vector space spanned by the $|g\rangle$. For a compact Lie group, the group algebra $\mathbb{C}[G]$ should be replaced with the space $L^2(G)$ of square-integrable functions on the group; apart from convergence issues, its description is the same as the group algebra. Then the Hilbert space on the whole discretized space is given by a copy of $\mathbb{C}[G]$ on each link, i.e.,

$$\mathcal{H}_{\text{tot}} = \bigotimes_{\text{links}} \mathbb{C}[G]. \tag{1}$$

We call this the "total" Hilbert space because we have not yet imposed the Gauss law. Since on each link we have $|G|$ basis states, $\dim(\mathbb{C}[G]) = |G|$, and therefore $\dim \mathcal{H}_{\text{tot}} = |G|^L$, where $L$ is the number of links. Basis states in the total Hilbert space are denoted as

$$|\{g\}\rangle \equiv \bigotimes_{l \in \text{links}} |g_l\rangle, \tag{2}$$

where $g_l \in G$ is an assignment of a group element to each link. In other words, the basis states in $\mathcal{H}_{\text{tot}}$ are indexed by classical configurations. Note that the graph discretizing space is assumed to have an (arbitrary) fiducial orientation. In particular, links can be traversed in either the positive or

negative orientation, but if a link $l$ is traversed in the negative orientation then the group element $g_l$ is assumed to be replaced by its inverse $g_l^{-1}$.

To perform a gauge transformation, one first assigns a group element $g_x$ to each site $x$. Given such an assignment, which we denote $\{g_x\}$, the corresponding operator $\mathcal{G}(\{g_x\})$ which performs the gauge transformation acts on the basis states as

$$\mathcal{G}(\{g_x\})|\{g\}\rangle = \bigotimes_{l=\langle xy\rangle \in \text{links}} |g_x g_l g_y^{-1}\rangle, \tag{3}$$

where $l = \langle xy\rangle$ is the link which connects sites $x$ and $y$ (in the positive orientation). In other words, on each link one performs the transformation $|g_l\rangle \to |g_x g_l g_y^{-1}\rangle$.

It is useful to introduce the operators $L_g$ and $R_g$ which perform left- and right-translations on the group element basis states on each link. These are defined as

$$L_g|h\rangle = |gh\rangle, \qquad R_g|h\rangle = |hg^{-1}\rangle. \tag{4}$$

Both $L$ and $R$ are faithful unitary representations of the gauge group $G$, known as the left- and right-regular representations [20,21]. In terms of these operators, the gauge transformation takes the form,

$$\mathcal{G}(\{g_x\}) = \bigotimes_{l=\langle xy\rangle \in \text{links}} L_{g_x} R_{g_y}, \tag{5}$$

where it is understood that the $L$ and $R$ operators act on the Hilbert space of the corresponding link in each tensor product factor. Then one considers as belonging to the physical, gauge-invariant Hilbert space $\mathcal{H}_{\text{phys}}$ only those states $|\psi\rangle$ which are gauge-invariant [1,22,23], i.e.,

$$\mathcal{G}(\{g_x\})|\psi\rangle = |\psi\rangle, \tag{6}$$

for any possible assignment $g_x$ of group elements to sites.

While our discussion of the Hilbert space is valid for both compact Lie groups and finite groups, in the rest of this work, we will only consider the Hamiltonian in the special case where the gauge group is a finite group and space is discretized as a hypercubic lattice. In this case, the Hamiltonian is given by [1,3,12,14,24,25]

$$H = \lambda_E \sum_{l \in \text{links}} h_E + \lambda_B \sum_{\square} h_B(g_\square). \tag{7}$$

Here $\square$ are the plaquettes and $g_\square = g_1 g_2 g_3^{-1} g_4^{-1}$ is the associated plaquette variable, i.e., the oriented product of the four links around the plaquette. For the magnetic Hamiltonian, to match the notation of [14] we choose

$$h_B(g_\square) = -2\mathrm{Re}\chi(g_\square), \tag{8}$$

where $\chi$ is the character of a faithful representation of the gauge group $G$.

On the other hand, the electric Hamiltonian acts separately on each link, where it is given by

$$h_E = \sum_{k \in \Gamma} (1 - L_k), \qquad (9)$$

where $L_k$ is the left translation operator defined in Eq. (4) and $\Gamma \subset G$ is a subset which does not contain the identity and is invariant under group conjugation and group inversion, i.e., $g\Gamma g^{-1} = \Gamma$ and $\Gamma^{-1} = \Gamma$. Invariance under conjugation ensures that $h_E$ is gauge-invariant. Because of these two requirements, it makes no difference whether one chooses $L_k$ or $R_k$ in Eq. (9). The electric Hamiltonian on each link Eq. (9) is nothing but a Laplacian on the finite group [14]. In relativistic theories, the electric and magnetic Hamiltonian are related, and one must have [12]

$$\Gamma = \{g \in G, g \neq 1, \max[\mathrm{Re}\chi(g)]\}; \qquad (10)$$

i.e., $\Gamma$ is the subset of non-identity elements of $G$ which maximise the real part $\mathrm{Re}\chi$ of the character of the magnetic Hamiltonian. Note that the set $\Gamma$ in Eq. (10) is invariant under both conjugation and inversion, as mentioned in the previous paragraph. As has been pointed out recently [14], the electric Hamiltonian on each link may be degenerate, but in this work we do not consider situations where this occurs. If this does not occur, then the electric Hamiltonian has a unique ground state on each link, given by an equal superposition over all group element states.

Finally, in Sec. IV we will numerically construct the Hamiltonian $H$ to compute some quantities of interest. We choose to measure the energy in units of $\lambda_E + \lambda_B$, so that in practice the Hamiltonian may be written as

$$H = (1 - \lambda) \sum_{l \in \mathrm{links}} h_E + \lambda \sum_{\square} h_B(g_\square), \qquad (11)$$

where $\lambda \in [0, 1]$ is a real parameter. The "strong coupling phase" where the electric Hamiltonian dominates then occurs for small $\lambda$, while for large $\lambda$ the physics is dominated by the magnetic Hamiltonian.

## III. ALMOST GAUGE-INVARIANT STATES

As we have seen in the previous section, only a small subset of the states in the total Hilbert space are physical. In some situations, it may therefore be advantageous to work directly in the physical subspace. This reduces memory requirements, but increases the complexity of operations because the Hamiltonian generally becomes somewhat nonlocal. Another advantage of working in a physical basis is that one does not need to correct errors which cause violations of the Gauss law, thus contributing to reduced memory and operational requirements. Whether

the advantages outweigh the disadvantages is application- and platform-dependent.

Several approaches have been devised to describe the physical subspace of lattice gauge theories. Recently, the physical subspace of pure gauge theories whose gauge group is a finite group has been described in terms of spin-network states [14,26,27]. This description has several advantages, but it can be cumbersome due to the need for Clebsch-Gordan coefficients as well as the formulation in the representation basis, which together make access to operators diagonal in the group element basis quite difficult.

Since the early days of lattice gauge theory, gauge-fixing has been a popular method to reduce the number of degrees of freedom of lattice gauge theories [28]. In this section, we will make extensive use of results in [15], where the gauge-invariant content of gauge-field configurations is described in terms of lattice holonomies based on a spanning rooted tree. This description allows the introduction of a particularly simple basis of states, which we call "holonomy states". They are *almost* gauge invariant, in the sense that they satisfy all Gauss law constraints except for one overall global transformation. Importantly, these states can be constructed on arbitrary graphs, which includes any lattice discretization, and in particular hypercubic lattices with both open and periodic boundaries. We then discuss several properties of these states, including how to compute matrix elements of the Hamiltonian in this basis and their relation to a basis of Wilson loops. Interestingly, as shown in Sec. III D, the electric Hamiltonian in this basis involves only $\mathcal{O}(|\Gamma|)$ terms (i.e., independent of $V$), against a naive expectation of $\mathcal{O}(V)$ terms where $V$ is the volume. Using this formulation, we also rederive a formula for the dimension of the physical subspace first obtained in [14].

Here and throughout we use notation appropriate for finite groups. But the results of this section remain valid also for compact Lie groups, with the sole caveat that one should replace the group sums $\frac{1}{|G|} \sum_{g \in G}$ with the (normalized) Haar measure $\int dg$.

The simplest way of constructing gauge-invariant states would be to start from a generic state $|\psi\rangle$ and project it onto the gauge-invariant sector. The appropriate projector can be constructed by averaging over all gauge transformations,

$$P = \frac{1}{|G|^V} \sum_{\{g_x\}} \mathcal{G}(\{g_x\}), \qquad (12)$$

where $V$ is the number of sites. However, this method is neither explicit nor efficient.

Alternatively, by repeatedly applying gauge transformations, the gauge field configurations split into gauge equivalence classes $X_1, X_2, \ldots$. One could then construct gauge equivalence states,

$$|X_i\rangle \equiv \frac{1}{\sqrt{|X_i|}} \sum_{\{g\} \in X_i} |\{g\}\rangle, \qquad (13)$$

by equally superimposing all configurations in one gauge equivalence class. The states $|X_i\rangle$ are gauge-invariant by construction and form in fact an orthonormal basis of gauge-invariant states. One could then work directly in the physical subspace spanned by the $|X_i\rangle$. The problem here is that the gauge equivalence classes *do not* all have the same size. This can be seen already in the simple examples presented in Sec. III A. Therefore in order to work with the states $|X_i\rangle$, one would need to know the sizes $|X_i|$ of the gauge equivalence classes, but it is not clear how to compute them.

In the next section, we give two simple examples to help gain intuition for the general construction of Sec. III B.

### A. A preliminary example

In this section, we consider two simple examples—a $2 \times 2$ and a $2 \times 3$ square lattice with open boundaries—in order to gain intuition for the general construction which will be presented in Sec. III B. We chose square lattices with open boundaries because they are the simplest example, although, as already remarked, our construction works with arbitrary graph discretizations. Several of the issues that we will encounter in the general case are already present in these simple cases.

Consider first the $2 \times 2$ lattice, shown in Fig. 1. This lattice has $V = 4$ sites and $L = 4$ links. A state in the group element basis is then of the form $|g_1, g_2, g_3, g_4\rangle$ where $g_i$ is the group element assigned to link $i$. In order to reduce the number of variables, we want to gauge-fix as many links as possible to the identity. Thus we choose a maximal tree, shown in Fig. 1 by the bold links, and ideally we would like to set all links in the tree to the identity. This leaves out exactly one link (in this case, the one labeled 3), and therefore the gauge-invariant content can be described by simply one variable.

This is how the story would go in the path-integral approach. However, the situation is slightly different in the Hamiltonian formulation. In fact, the "gauge-fixed" state in this case takes the form $|1, 1, g_3, 1\rangle$, and it can be easily checked that it is *not* invariant under gauge transformations of the usual form Eq. (5). Instead, we construct (in this case, the only) untraced Wilson loop in this lattice, which is $h \equiv g_4 g_3 g_2^{-1} g_1^{-1}$, i.e., the plaquette. Note that we have chosen the orientation of the path defining $h$ so that it
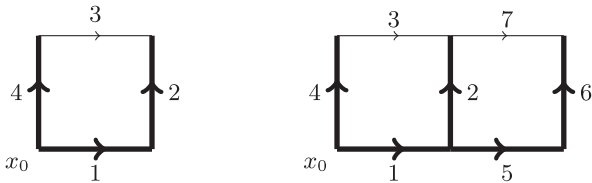
traverses link 3 in its positive orientation. Link 3 is precisely the one which we *do not* want to fix to the identity. Also, since the Wilson loop is untraced, the base point (marked $x_0$ in Fig. 1) is important. As we have seen above, the gauge-invariant content on this lattice can be described by a single variable, which is the holonomy $h$. Therefore we define a state $|h\rangle$ associated to $h$ by equally superimposing all configurations with the same $h$,

$$|h\rangle \equiv \frac{1}{\sqrt{|G|}^3} \sum_{\substack{g_1, g_2, g_3, g_4 \in G \\ g_4 g_3 g_2^{-1} g_1^{-1} = h}} |g_1, g_2, g_3, g_4\rangle. \qquad (14)$$

In his case, it is possible to check explicitly that the states $|h\rangle$ for $h \in G$ are orthonormal; in particular the normalization factor is *the same* for all $h$. It is also not hard to check that the state $|h\rangle$ is invariant under all gauge-transformations except those at the base point $x_0$. This is not surprising, since $h$ is an holonomy with base point $x_0$. Therefore, we can think of $h$ as a "more" gauge-invariant version of the simple link variable $g_3$ which is the only one which we do not gauge-fix. In fact, on the gauge-fixed configuration, $h = g_3$.

But, as we said, the state $|h\rangle$ is not fully gauge-invariant. There is one remaining gauge transformation, the one at the base point $x_0$, under which $|h\rangle$ is *not* invariant. This acts simply by conjugation, i.e., $|h\rangle \rightarrow |g_{x_0} h g_{x_0}^{-1}\rangle$. This means that it simply groups together the states $|h\rangle$ in the same conjugacy class. Therefore the gauge-invariant states on this lattice are labelled by the conjugacy classes $C$ of the gauge group $G$, and they are of the form,

$$|C\rangle \equiv \frac{1}{\sqrt{|C|}} \sum_{h \in C} |h\rangle. \qquad (15)$$

These are the gauge-equivalence states of Eq. (13). The states $|C\rangle$ are again orthonormal, but their normalization is different for each state. In this case, it is simply given by the size of the conjugacy classes. As we will see in the other example presented in this section, computing the equivalent normalizations is in general not easy. In particular, as anticipated in the introduction to Sec. III, this shows that even in this simple example the gauge-equivalence classes do not all have the same size, as the conjugacy classes of a non-Abelian group generally have different sizes. The result remains true also for compact Lie groups. In that case one cannot simply count the elements (which would be infinity); the appropriate notion which generalizes counting is volume under the Haar measure. Then the fact that different conjugacy classes have different sizes is made precise by the Weyl integration formula [21].

Alternatively, the gauge-invariant states on this lattice can be described in terms of wavefunctions of the holonomy as



FIG. 1. A $2 \times 2$ (*left*) and $2 \times 3$ (*right*) square lattice with open boundaries. The arrows depict the standard orientation of the links. In each case, the bold links form a possible maximal tree.

$$|\psi\rangle = \sum_{h \in G} \psi(h)|h\rangle, \qquad (16)$$

where the wave function satisfies $\psi(ghg^{-1}) = \psi(h)$. Such functions are known as "class functions" and by standard theorems [20,21] they can be expanded in a basis of irreducible characters,

$$\psi(h) = \sum_i \psi_i \chi_i(h), \qquad (17)$$

where the $\psi_i$ are arbitrary complex coefficients and the $\chi_i$ are the irreducible characters of $G$. These are simply traces of irreps of $G$, i.e., for this lattice with only one independent holonomy, the $\chi_i(h)$ are nothing but all the traced Wilson loops in all irreducible representations of $G$. Note that it is generally true that the number of conjugacy classes of a finite group equals the number of its irreps [20], so that the two descriptions of the gauge-invariant states are consistent. While in this case it is possible to express the gauge-invariant states as functions of traced Wilson loops, this is not true in general. We will explore this issue further in the rest of this section as well as in Sec. III C.

We now turn to the $2 \times 3$ square lattice shown on the right side of Fig. 1. In this case we have $V = 6$ sites and $L = 7$ links. A maximal tree like the one shown in Fig. 1 includes 5 links. Therefore after fully "gauge-fixing", we are left with two group-valued variables related to the links labeled 3 and 7 in Fig. 1. But again we cannot simply gauge-fix elements to the identity, as the resulting state would not be gauge-invariant. Instead we define appropriate holonomies. In this case we have two

$$h_1 \equiv g_4 g_3 g_2^{-1} g_1^{-1}, \qquad h_2 \equiv g_1 g_2 g_7 g_6^{-1} g_5^{-1} g_1^{-1}. \qquad (18)$$

Importantly, we chose the two holonomies to have the *same base point* $x_0$. Each holonomy goes through precisely one link which is not in the maximal tree; $h_1$ is associated with link 3 and $h_2$ with link 7. Now we define holonomy states,

$$|h_1, h_2\rangle \equiv \frac{1}{\sqrt{|G|}^5} \sum_{\{g\} \in [h_1, h_2]} |\{g\}\rangle, \qquad (19)$$

by equally superimposing all configurations with the same holonomies. Again it is tedious but not hard to check that the states $|h_1, h_2\rangle$ are orthonormal. In particular, they again all have the same normalization, which is the same as saying that all classes of states with the same holonomies contain an equal number of states. We will show that this statement is true in general in Sec. III B.

The holonomy states $|h_1, h_2\rangle$ are invariant by construction under all gauge transformations except the one at the base point $x_0$. Under this gauge transformation, they transform by simultaneous conjugation, i.e., as $|h_1, h_2\rangle \rightarrow |g_{x_0} h_1 g_{x_0}^{-1}, g_{x_0} h_2 g_{x_0}^{-1}\rangle$. Unlike the previous case with only

one holonomy, this condition admits no simple mathematical characterization.[1] One could still form gauge-equivalence states by superimposing holonomy states related by simultaneous conjugation, but it is not clear how (apart from brute force) to compute the size of the simultaneous conjugation classes and therefore the normalization of the gauge-equivalence states.

Nonetheless, we can simply describe the gauge-invariant Hilbert space as the space of states of the form,

$$|\psi\rangle = \sum_{h_1, h_2 \in G} \psi(h_1, h_2)|h_1, h_2\rangle, \qquad (20)$$

where the wave function $\psi$ satisfies $\psi(gh_1 g^{-1}, gh_2 g^{-1}) = \psi(h_1, h_2)$ for all $g \in G$. Unfortunately, again, functions which are invariant under simultaneous conjugation admit no simple characterization. Given a character $\chi$ of $G$, all the traced Wilson loops $\chi(h_1)$, $\chi(h_2)$, $\chi(h_1 h_2)$ [or more generally $\chi(h)$ where $h$ is an arbitrary product of $h_1$ and $h_2$] are gauge-invariant, but there is no guarantee that they *span* the space of gauge-invariant functions. In fact, as we will see in Sec. III C, this is not true in general.

In the next section, we generalize the construction of this chapter to the case where space is discretized as an arbitrary graph. In particular, we construct holonomy states in full generality. In Sec. III C we discuss under what conditions Wilson loops are sufficient to describe the gauge-invariant functions. Then in Sec. III D we show how to construct matrix elements of the Hamiltonian in the holonomy basis.

### B. The holonomy basis

In this section, we generalize the construction exemplified in the previous section to arbitrary graph geometries. We describe a basis of "holonomy states" which can be implemented without requiring knowledge of any coefficients or normalizations. They satisfy all Gauss law constraints except one, which is then treated separately.

In what follows, we use the word "path" in a loose sense. In particular, we say that a path $\gamma$ between two sites $x$ and $y$ in a graph $\Lambda$ is a sequence of sites $x_0, x_1, \ldots, x_n \in \Lambda$ such that $\langle x_i, x_{i+1}\rangle$ is a link in $\Lambda$ for each $i$ and $x_0 = x$, $x_n = y$. The links can be traversed in either the positive or negative orientations indifferently. Then a path $\gamma$ is also naturally associated with an ordered set of links in the graph, i.e., the links $\langle x_i, x_{i+1}\rangle$.

Given a gauge field configuration, we can associate to each path $\gamma$ given by links $l_1, l_2, \ldots, l_n$ a Wilson line $h_\gamma \in G$ defined as

$$h_\gamma = \prod_{i=1}^n g_{l_i}^{\sigma(l_i, \gamma)}, \qquad (21)$$

---

[1]In particular, it is not conjugation in the group $G \times G$, which would allow the two holonomies to be conjugated by different group elements.

where $\sigma(l_i, \gamma) = \pm 1$ depending on whether $\gamma$ traverses $l_i$ in the positive or negative orientation. Note that, importantly, no trace is taken. If the path $\gamma$ starts and ends at the same point, we refer to $h_\gamma$ as a "holonomy". Since the holonomies are defined without a trace, the base point is important.

Now we follow closely the procedure described in [15]. Starting from the graph $\Lambda$ which discretizes space, pick a maximal spanning tree $T$, i.e., a subgraph of $\Lambda$ which includes all vertices of $\Lambda$ but has no closed loops. (A connected graph always has a spanning tree [29].) Pick an arbitrary special site $x_0$, which we call the *root*. Examples of spanning trees are shown in Figs. 1 and 2. By construction $T$ has $V$ vertices (the same as $\Lambda$) and, by the general properties of tree graphs, $V - 1$ links [29]. Therefore $\Lambda$ has $L$ links in total, of which $V - 1$ are in $T$ and $M \equiv L - V + 1$ outside $T$. Note also that between any two sites $x$ and $y$ in $\Lambda$ there is a unique path (i.e., a unique set of links) lying entirely within $T$ [29]. The construction is illustrated in Fig. 2. Given a gauge field configuration, construct a specific set of holonomies associated to it, which we call the *fundamental holonomies*, as follows:

(1) For each link $\tilde{l} = \langle xy \rangle \in \Lambda \backslash T$, i.e., *outside* $T$, construct the closed path $\gamma$ which starts from the root $x_0$, follows the unique path in $T$ from $x_0$ to $x$, follows $\tilde{l}$ to reach $y$ and then follows the unique path in $T$ from $y$ back to $x_0$.

(2) Construct the holonomy associated to $\gamma$.

(3) Note that with this procedure we have constructed $M = L - V + 1$ independent holonomies, each associated to a link outside the tree. We call these the *fundamental holonomies*. Any other holonomy can be obtained by a finite product of fundamental holonomies.

In particular, note that the fundamental holonomies are invariant under all gauge transformations which equal the identity at the root, i.e., $g_{x_0} = 1$. They are *not* invariant under gauge transformations at the root. For ease of reference we call *internal* gauge transformations those which equal identity at the root, i.e., $g_{x_0} = 1$. We call the remaining gauge transformation, the one based at the root $x_0$, the *external* gauge transformation. Then we say that the fundamental holonomies are invariant under the

internal gauge transformations, which form a group $\mathcal{G}_{\text{int}}$; on the other hand, under the external gauge transformation all holonomies transform in the same way, i.e., $h_\gamma \rightarrow g_{x_0} h_\gamma g_{x_0}^{-1}$. The external gauge transformations also form a group which we call $\mathcal{G}_{\text{ext}}$.

Among all gauge field configurations with the same set of $M$ holonomies $\{h\}$, there is one which is particularly simple. It is the one where all links in the maximal tree $T$ are set to the identity, and we refer to it as the "gauge-fixed" configuration. Formally we denote it as

$$|\{h\}_{\text{GF}}\rangle \equiv \bigotimes_{l \in \text{links}} |g_l\rangle \qquad g_l = \begin{cases} h_l & l \in \Lambda \backslash T \\ 1 & l \in T \end{cases}, \qquad (22)$$

where here $h_l$ is the holonomy associated to link $l$ outside the tree. It is clear that the gauge field configuration $|\{h\}_{\text{GF}}\rangle$ has fundamental holonomies $\{h\}$.

As is clear from the construction, knowledge of the fundamental holonomies and of the gauge field on the tree is sufficient to reconstruct the gauge field configuration. Since also general (nonfundamental) holonomies are invariant under internal gauge transformations, this also implies that the fundamental holonomies with root $x_0$ are sufficient to reconstruct all holonomies with the same root $x_0$. In fact, since they can be equivalently evaluated on the gauge-fixed configuration, all holonomies with root $x_0$ are given by products of the fundamental holonomies. Moreover, any holonomy with arbitrary base point is given, up to conjugation, by a product of fundamental holonomies; if $\gamma$ is a closed path with base point $x \neq x_0$ and $\omega$ is the path from $x_0 \rightarrow x$ in the spanning tree, then $\omega \circ \gamma \circ \omega^{-1}$ is another closed path, but with base point $x$. An example of this occurs in Fig. 2. In what follows, we will often refer to the fundamental holonomies simply as the "holonomies".

Now we show an important result for what follows:

*Proposition 1.* Let $|\{g\}\rangle$ be a gauge field configuration with fundamental holonomies $\{h\}$. Then $|\{g\}\rangle$ is internally gauge-equivalent to the gauge-fixed configuration; i.e., there is an internal gauge transformation $\mathcal{G}$ such that $|\{g\}\rangle = \mathcal{G}|\{h\}_{\text{GF}}\rangle$.

Variations on this result have certainly appeared before in the literature; see for example [30] for a similar statement and proof.
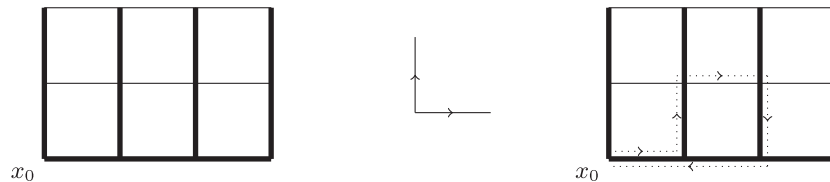


FIG. 2. Illustration of the construction of the holonomies on a $3 \times 4$ square lattice with open boundaries. The reference orientation of the links is shown in the center figure. On the left, the lattice together with a maximal tree (the bold links) and root $x_0$. On the right, a path making up an holonomy according to the construction described in the text. We chose to illustrate this case for simplicity, but our construction works for the general case where space is discretized as an arbitrary graph.

*Proof.* We will exhibit a sequence of internal gauge transformations which gauge-fixes to the identity the gauge field in the tree. Note that a product of internal gauge transformations is also an internal gauge transformation, and that internal gauge transformations preserve the holonomies. Therefore the end result of this procedure is the gauge-fixed configuration.

Consider the spanning tree $T$. Call *active* the sites at which we have not yet fixed the gauge transformation, and *inactive* otherwise. If the gauge field $g_l$ at link $l$ has been set to the identity by a gauge transformation, we mark it as *frozen*. Initially, the root site is inactive and all other sites are active. A *leaf* of the tree $T$ is a site which is connected only to one other site. Every tree has at least one leaf [29]. Then pick a leaf $x$ and consider the path from the root $x_0$ to $x$ (this is unique [29]), which takes the form $x_0, x_1, x_2, \ldots, x_k$ with $x_k = x$. Then use the gauge freedom at $x_1$ to set the gauge field on link $l = \langle x_0 x_1 \rangle$ to the identity; then mark $x_1$ as inactive and $l$ as frozen. Note that since both sites connected to $l$ are inactive, its value cannot be changed by the following gauge transformations. Then continue to use the gauge freedom at site $x_i$ to fix the gauge field on link $\langle x_{i-1} x_i \rangle$. Thus all the links in the path between $x_0$ and $x_k$ are frozen and all the sites in the path are inactive. Now among inactive sites, choose one (say $y_0$) which is connected to a link which is not frozen. If it does not exist, then all links are frozen and the algorithm terminates. Otherwise, consider the subgraph formed by $y_0$ and all sites and edges reachable via nonfrozen links. Since this is a subgraph of $T$, it is again a tree and therefore will have a leaf. Thus we can repeat the previous algorithm whereby we consider the path from $y_0$ to the leaf and fix the gauge field along the path. Then one again picks another inactive site recursively until all links in the tree are frozen. ∎

Using the previous proposition, we can show the following:

*Proposition 2.* Two gauge field configuration states $|\{g\}\rangle$ and $|\{g'\}\rangle$ have the same fundamental holonomies if and only if they are related by an internal gauge transformation.

*Proof.* One direction is obvious. We know that internal gauge transformations do not change the holonomies; therefore if $|\{g\}\rangle$ and $|\{g'\}\rangle$ are related by an internal gauge transformation, then they have the same holonomies. For the other direction, we make use of Proposition 1 according to which both $|\{g\}\rangle$ and $|\{g'\}\rangle$ are related to the gauge-fixed configuration via an internal gauge transformation. The result then follows because internal gauge transformations form a group. ∎

All the results until now deal with gauge field configurations and are therefore well-known. Now we deal more specifically with the consequences for the physical Hilbert space.

We define a *holonomy class* to be the equivalence class of gauge field configurations with the same fundamental holonomies. We have said before that gauge equivalence classes do not all have the same number of elements. The remarkable fact about *holonomy classes* is that they *do* all have the same number of elements. Before proving this fact, we show a refinement of the previous propositions,

*Proposition 3.* Within each holonomy class, configuration states are in one-to-one correspondence with internal gauge transformations.

*Proof.* Given a configuration state $|\{g\}\rangle$ with holonomies $\{h\}$, from Proposition 1 we know that there is an internal gauge transformation $\mathcal{G}$ such that $|\{g\}\rangle = \mathcal{G}|\{h\}_{\mathrm{GF}}\rangle$. Then we define a map from configurations to internal gauge transformations by $|\{g\}\rangle \to \mathcal{G}$, and we have to show that it is bijective.

First of all, we have to show that the map is well-defined. In particular, we have to show that if we have two gauge transformations $\mathcal{G}$ and $\mathcal{G}'$ such that

$$|\{g\}\rangle = \mathcal{G}|\{h\}_{\mathrm{GF}}\rangle, \tag{23}$$

$$|\{g\}\rangle = \mathcal{G}'|\{h\}_{\mathrm{GF}}\rangle, \tag{24}$$

then $\mathcal{G} = \mathcal{G}'$. In fact uniqueness follows from the proof of Prop. 1 because gauge transformations on different sites commute and paths in a tree are unique. But we prove this result again here in a simpler manner. Since internal gauge transformations form a group, they have inverses. Therefore from Eq. (23) we have

$$|\{h\}_{\mathrm{GF}}\rangle = \mathcal{G}^{-1}\mathcal{G}'|\{h\}_{\mathrm{GF}}\rangle. \tag{25}$$

Therefore the gauge-fixed configuration is fixed by the internal gauge transformation $\mathcal{G}'\mathcal{G}^{-1}$. But in the gauge-fixed configuration, each link in the tree is set to the identity. For a link $l = \langle xy \rangle$ in the tree, $\mathcal{G}'\mathcal{G}^{-1}$ acts as $|1\rangle \to |g_x g_y^{-1}\rangle = |1\rangle$ for some group elements $g_x, g_y \in G$. Therefore $g_x = g_y$. But since the tree is maximal, i.e., it includes all vertices, this then implies that $\mathcal{G}'\mathcal{G}^{-1}$ is given by the same group element on *all* sites. But since it is an internal gauge transformation, it is equal to the identity on the root $x_0$; therefore it must be given by the identity on *all* sites, i.e., it is the identity transformation. Therefore $\mathcal{G}' = \mathcal{G}$.

It remains to show that the map is bijective. But this can be done by exhibiting an explicit inverse, which is simply given by the map $\mathcal{G} \to |\{g\}\rangle \equiv \mathcal{G}|\{h\}_{\mathrm{GF}}\rangle$ (note that the inverse is defined on the whole codomain). ∎

Incidentally, this proves that the group $\mathcal{G}_{\mathrm{int}}$ of internal gauge transformations is really isomorphic to $G^{V-1}$. On the other hand, the group of external gauge transformations $\mathcal{G}_{\mathrm{ext}}$ is *not* isomorphic to $G$. In fact it is clear that if $g$ is in the center $Z(G)$ of $G$, then the corresponding external gauge transformation $\mathcal{G}(g)$ is simply the identity. Therefore $\mathcal{G}_{\mathrm{ext}} \cong G/Z(G)$. Similarly, the full group of gauge transformations is isomorphic to $G^V/Z(G)$ where $Z(G)$ is the

"diagonal" subgroup of constant transformations equal to a central element (which act as the identity).

We can now show that all holonomy classes have the same number of states:

*Proposition 4.* Each holonomy class contains $|G|^{V-1}$ configurations.

*Proof.* Suppose that we want to construct a configuration with prescribed fundamental holonomies. These are always defined with respect to a spanning tree, say $T$. Following Propositions 1 and 2, we are free to set the gauge field configuration on each link in the tree arbitrarily. Then for each link not in $T$, we assign to it the unique group element which fixes its fundamental holonomy to the prescribed one. Hence are free to set all and only the links on the tree $T$. Since there are $V-1$ such links, we can construct $|G|^{V-1}$ states with the same fundamental holonomies. ∎

This statement can also be proved from an alternative point of view:

*Proof.* From Proposition 3 within each holonomy class configurations are in one-to-one correspondence with internal gauge transformations. Therefore it suffices to count the number of internal gauge transformations. Since we are free to choose $|G|$ elements for $V-1$ sites (i.e., all sites except the root), then we have $|G|^{V-1}$ internal gauge transformations and therefore the same number of configurations in each holonomy class. ∎

This means, in particular, that we can define the "holonomy class states",

$$|\{h\}\rangle \equiv |h_1, h_2, ..., h_M\rangle \equiv \frac{1}{\sqrt{|G|}^{V-1}} \sum_{\{g\} \in [h]} |\{g\}\rangle, \quad (26)$$

by an equal superposition of all configuration states with the same holonomies. Here $[h]$ denotes an holonomy class, i.e., a definite choice of $h_1, h_2, ..., h_M \in G$. Since all holonomy classes have the same size, the normalization for these states is known and in fact the states $|\{h\}\rangle$ are orthonormal,

$$\langle\{h'\}|\{h\}\rangle = \prod_{i=1}^{M} \delta(h'_i, h_i). \quad (27)$$

We call the Hilbert space spanned by the holonomy states $\mathcal{H}_{hol}$. Since each holonomy can take $|G|$ values and we have $M = L - V + 1$ holonomies, then $\dim \mathcal{H}_{hol} = |G|^{L-V+1}$. In particular, $\mathcal{H}_{hol}$ admits a remarkably simple description as a tensor product space,

$$\mathcal{H}_{hol} = \mathbb{C}[G]^{\otimes M}. \quad (28)$$

In other words, $\mathcal{H}_{hol}$ is simply the Hilbert space spanned by states of the form $|h_1, h_2, ...h_M\rangle$ where each holonomy $h_i \in G$ can take any value in $G$ without restrictions. Therefore the holonomy states satisfy almost all gauge constraints, thus reducing the size of the Hilbert space by an

exponential factor, without requiring the computation of normalizations or Clebsch-Gordan coefficients. One is simply left with one overall global constraint, i.e., the external gauge transformation. Note in particular that the description of the holonomy space is entirely independent of the embedding of the holonomies into any particular graph, only *operators* depend on the embedding.

The holonomy states are generally not gauge-invariant because of the external gauge transformations $\mathcal{G}_{ext}$. In fact a transformation $\mathcal{G}(g) \in \mathcal{G}_{ext}$ acts in holonomy space as

$$\mathcal{G}(g)|h_1, h_2, ..., h_M\rangle = |gh_1g^{-1}, gh_2g^{-1}, ..., gh_Mg^{-1}\rangle. \quad (29)$$

In holonomy space, the external gauge transformation takes the role of a global symmetry which acts by simultaneous conjugation of all holonomies. Therefore the Hilbert space of holonomies is an intermediate space between the total Hilbert space, and the physical Hilbert space,

$$\mathcal{H}_{tot} \xrightarrow{\mathcal{G}_{int}} \mathcal{H}_{hol} \xrightarrow{\mathcal{G}_{ext}} \mathcal{H}_{phys}. \quad (30)$$

Here the arrows denote projection onto the states invariant under $\mathcal{G}_{int}$ and $\mathcal{G}_{ext}$ respectively. The projector from $\mathcal{H}_{hol}$ to $\mathcal{H}_{phys}$ is then constructed by averaging over external gauge transformations; i.e., it is given by the operator,

$$P|h_1, ..., h_M\rangle = \frac{1}{|G|} \sum_{g \in G} |gh_1g^{-1}, ..., gh_Mg^{-1}\rangle. \quad (31)$$

In terms of the $L$ and $R$ operators defined in Sec. II (now acting on the Hilbert space of each holonomy), it is given by

$$P = \frac{1}{|G|} \sum_{g \in G} (L_g R_g)^{\otimes M}. \quad (32)$$

It is not hard to show that $P$ is indeed a projector. We can then use it to rederive a formula for the dimension of the physical subspace first obtained in [14]:

*Theorem 1.* If $C$ are the conjugacy classes of $G$, and $|C|$ their sizes, then on a graph with $V$ vertices and $L$ edges one has $\dim \mathcal{H}_{phys} = \sum_C \left(\frac{|G|}{|C|}\right)^{L-V}$.

*Proof.* As we have seen, $P$ projects onto the gauge-invariant sector. As usual, the dimension of the image of the projector (in this case, the gauge-invariant subspace) equals the trace of the projector. Therefore,

$$\dim \mathcal{H}_{phys} = \text{tr}P = \frac{1}{|G|} \sum_{g \in G} [\text{tr}(L_g R_g)]^M. \quad (33)$$

By the Peter-Weyl decomposition [21,31], we find

$$L_g R_g = \bigoplus_j \rho_j(g)^* \otimes \rho_j(g), \quad (34)$$

where $\rho_j$ are the irreps of $G$. Then taking traces, one finds

$$\text{tr}(L_g R_g) = \sum_j \chi_j(g)^* \chi_j(g) = \frac{|G|}{|C(g)|}, \qquad (35)$$

where we used one of the orthogonality relations of characters [20] and $C(g)$ is the conjugacy class of $g$. Therefore we find

$$\dim \mathcal{H}_{\text{phys}} = \text{tr}P = \frac{1}{|G|} \sum_{g \in G} \frac{|G|^{L-V+1}}{|C(g)|^{L-V+1}} \qquad (36)$$

$$= \sum_C \sum_{g \in C} \frac{|G|^{L-V}}{|C|^{L-V+1}} = \sum_C \frac{|G|^{L-V}}{|C|^{L-V}}, \qquad (37)$$

which proves the result. ∎

In fact this statement could also have been proven by computing the trace of the projector from the total Hilbert space, Eq. (12), by combining the argument presented here with the one in [14].

Note that in an Abelian group, all elements commute and therefore $gh_ig^{-1} = h_i$ so that the external gauge transformation does nothing. In this case, therefore the space of holonomies coincides with the physical Hilbert space, $\mathcal{H}_{\text{hol}} \cong \mathcal{H}_{\text{phys}}$.

The statement that all holonomy classes have the same size may also be understood as the statement that the transformation from configurations to holonomies has constant Jacobian. Consider in fact a gauge-invariant function of configurations $f(\{g\})$, i.e., $f : G^L \to \mathbb{C}$. This may also be thought of as the wave function of a gauge-invariant state. Since $f$ is gauge-invariant, it depends only on the holonomies of $\{g\}$. Therefore we can define a function of the holonomies $F : G^M \to \mathbb{C}$ via $F(\{h\}) \equiv f(\{h\}_{\text{GF}})$. Then we have the change of variable formula,

$$\frac{1}{|G|^L} \sum_{\{g\}} f(\{g\}) = \frac{1}{|G|^M} \sum_{\{h\}} F(\{h\}). \qquad (38)$$

Here $\frac{1}{|G|} \sum_{g \in G}$ should be thought of as the normalized Haar measure on the finite group; the same formula also holds for compact Lie groups by replacing this with their Haar measure.

It is also interesting to compare the dimension of the two spaces. One has

$$\frac{\dim \mathcal{H}_{\text{phys}}}{\dim \mathcal{H}_{\text{hol}}} = \frac{1}{|G|} \sum_C \frac{1}{|C|^{L-V}} \geq \frac{1}{|G|}. \qquad (39)$$

Therefore, in going to the physical subspace one gains at most a factor of $\approx |G|$ in terms of "memory" compared to the holonomy space, paying a price in terms of having to store the Clebsch-Gordan coefficients which express the gauge-invariant states in terms of the holonomy states.

## C. Wilson loops as a basis

For $SU(N)$ gauge theories, it is standard to expand gauge-invariant functions in terms of traced Wilson loops. These are generally overcomplete, but at least for SU(2) it was possible to find a subset of loops which form a proper basis [32]. As anticipated in Sec. III A, in general the traced Wilson loops do not even span the space of gauge-invariant functions.

In this section, we review the issue of whether a basis of gauge-invariant states may be found in terms of traced Wilson loops for a general finite or compact Lie gauge group $G$. This issue is closely related to the discussion of this section, and the results discussed here have been obtained by several authors [15,30,33,34]. From the previous discussion, in terms of the holonomy basis, gauge-invariant states take the form,

$$|\psi\rangle = \sum_{\{h\}} \psi(h_1, h_2, ..., h_M) |h_1, h_2, ..., h_M\rangle, \qquad (40)$$

where the wave function in the holonomy basis satisfies

$$\psi(gh_1g^{-1}, gh_2g^{-1}, ..., gh_Mg^{-1}) = \psi(h_1, h_2, ..., h_M), \qquad (41)$$

for all $g \in G$. As we saw in Sec. III A, for $M = 1$ this is simply a class function on $G$ (a basis of which is given by the characters of $G$), but for $M > 1$ this condition admits no such simple mathematical description. One may nonetheless attempt to construct a basis of such scalar functions, which would then also provide a basis of gauge-invariant states. A simple observation is that a set of functions which satisfies Eq. (41) are of the form $f(h)$ where $f$ is a class function on $G$ and $h$ is a product of fundamental holonomies, i.e.,

$$h = h_{i_1} h_{i_2} h_{i_3} \cdots, \qquad (42)$$

where each fundamental holonomy appears an arbitrary number of times. But since the irreducible group characters form a basis of class functions, it is equivalent to consider a basis of scalar functions of the form $\chi(h)$, where $\chi$ is an irreducible character and $h$ a product of holonomies. These are precisely all the possible traced Wilson loops in all the irreducible representations of $G$.

One might then attempt to prove that all functions satisfying Eq. (41) may be expanded in a basis of traced Wilson loops; but this turns out not to be possible. This is not only because such a basis would be overcomplete, but more importantly because *the traced Wilson loops do not necessarily span the space of functions satisfying* Eq. (41). In fact for some gauge groups $G$ it is possible to construct

*orthogonal* gauge-invariant states with the *same* Wilson loops [30].

Nonetheless, the traced Wilson loops *do* span the space of gauge-invariant functions in some cases of interest (while still being overcomplete). For example if the group $G$ is Abelian, then the condition Eq. (41) is trivial. For $\mathbb{Z}_N$ and U(1) therefore we obtain the even stronger result that Wilson loops in any one faithful irreducible representation are sufficient to determine the gauge-invariant content of configurations. But this result is in a sense not very interesting, as the above description in terms of holonomy states is already fully sufficient to describe the gauge-invariant states.

More generally, note that if $G$ is either finite or compact Lie and $\chi(g) = \chi(g')$ for all irreducible characters $\chi$, then $g$ and $g'$ are conjugate [20,21]. Suppose that one has two states with holonomies $\{h\}$ and $\{h'\}$. Then one can consider all possible independent products of holonomies, which we call $\tilde{h}_\alpha$ and $\tilde{h}'_\alpha$ indexed by $\alpha$. If the Wilson loops form a basis of gauge-invariant states then they must agree on $\{h\}$ and $\{h'\}$ if and only if the two are gauge-equivalent. From the previous result we know that if $\chi(\tilde{h}_\alpha) = \chi(\tilde{h}'_\alpha)$ for all $\chi$ (i.e., the Wilson loops agree), then $\tilde{h}_\alpha$ and $\tilde{h}'_\alpha$ are conjugate. But $\{h\}$ and $\{h'\}$ are gauge-equivalent if and only if $h'_i = gh_ig^{-1}$ for some $g$, which must be the *same* for all holonomies; therefore it is not sufficient that $\tilde{h}_\alpha$ and $\tilde{h}'_\alpha$ are conjugate for all $\alpha$, but they must be conjugate *by the same group element for all $\alpha$.* In order to guarantee this, one must exclude the presence of certain group homomorphisms. In fact suppose that $\phi$ is a class-preserving outer automorphism [15,30,35,36], i.e., $\phi: G \to G$ is an invertible group homomorphism such that $g$ and $\phi(g)$ are always in the same conjugacy class, but $\phi(g)$ is *not* of the form $\phi(g) = kgk^{-1}$ for some fixed $k \in G$ and all $g \in G$. Then the configurations $\{g\}$ and $\{\phi(g)\}$ have holonomies $\{h\}$ and $\{\phi(h)\}$ and identical Wilson loops, $\chi(h) = \chi(\phi(h))$ on account of $\phi$ preserving conjugacy classes. Yet the configurations $\{g\}$ and $\{\phi(g)\}$ are not gauge-equivalent, since $\phi(g) \neq kgk^{-1}$ for fixed $k \in G$, and therefore their holonomies are not related by an external gauge transformation. But note that it is *not* sufficient to exclude automorphisms $\phi: G \to G$, because one may consider gauge field configurations lying entirely within a subgroup of $G$. More generally, one must prove that for any two (closed) subgroups $H_1$ and $H_2$ of $G$, an isomorphism $\phi: H_1 \to H_2$ such that $g$ and $\phi(g)$ are conjugate in $G$, must be of the form $\phi(g) \equiv kgk^{-1}$ for all $g \in H_1$ and some $k \in G$ [15].

Putting together the results of several authors [15,33,34], one finds that Wilson loops determine gauge-invariant states if the gauge group $G$ is a finite direct product of Abelian groups together with any finite number of copies of the groups U(N), SU(N), O(N), SO(N), Sp(N) (i.e., the subgroup of GL(2N)). This property is in fact preserved by taking direct products, but not by subgroups, and not even

by quotients or central extensions [34]. Thus the question in general needs to be re-examined for every gauge group. In particular, these results *do not* cover the general case of Lie groups (for example it is known to be false for some noncompact Lie groups [33]) or even compact Lie groups; for example, one could construct counterexample gauge groups of the form $G \ltimes H$, where $G$ is a finite group and $H$ a compact Lie group. In fact groups of this form have attracted interest recently [37,38]. Among finite groups of interest, Wilson loops determine gauge-invariant states for the symmetric group $S_N$ [15], as well as for the dihedral group $D_4$ and the quaternion group $Q$ which we consider in Sec. IV.

As is clear from the above discussion, when we say that Wilson loops determine gauge-invariant states, we must in general consider Wilson loops in *all* irreducible representations. Whether it is sufficient to consider Wilson loops in any one (faithful) irrep is a separate question. In particular, Wilson loops in the fundamental representation are sufficient for the usual cases of U(N) and SU(N), but not (for example) for SO(2N) [15].

### D. Matrix elements of the Hamiltonian

The only remaining piece in order to be able to use the holonomy basis is to compute the matrix elements of relevant operators. All operators which are gauge-invariant can be expressed in the holonomy basis. In particular, we compute the matrix elements of the Hamiltonian, which will be useful in Sec. IV. In this section, we restrict to the case of interest where the gauge group is finite. The Hamiltonian for this case was discussed in Sec. II. Moreover, as already remarked, in order to define the magnetic Hamiltonian a graph structure is not sufficient; therefore in this section we restrict ourselves to hypercubic lattices with either open or periodic boundaries. Of course one could also define the magnetic Hamiltonian on other lattices with a notion of minimal cycle (for example triangular lattices, etc.). On the other hand, the results for the electric Hamiltonian apply to general graphs.

The magnetic Hamiltonian is diagonal in the group-element basis and is therefore also diagonal in the holonomy basis. Each (untraced) plaquette can be expressed (up to conjugation) as a product of the fundamental holonomies. Once this is done, it is easy to compute the value of the magnetic Hamiltonian on each holonomy state.

On the other hand, the electric Hamiltonian is *not* diagonal. Yet its matrix elements are not hard to compute. On each link, the electric Hamiltonian is given by

$$h_E = |\Gamma|\mathbb{1} - A, \qquad A \equiv \sum_{k \in \Gamma} L_k, \qquad (43)$$

where we have removed the constant for ease of computation. Since $h_E$ is gauge-invariant, so is $A$ (this is also not hard to check explicitly using the properties of $\Gamma$ given in

Sec. II). Now consider the action of $A$ on an arbitrary holonomy state, $A|\{h\}\rangle$. As we have seen from Proposition 3, configurations with given holonomies are in one-to-one correspondence with internal gauge transformations. Therefore we can express the holonomy state as

$$|\{h\}\rangle = \frac{1}{\sqrt{|G|}^{V-1}} \sum_{\{g\} \in [h]} |\{g\}\rangle = \frac{1}{\sqrt{|G|}^{V-1}} \sum_{\mathcal{G} \in \mathcal{G}_{\text{int}}} \mathcal{G}|\{h\}_{\text{GF}}\rangle. \tag{44}$$

Since $A$ is gauge-invariant it commutes with $\mathcal{G}$, and we then see that

$$A|\{h\}\rangle = \frac{1}{\sqrt{|G|}^{V-1}} \sum_{\mathcal{G} \in \mathcal{G}_{\text{int}}} \mathcal{G}A|\{h\}_{\text{GF}}\rangle. \tag{45}$$

Since applying $L_k$ to a configuration state produces another configuration state, applying $A$ to the gauge-fixed state $|\{h\}_{\text{GF}}\rangle$ produces $|\Gamma|$ configuration states $|\{g'\}\rangle, |\{g''\}\rangle, \ldots$ with holonomies $\{h'\}, \{h''\}, \ldots$. But then acting with all internal gauge transformations on a configuration state $|\{g'\}\rangle$ with holonomies $\{h'\}$, produces precisely the holonomy state $|\{h'\}\rangle$, i.e.,

$$|\{h'\}\rangle = \frac{1}{\sqrt{|G|}^{V-1}} \sum_{\mathcal{G} \in \mathcal{G}_{\text{int}}} \mathcal{G}|\{g'\}\rangle. \tag{46}$$

As such, the action of $A$ on holonomy states takes the form,

$$A|\{h\}\rangle = |\{h'\}\rangle + |\{h''\}\rangle + \cdots. \tag{47}$$

Therefore the matrix elements of $A$ can be computed as follows:

(1) For each holonomy state $|\{h\}\rangle$ produce the gauge-fixed configuration $|\{h\}_{\text{GF}}\rangle$.
(2) For each $k \in \Gamma$, consider the configuration $|\{g'\}\rangle = L_k|\{h\}_{\text{GF}}\rangle$ and compute its holonomies $\{h'\}$.
(3) For each of these, set the matrix element $\langle\{h'\}|A|\{h\}\rangle = 1$.

All matrix elements which are not set to one are equal to zero. This is repeated to construct $A$ (and therefore $h_E$) for each link, and then the electric Hamiltonian is obtained by addition, $H_E = \sum_{\text{links}} h_E$. Note that all operators, including $H_B$ and $H_E$, are expressed in the holonomy basis as $|G|^M \times |G|^M$ matrices.

Since the holonomies are defined as products of the gauge field on up to $\mathcal{O}(V)$ links, one might have naively expected that, even on one link, the electric Hamiltonian $h_E$ would have involved up to $\mathcal{O}(V)$ terms. Instead, as we have seen, $h_E$ involves only $\mathcal{O}(|\Gamma|)$ terms (independent of $V$), which is favorable.

From this description, we see that the electric Hamiltonian in the holonomy basis is guaranteed to be sparse. In particular, for each $h_E$ on each row one has $|\Gamma|$

non-zero elements. Since we have $L$ links, then $H_E$ has at most $|\Gamma|L$ nonzero elements per row, therefore it will overall have order $\approx |\Gamma|L|G|^M$ nonzero elements, i.e., the proportion of nonzero elements is an exponentially small number, $\approx |\Gamma|L/|G|^M$. Notably, $H_E$ is also integer-valued.

If the group $G$ is Abelian, the matrix elements can be described more explicitly. In fact, in the Abelian case the $L_k$ operator on each link is actually gauge-invariant, and therefore maps holonomy states to holonomy states. For clarity, we denote the $L_k$ operator acting on link $l$ as $L_k|_l$. As we will show in a moment, if a link $l$ occurs in the path of a fundamental holonomy more than once, then it occurs exactly twice and with opposite orientation. Therefore, since the group is Abelian, one in fact has

$$L_k|_l|\{h\}\rangle = |\{\tilde{h}\}\rangle, \qquad \tilde{h}_\gamma = k^{\sigma(l,\gamma)}h_\gamma, \tag{48}$$

where $\sigma(l,\gamma) = \pm 1$ depending on whether the holonomy path $\gamma$ crosses $l$ in the positive or negative orientation, and is zero if $\gamma$ does not cross $l$. Therefore the matrix elements of $L_k$, from which those of $h_E$ can be easily deduced, are given by

$$\langle\{h'\}|L_k|_l|\{h\}\rangle = \prod_{\gamma \not\ni l} \delta(h'_\gamma, h_\gamma) \prod_{\gamma \ni l} \delta(h'_\gamma h_\gamma^{-1}, k^{\sigma(l,\gamma)}), \tag{49}$$

where the first product runs over those fundamental holonomies which do not go through link $l$, and the second product goes over those that do. It remains to prove the above statement, with which we close this section.

*Proof.* A path $\gamma$ defining a fundamental holonomy is of the form $\gamma = \gamma_1 \circ \tilde{l} \circ \gamma_2$, where $\gamma_1$ and $\gamma_2$ lie entirely inside the chosen spanning tree and $\tilde{l}$ is a link outside the tree. By construction, each link appears in $\gamma_i$ at most once. Therefore it can appear at most twice in $\gamma$. So we have to show that if a link $l$ appears in both $\gamma_1$ and $\gamma_2$, then it appears with opposite orientations. Suppose on the contrary that it appears with the *same* orientation. If $l$ connects the sites $x$ and $y$ according to the orientation defined by the paths (in either the positive or negative orientation), then there are subpaths $\gamma'_i \subset \gamma_i$ (therefore lying entirely inside the maximal tree) such that $\gamma'_1$ connects the root $x_0$ to $x$ and $\gamma'_2$ connects $y$ to $x_0$. Since $x$ is the source site for $l$ according to the orientation defined by the path, by construction $\gamma'_1$ does not go through $l$; similarly, $y$ is the target link of $l$, so $\gamma'_2$ also does not go through $l$. But then the closed path $\gamma'_1 \circ l \circ \gamma'_2$ lies entirely within the tree. Therefore it must be the *trivial* closed path. But this is impossible, since neither $\gamma'_1$ nor $\gamma'_2$ contains $l$. ∎

## E. Topology and center symmetry

In this section, we discuss further structure of the Hilbert space and how it interacts with the holonomy basis. Similarly to the toric code [39], we discuss two kinds of

topological operators, those which are diagonal in the group element basis (i.e., holonomies wrapping around noncontractible loops on the lattice) and those which are diagonal in the dual basis where the electric Hamiltonian is diagonal (these are related to center symmetry). A discussion of these issues in a similar language was given in [40] for Abelian gauge theories on a ladder. Here we extend the discussion to non-Abelian gauge theories on an arbitrary geometry.

First of all we discuss why in Yang-Mills gauge theories the electric Hamiltonian does not commute with Wilson lines or loops. Suppose that we want to fix a certain holonomy $h_i$, or even a product of holonomies. This is only possible if two separate conditions are satisfied:

(1) *Compatibility with gauge-invariance.* As we have seen, under external gauge transformations the holonomy $h_i$ transforms as $h_i \rightarrow gh_ig^{-1}$ for $g \in G$. In order to fix this holonomy, it must be invariant under gauge transformations, i.e., $gh_ig^{-1} = h_i$ for all $g \in G$. In other words, $h_i$ must lie in the center of the gauge group $Z(G)$. The same is true for any product of holonomies, since it transforms in the same way.

(2) *Compatibility with time evolution.* An holonomy can only be consistently fixed if it cannot change value under time evolution. Since the magnetic Hamiltonian is diagonal in the holonomy basis, it does not cause transitions between states with different holonomies. The electric Hamiltonian, on the other hand, always causes transitions between different holonomy states. In particular, since the matrix elements of the electric Hamiltonian on each link are either 0 or 1, it suffices to show that this is true at any one link (in other words, there cannot be cancellations when summing over all links). But then consider the electric Hamiltonian on the link outside the maximal tree which defines the holonomy $h_i$; the electric Hamiltonian on that link always has nonzero matrix elements between states with different values of $h_i$, while leaving unchanged all other holonomies. Thus fixing any holonomy, or any product of holonomies, is incompatible with time-evolution according to the Hamiltonian Eq. (7).

One may nonetheless consider other topological operators, which correspond to *dual* winding. Consider a unitary operator of the form,

$$U(\{g\}) = \bigotimes_{l \in \text{links}} L_{g_l}, \qquad (50)$$

where $\{g\}$ is a configuration, i.e., an assignment of a group element $g_l$ to each lattice link $l$. Thus in principle $U(\{g\})$ acts on all lattice links, but we will soon impose further conditions which will restrict its form. Eq. (50) generalizes an operator for the Abelian case in [40]. As we will soon see, it is irrelevant whether one chooses left or right translations to define $U$. Again $U$ is only useful if it

is compatible with the gauge symmetry (i.e., it must be gauge-invariant) and time evolution (i.e., it must be a symmetry of the Hamiltonian).

Consider an arbitrary gauge transformation $\mathcal{G}$, as defined in Eq. (5). Then one can check that

$$
\begin{aligned}
\mathcal{G}U(\{h\})\mathcal{G}^{-1} &= \bigotimes_{l=\langle xy \rangle \in \text{links}} L_{g_x} R_{g_y} L_{h_l} R_{g_y^{-1}} L_{g_x^{-1}} \\
&= \bigotimes_{l=\langle xy \rangle \in \text{links}} L_{g_x h_l g_x^{-1}}.
\end{aligned}
\qquad (51)
$$

Therefore gauge-invariance requires $g_x g_l g_x^{-1} = g_l$ for all $g_x$. Thus $g_l$ must lie in the center of the gauge group $G$. In other words, the operator $U(\{g\})$ is gauge-invariant if and only if $\{g\}$ is a configuration made entirely of central elements. Note that if $g$ is central, then $R_g = L_{g^{-1}}$ so restricting to left translations does not lead to loss of generality.

Now consider time evolution. Using the invariance of $\Gamma$ under conjugation, it is not hard to see that, for arbitrary $g$ and on any link,

$$L_g h_E = h_E L_g. \qquad (52)$$

Therefore $U(\{g\})$ always commutes with the electric Hamiltonian. Note that this is the "dual" statement to the fact that the magnetic Hamiltonian preserves the holonomies.

We then need to consider conditions under which $U(\{g\})$ commutes with the magnetic Hamiltonian. Until now we could work on an arbitrary graph geometry, but the magnetic Hamiltonian requires a bit more structure, so we specialize to a hypercubic lattice. Nonetheless the generic construction worked out in this section can be used to find appropriate operators on arbitrary geometries. As we have seen above, the configuration $\{g\}$ is required to be central for gauge-invariance. Then since $U(\{g\})$ is gauge-invariant, it maps holonomy states to holonomy states. Therefore the fact that the magnetic Hamiltonian commutes with $U(\{g\})$ is equivalent to stating that $H_B$ takes the same value on the configuration states $|\{g'\}\rangle$ and $U(\{g\})|\{g'\}\rangle$ for arbitrary $\{g'\}$. Since the $g_l$s are central, plaquettes transform as

$$g_1' g_2' g_3'^{-1} g_4'^{-1} \rightarrow g_1 g_2 g_3^{-1} g_4^{-1} g_1' g_2' g_3'^{-1} g_4'^{-1}. \qquad (53)$$

As we show in Appendix A the magnetic Hamiltonian on each plaquette, $h_B = -2\text{Re}\chi$ has a unique minimum at the identity, since $\chi$ is assumed to be the character of a faithful representation. Since Eq. (53) must hold for all configurations, it must in particular hold for configurations $\{g'\}$ where the plaquette variable $g_1' g_2' g_3'^{-1} g_4'^{-1}$ is equal to the identity. Thus it is a necessary and sufficient condition that $g_1 g_2 g_3^{-1} g_4^{-1} = 1$ for all plaquettes. In other words, the configuration $\{g\}$ must be a flat configuration living in

the center of the gauge group $G$. Note that this does *not* mean in general that $\{g\}$ is gauge-equivalent to the identity.

Up until this point, we have therefore constructed a large number of gauge-invariant symmetry operators $U(\{g\})$. But, as we will now show, most of these act in the same way in the gauge-invariant subspace. Suppose in fact that we construct another operator $U(\{\tilde{g}\})$ which is related to $U(\{g\})$ by a gauge transformation of the $g_l$s, i.e., $\tilde{g}_{\langle xy\rangle} = k_x g_{\langle xy\rangle} k_y^{-1}$ for some $k_x \in G$. Note that a gauge transformation of $g_l$s does *not* correspond to a gauge transformation of $U$. Gauge invariance requires that both $\{g\}$ and $\{\tilde{g}\}$ are central; it is not hard to show that one may then choose all $k_x$ to be central also. Then one finds that

$$U(\{\tilde{g}\}) = \mathcal{G}U(\{g\}), \qquad (54)$$

where $\mathcal{G}$ is a gauge transformation of the usual form [i.e., Eq. (5)] with (central) group elements $k_x$. In other words, performing a gauge transformation of $\{g\}$ leads to $U$ operators which coincide on the physical subspace and should therefore be viewed as physically equivalent. One can therefore perform gauge transformations of the $\{g\}$ to reduce them to the minimal amount of degrees of freedom.

For example, on a $d$—dimensional periodic hypercubic lattice the independent $U$ operators are the 't Hooft operators responsible for center symmetry. In this case, a flat central configuration $\{g\}$ can be gauge-transformed to one which is the identity everywhere except on the links which pierce a $d-1$ dimensional hypersurface. On these links, it is equal to a constant central element. The 't Hooft operator then takes the form [12],

$$U_S(g) = \bigotimes_{l \in S} L_g, \qquad (55)$$

where $g$ is central and $S$ is the set of links which pierces a $d-1$ dimensional surface. Since $U$ operators related by a gauge transformation of the $g_l$s are physical equivalent, the location of the surface $S$ is immaterial. Therefore the $U$ operators are topological. On a $d$-dimensional periodic hypercubic lattice, one therefore has $d$ independent such operators which cannot be deformed into each other. An example of such an operator on the square lattice is given in Fig. 3. Here the surface $S$ is one dimensional; i.e., it is a line.

In the confining phase of Yang-Mills theory, one expects center symmetry to be unbroken. One could then further reduce the size of the relevant subspace by considering only states which are center-invariant. In fact it is rather simple to construct such a projector, and one can even compute the size of the relevant gauge-invariant, center-invariant subspace following an argument similar to Theorem 1. However, this goes against the philosophy of the holonomy states, whose main advantage is precisely the fact that the description of their Hilbert space does not require Clebsch-Gordan coefficients.
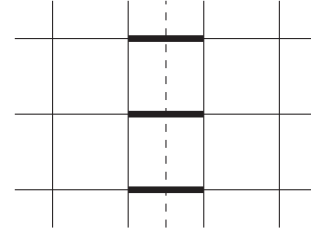


FIG. 3. A square lattice. One of the two 't Hooft operators is supported on the bold links, i.e., those crossing the dashed line. This 't Hooft operator thus multiplies by a central element any loop winding around the horizontal direction.

## IV. GROUND STATE WAVE FUNCTION

The determination of the ground state wave function of Yang-Mills theory is not only a problem of theoretical interest, but also relevant in the quantum simulation setting where some quantum algorithms require a sufficiently good approximation to the ground state as a starting point.

The ground states for the electric and magnetic Hamiltonians can be easily described in the holonomy basis. While the electric Hamiltonian can have ground state degeneracy [14], we do not consider this case here. On each link, the ground state of the electric Hamiltonian is then given by an equal superposition of all group elements. Tensoring across all links, we find the electric ground state,

$$|0_E\rangle = \frac{1}{\sqrt{|G|}^M} \sum_{\{h\}} |\{h\}\rangle. \qquad (56)$$

Since it corresponds to the constant wave function it is gauge-invariant, and it is the maximal entropy state in the holonomy basis. On the other hand the magnetic Hamiltonian generally has several ground states. Usually one chooses $h_B = -2\mathrm{Re}\chi$ where $\chi$ is the character of a faithful representation, so that it is minimal on the identity element only (this we prove in Appendix A). Then the ground states are those with all plaquettes equal to the identity. The simplest gauge-invariant ground state is rather easy to write in the holonomy basis, i.e.,

$$|0_B\rangle = |1, 1, \dots, 1\rangle. \qquad (57)$$

This is gauge invariant because the identity is a singlet under conjugation. This is also a state of minimal entropy in the holonomy basis. Note that the preparation of both $|0_E\rangle$ and $|0_B\rangle$ on a quantum device is standard.

In the group element basis, it has been argued [16–19] that the confining ground state of lattice Yang-Mills theory in $2+1$ dimensions (here we take a periodic square lattice) can be approximated by a trial wave function of the form,

$$\psi_\alpha(\{g\}) = \exp\left(-\alpha \sum_{\square} \mathrm{Re}\chi(g_\square)\right), \qquad (58)$$

TABLE I. Parameters for the two lattices used in the simulations for the two gauge groups $D_4$ and $Q$. All quantities happen to be either approximately or exactly the same for both groups. The labels `time`$(H_E)$, `size`$(H_E)$ and `sparsity`$(H_E)$ refer respectively to the time taken to construct $H_E$ (on a common laptop), its size and the proportion of its nonzero elements.

| Size | $V$ | $L$ | $M$ | $\dim \mathcal{H}_{\text{tot}}$ | $\dim \mathcal{H}_{\text{hol}}$ | $\dim \mathcal{H}_{\text{phys}}$ | `Time` $(H_E)$ | `Size` $(H_E)$ | `Sparsity` $(H_E)$ |
|---|---|---|---|---|---|---|---|---|---|
| $2 \times 2$ | 4 | 8 | 5 | $2^{24}$ | $2^{15}$ | $8960 \approx 2^{13}$ | 0.1 s | 13 MB | 0.15% |
| $2 \times 3$ | 6 | 12 | 7 | $2^{36}$ | $2^{21}$ | $536576 \approx 2^{19}$ | 10 s | 1.2 GB | 0.0035% |

where $\alpha$ is a variational parameter and $\chi$ is the group character which appears in the magnetic Hamiltonian [and, for relativistic theories, also in $h_E$ via Eq. (10)]. The ansatz Eq. (58) interpolates between the electric ground state ($\alpha = 0$) and the magnetic ground state ($\alpha = -\infty$). For each value of the coupling $\lambda$ of the Hamiltonian $H$ [recall Eq. (11)], $\alpha$ is determined by numerically minimizing the average energy per link,

$$E(\alpha) \equiv \frac{1}{L} \frac{\langle \psi_\alpha | H | \psi_\alpha \rangle}{\langle \psi_\alpha | \psi_\alpha \rangle}. \tag{59}$$

Due to the peculiarities of two space dimensions, this minimization is rather easy and can be performed directly in infinite volume. The details are described in Appendix B.

The trial wave function in Eq. (58) has been used in several variational calculations for Lie group gauge theories both in $2 + 1$ and higher dimensions [41–48]. It can also be systematically improved by adding larger Wilson loops, for example a $2 \times 1$ rectangular loop. Other approaches have been developed towards variational vacuum wave functionals, but they are generally more complicated and often not easily applicable to finite gauge groups. The wave function Eq. (58) is particularly appealing due to its simplicity. Note that our chosen trial wave function is based on the idea of an expansion in terms of Wilson loops, and it is possible that it might not be applicable to gauge theories where Wilson loops do not determine the gauge invariant states (as discussed in Sec. III C). However the groups considered here do not suffer from this problem.

In [49–52], Monte Carlo methods were employed to compare different proposals for ground state wave functions [all closely related to Eq. (58)] on some sets of configurations for non-Abelian Lie groups. Here we simply compute the overlap (i.e., the fidelity) between the trial wave function and the exact finite-volume ground state. Of course, due to the exponential size of the Hilbert space, we are limited to small lattices.

In what follows, we consider two different finite gauge groups of order 8. These are the dihedral group $D_4$ and the quaternion group $Q$, which are subgroups of O(2) and SU(2) respectively. In both cases, we choose the character of their unique faithful irrep for the magnetic Hamiltonian, and then set the electric Hamiltonian according to Eqs. (9) and (10). Using holonomy states, we construct their Hamiltonian on two small lattices, $2 \times 2$ and $2 \times 3$ with periodic boundary conditions. Note that, using spin network states, only a $2 \times 2$ lattice was possible [14]. The relevant parameters for the two cases are given in Table I. Since $D_4$ and $Q$ (accidentally) share the same character table, they also share the same physical subspace dimension (on any graph). Moreover, as we will soon see, despite the fact that their Hamiltonians are *not* equal, the results for the two groups are very similar.

Given each lattice, we construct the Hamiltonian in the holonomy basis following the recipe provided in Sec. III D. We can then determine some of the lowest eigenstates and compute several quantities of interest. In particular, we can compute the overlap of the true ground state with the trial wave function Eq. (58). Our code was implemented in Python compiled with Numba [53]; the Hamiltonian (and all other matrices) were implemented in sparse matrix format with SciPy [54]. In terms of constructing operators, the electric Hamiltonian was by far the slowest, and we have reported the time required to construct it (on a common laptop) in Table I. This is of course not a precise benchmark but serves to give an idea of how efficient the code can be in practice. Among the various matrices which we need to store, $H_E$ is by far the largest in terms of memory use, and again we list its size in Table I.

An important consideration is that one still needs to impose invariance under the external gauge transformation. Generally speaking, the holonomy space is a constrained Hilbert space much like the total Hilbert space. Thus they should be treated similarly. On a quantum device, one would map the holonomy Hilbert space $\mathcal{H}_{\text{hol}}$ directly onto the quantum register. On a classical device, a gauge-invariant basis could be constructed by numerically solving the equation $P|\psi\rangle = |\psi\rangle$, where $P$ is the projector onto the gauge-invariant subspace in the holonomy basis of Eq. (32). This is already advantageous compared to performing the same operation in the total Hilbert space, as the memory requirements for storing the states are substantially reduced in the holonomy basis.

However, in our case, it is sufficient to diagonalize the Hamiltonian directly in the holonomy basis. One could then apply the projector $P$, which is a cheap operation, and filter out identical states. But even this is not required in most cases. In fact suppose that $|\psi\rangle$ is a nondegenerate energy eigenstate of the Hamiltonian $H$; since the electric and magnetic Hamiltonians do not commute, this is the typical situation. Since $P$ commutes with $H$, then if $P|\psi\rangle$ is
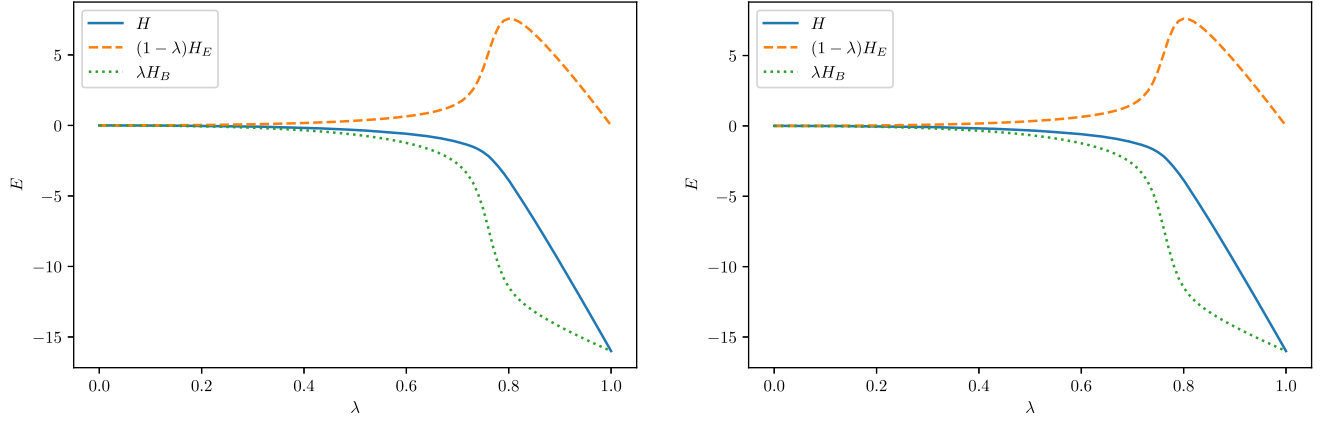
FIG. 4. Ground state energy on a $2 \times 2$ lattice for $D_4$ (*left*) and $Q$ (*right*), as well as expectation value of $H_E$ and $H_B$ in the ground state. The two plots are almost indistinguishable.

not zero, it is an energy eigenstate with the same energy and therefore $P|\psi\rangle = |\psi\rangle$ (since the state is nondegenerate and the eigenvalues of $P$ are either 0 or 1). In other words, in this case states are either gauge-invariant or not, and no superpositions occur. It is therefore quite easy to check for gauge-invariance. In particular, in all our simulations, the finite-volume ground state of the Hamiltonian was always found to be unique and gauge-invariant.

We now discuss the results of our numerical simulations. In $2 + 1$ dimensions, gauge theories with non-Abelian finite groups generally have a two-phase structure, with a confining phase for small $\lambda$ (where the electric Hamiltonian dominates) and a deconfined phase for large $\lambda$ (where the magnetic Hamiltonian dominates), separated by a first-order transition [55–59].

Figure 4 shows the ground state energy, as well as the expectation values of the electric and magnetic Hamiltonians in the ground state. It has been previously shown that the location of the phase transition may be reasonably identified with the point of sharpest change of the expectation value of the electric Hamiltonian [14]. With this prescription, we find the pseudocritical coupling $\lambda_c^{(2\times2)} = 0.76(1)$ for both groups

on the $2 \times 2$ lattice (for $D_4$, in agreement with [14]). On the $2 \times 3$ lattice the pseudocritical coupling increases slightly to $\lambda_c^{(2\times3)} = 0.77(1)$, again for both groups.

On the other hand, Fig. 5 shows the variational energy $E(\alpha)$ defined in Eq. (59) for the group $D_4$ in two cases, $\lambda = 0.5$ and $\lambda = 0.9$. In particular, $E(\alpha)$ achieves a minimum for small $\lambda$, but not for large $\lambda$, where the energy asymptotes to its minimum at $\alpha = -\infty$. Therefore the minimization procedure is well-defined only for $\lambda$ sufficiently small. This should not be thought of as a limitation of the ground state ansatz Eq. (58), but rather as a feature. Since the trial wave function defines a confining ground state, it is not expected to be valid in the deconfined phase. Thus the failure of the minimization should be taken as a *prediction* for the location of the phase transition. A similar situation was encountered in [60]. With this definition, we locate the phase transition for both groups at $\lambda_c = 0.79(1)$. Note that this value obtained via the infinite-volume trial wave function is completely independent from the one obtained from the exact finite-volume numerical data, yet quite close.

In the confining phase, we can compare the exact finite-volume ground state with the ground state ansatz Eq. (58).
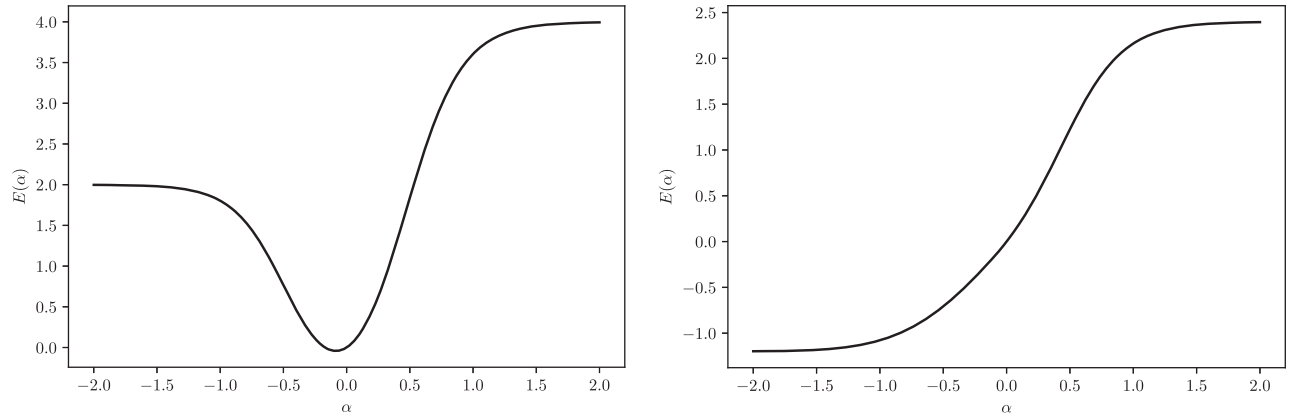


FIG. 5. Variational energy $E(\alpha)$ of Eq. (59) for $D_4$ gauge theory as a function of $\alpha$ for two values of $\lambda$, i.e., $\lambda = 0.5$ (*left*) and $\lambda = 0.9$ (*right*). For small $\lambda$, $E(\alpha)$ achieves a minimum, while for large $\lambda$ the energy asymptotes to $\alpha = -\infty$.
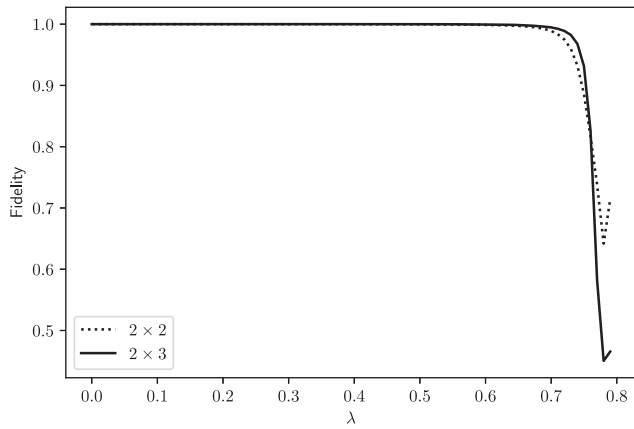
FIG. 6. Fidelity between the exact finite volume ground state and the ground state ansatz Eq. (58) for $D_4$ gauge theory on a $2 \times 2$ and $2 \times 3$ lattice, for $\lambda \in [0, 0.79]$. The picture for $Q$ is quite similar. The agreement is excellent for small $\lambda$ but drops substantially near the phase transition, especially for the larger lattice.
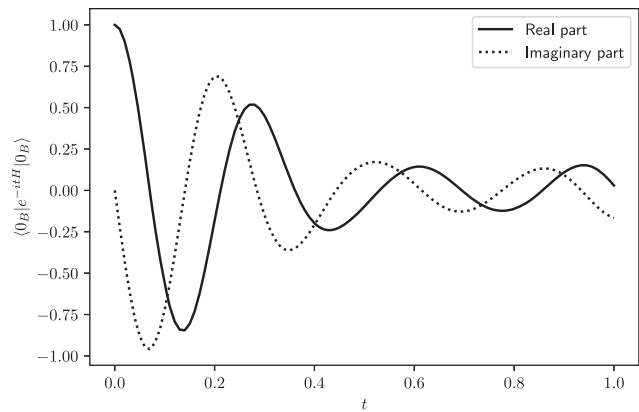


FIG. 7. Expectation value of the time-evolution operator with the full Hamiltonian $H$ at $\lambda = 0.4$ in the magnetic ground state $|0_B\rangle$ [Eq. (57)], for $D_4$ gauge theory on the $2 \times 2$ lattice.

Given two normalized states $|\psi_1\rangle$ and $|\psi_2\rangle$, a natural measure of their similarity is their *fidelity*, i.e., the absolute value squared of their overlap $|\langle \psi_1 | \psi_2 \rangle|^2$. In Fig. 6 we plot the fidelity between the exact finite-volume ground state and the trial wavefunction in the confining phase. In particular, we see that the trial wave function provides an almost perfect approximation to the exact ground state for much of the confining phase, but near the phase transition the approximation is less good. Moreover, the fidelity decreases more sharply on the larger $2 \times 3$ volume.

Overall, we find that the ansatz Eq. (58) provides a good approximation for the ground state of the two theories, especially in that it is able to predict the existence of the phase transition, as well as its location with reasonable accuracy. However, it does not fully capture the correlations near the phase transition, which is the most interesting location. Further, extrapolating from the trend in Fig. 6, one expects the fidelity near the transition to further decrease on larger volumes. It would be interesting to study the effect of adding larger loops to the trial wavefunction; we would expect them to improve its accuracy.

We also examined what happens to the 't Hooft operators defined in Sec. III E. Both $D_4$ and $Q$ have center isomorphic to $\mathbb{Z}_2$. Therefore on a periodic square lattice, one can construct two independent 't Hooft operators according to Eq. (55). Numerically, we found in all cases that the finite-volume ground state (but not excited states) is invariant under both operators for all values of $\lambda$. Finally, in Fig. 7 a quench for $D_4$ on the $2 \times 2$ lattice where the magnetic ground state is time-evolved using the full Hamiltonian $H$ at $\lambda = 0.4$. The results for $Q$ are similar.

Calculations of the kind presented here could be used as benchmarks for quantum simulators. As is clear from the trend in Table I, using our current implementation it is not possible to extend the calculation to a $3 \times 3$ lattice due to

the large memory requirements. One way to substantially reduce the size of the Hilbert space, and therefore memory requirements, is to give up periodicity, which may even be more appropriate for comparison with quantum simulators. Nonetheless, with our current implementation it should be possible to exactly diagonalize non-Abelian groups of order up to approximately 16 on a $2 \times 2$ periodic lattice.

## V. CONCLUSIONS

In this work, we have discussed how to represent the physical, gauge-invariant Hilbert space of pure lattice gauge theories in terms of a class of "holonomy states". Based on this construction, we discussed the Wilson loop representation of the gauge-invariant space as well as the topological structure of the Hilbert space. On small lattices, we could use this construction to exactly diagonalize the Hamiltonian. In particular, we computed the exact overlap between the finite-volume ground state and a conjectured ansatz and discussed to what extent they agree.

As we have seen, the holonomy basis may be useful to study small systems in classical simulations, and possibly also in current-era quantum devices. It would be interesting to understand whether the holonomy basis can be efficiently implemented on a quantum device. The crucial difficulty here is the construction of the electric Hamiltonian. It would also be interesting to study how to efficiently prepare the ground state wave function Eq. (58) on a quantum device.

As mentioned in the beginning, whether working in the total or physical Hilbert space is more advantageous may depend on what quantum device is used for the simulation. It is currently unclear what kind of hardware architecture might prevail in the long term and therefore allow quantum simulation of gauge theories at scale. It is also possible that multiple architectures could coexist for different purposes, much like CPUs and GPUs for classical computing. On an architecture where qubits are cheap, it is unlikely that

working in terms of physical variables is helpful. On the other hand, on an architecture where qubits are expensive but, for example, gates can be accurately applied, then it might be useful to work in a physical basis.

It would be interesting to further explore other properties of the holonomy basis. For example, it is unclear to what extent the properties of the holonomy states depend on the choice of maximal tree. Moreover, one may even think of applying various truncation methods directly in the holonomy basis, once most of the gauge symmetry has been removed.

The most natural next step is the inclusion of matter fields. Developing methods to compute the dimension of the physical subspace is particularly interesting in this case. For example, in the presence of a gauge anomaly there are no gauge invariant states [61].

## ACKNOWLEDGMENTS

## APPENDIX A: PROPERTIES OF FAITHFUL IRREPS

In Sec. II we defined the magnetic Hamiltonian on each plaquette to be the function $h_B \equiv -2\mathrm{Re}\chi$, where $\chi$ is the character of a faithful representation of the gauge group. Several times we have used the result that $h_B$ has a unique minimum at the identity element, a result which we now prove. The results in this section are valid for finite-dimensional representations of both finite groups as well as compact Lie groups. First we show a related result:

*Proposition 5.* Let $\rho$ be a representation of $G$ and $\chi$ its character. Then for all $g \in G$, $|\chi(g)| \leq \chi(1)$ with equality if and only if $\rho(g)$ is proportional to the identity matrix.

*Proof.* Since $G$ is either finite or compact Lie its finite-dimensional representations may be chosen to be unitary [20,21]. Let $\chi = \mathrm{tr}\rho$, where $\rho$ is a unitary representation. Then $\rho(g)$ is a unitary matrix, and as a result, its eigenvalues $\lambda_i$ are all unit-norm, i.e., $|\lambda_i| = 1$. Then by the triangle inequality,

$$|\chi(g)| = \left|\sum_i \lambda_i\right| \leq \sum_i |\lambda_i| = \dim \rho = \chi(1), \quad \text{(A1)}$$

with equality if and only if all the $\lambda_i$ are equal, in which case $\rho(g)$ is proportional to the identity matrix. ∎

Now we refine the previous statement of equality:

*Proposition 6.* If $\chi(g) = \chi(1)$, then $\rho(g) = \mathbb{1}$ is the identity matrix.

*Proof.* Since $\chi(g) = \chi(1)$, then also $|\chi(g)| = \chi(1)$ and therefore by the previous Proposition, $\rho(g) = \lambda\mathbb{1}$ for some $\lambda$ with $|\lambda| = 1$. But then

$$\chi(g) = \lambda\mathrm{tr}(\mathbb{1}) = \lambda \dim \rho = \lambda\chi(1). \quad \text{(A2)}$$

Therefore by the hypothesis $\lambda = 1$ and $\rho(g)$ is the identity matrix. ∎

Now we finally prove the required statement:

*Proposition 7.* If $\chi$ is the character of a faithful representation, then $\mathrm{Re}\chi(g)$ has a unique maximum at the identity element.

*Proof.* Using Prop. 5, we see that

$$\mathrm{Re}\chi(g) \leq |\chi(g)| \leq \chi(1) = \mathrm{Re}\chi(1). \quad \text{(A3)}$$

Therefore $\mathrm{Re}\chi(g)$ achieves its maximum at the identity element. Now suppose that $\mathrm{Re}\chi(g) = \mathrm{Re}\chi(1)$. By the above inequality, this implies that $\mathrm{Re}\chi(g) = |\chi(g)|$, so $\chi(g)$ is real and positive. Therefore we also have $\chi(g) = \chi(1)$, which, by Prop. 6, implies that $\rho(g)$ is the identity matrix (where $\rho$ is the representation with character $\chi$). But since $\rho$ is faithful by assumption, then $g = 1$ is the identity, proving uniqueness. ∎

## APPENDIX B: VARIATIONAL MINIMIZATION

As remarked in Sec. IV, in order to determine the variational parameter $\alpha$ of the ground state wave function Eq. (58) one numerically minimizes the expectation value of the Hamiltonian Eq. (59). The minimization can be performed directly in infinite volume and is quite simple due to the peculiarities of two space dimensions.

For ease of reference we define the following functions which will appear in what follows:

$$F_1(\alpha) \equiv \sum_{g \in G} \exp(-2\alpha\mathrm{Re}\chi(g)), \quad \text{(B1)}$$

$$F_2(\alpha) \equiv \sum_{g \in G} \exp(-2\alpha\mathrm{Re}\chi(g))\mathrm{Re}\chi(g), \quad \text{(B2)}$$

$$F_3(\alpha) \equiv \sum_{g_1,g_2 \in G} \sum_{k \in \Gamma} \exp[-\alpha\mathrm{Re}(\chi(g_1) + \chi(g_2) + \chi(g_1 k) + \chi(g_2 k^{-1}))]. \quad \text{(B3)}$$

The key ingredient on a two-dimensional infinite square lattice is the change of variables from links to plaquettes,

$$\frac{1}{|G|^L}\sum_{\{g\}} \longrightarrow \frac{1}{|G|^K}\sum_{\{g_\square\}}, \quad \text{(B4)}$$

where $K = L/2$ is the number of plaquettes. This is a classic result; in fact the change of variables from links to plaquettes is a lattice Bianchi identity [19,62] which is trivial in two dimensions [48,63]. From a more modern perspective, this can also be obtained from the more general change of variable formula for the holonomies Eq. (38) by choosing an appropriate maximal tree for the infinite square

lattice and then performing an appropriate change of variables from holonomies to plaquettes.

We can then compute first of all the normalization of the trial state,

$$
\begin{aligned}
\langle \psi_\alpha | \psi_\alpha \rangle &= \sum_{\{g\}} \exp\left(-2\alpha \sum_\Box \mathrm{Re}\chi(g_\Box)\right) \\
&= \frac{|G|^L}{|G|^K} \sum_{\{g_\Box\}} \exp\left(-2\alpha \sum_\Box \mathrm{Re}\chi(g_\Box)\right) \\
&= \frac{|G|^L}{|G|^K} F_1(\alpha)^K,
\end{aligned}
\tag{B5}
$$

since we can now sum each plaquette independently. A very similar calculation leads to

$$
\langle \psi_\alpha | H_B | \psi_\alpha \rangle = -2K \frac{|G|^L}{|G|^K} F_1(\alpha)^{K-1} F_2(\alpha).
\tag{B6}
$$

For the electric Hamiltonian, one needs to consider the two plaquettes hinging on each link, and then one finds

$$
\langle \psi_\alpha | h_E | \psi_\alpha \rangle = |\Gamma| \langle \psi | \psi \rangle - \frac{|G|^L}{|G|^K} F_1(\alpha)^{K-2} F_3(\alpha).
\tag{B7}
$$

Therefore overall the target function to minimize is

$$
E(\alpha) \equiv \frac{1}{L} \frac{\langle \psi_\alpha | H | \psi_\alpha \rangle}{\langle \psi_\alpha | \psi_\alpha \rangle} = (1-\lambda)\left[|\Gamma| - \frac{F_3(\alpha)}{F_1(\alpha)^2}\right] - \lambda \frac{F_2(\alpha)}{F_1(\alpha)}.
\tag{B8}
$$

Since the $F$ functions are rather simple to compute numerically, for each $\lambda$ it is easy to minimize $E$ in order to find the optimal $\alpha$.

[1] J. Kogut and L. Susskind, Hamiltonian formulation of Wilson's lattice gauge theories, Phys. Rev. D **11**, 395 (1975).

[2] P. Hasenfratz and F. Niedermayer, Unexpected results in asymptotically free quantum field theories, Nucl. Phys. **B596**, 481 (2001).

[3] E. Zohar and M. Burrello, Formulation of lattice gauge theories for quantum simulations, Phys. Rev. D **91**, 054506 (2015).

[4] S. Notarnicola, E. Ercolessi, P. Facchi, G. Marmo, S. Pascazio, and F. V. Pepe, Discrete abelian gauge theories for quantum simulations of QED, J. Phys. A **48**, 30FT01 (2015).

[5] E. Ercolessi, P. Facchi, G. Magnifico, S. Pascazio, and F. V. Pepe, Phase transitions in $Z_n$ gauge models: Towards quantum simulations of the Schwinger-Weyl QED, Phys. Rev. D **98**, 074503 (2018).

[6] A. Alexandru, P. F. Bedaque, R. Brett, and H. Lamm, Spectrum of digitized QCD: Glueballs in a S(1080) gauge theory, Phys. Rev. D **105**, 114508 (2022).

[7] K. Liu, J. Nissinen, R.-J. Slager, K. Wu, and J. Zaanen, Generalized liquid crystals: Giant fluctuations and the vestigial chiral order of I, O and T matter, Phys. Rev. X **6**, 041025 (2016).

[8] A. J. Beekman, J. Nissinen, K. Wu, K. Liu, R.-J. Slager, Z. Nussinov, V. Cvetkovic, and J. Zaanen, Dual gauge field theory of quantum liquid crystals in two dimensions, Phys. Rep. **683**, 1 (2017).

[9] A. J. Beekman, J. Nissinen, K. Wu, and J. Zaanen, Dual gauge field theory of quantum liquid crystals in three dimensions, Phys. Rev. B **96** (2017).

[10] K. Liu, J. Nissinen, J. de Boer, R.-J. Slager, and J. Zaanen, Hierarchy of orientational phases and axial anisotropies in the gauge theoretical description of generalized nematic liquid crystals, Phys. Rev. E **95**, 022704 (2017).

[11] N. Manjunath and M. Barkeshli, Crystalline gauge fields and quantized discrete geometric response for Abelian topological phases with lattice symmetry, Phys. Rev. Res. **3**, 013040 (2021).

[12] D. Harlow and H. Ooguri, Symmetries in quantum field theory and quantum gravity, Commun. Math. Phys. **383**, 1669 (2021).

[13] O. Friedrich, A. Singh, and O. Doré, Toolkit for scalar fields in universes with finite-dimensional Hilbert space, Classical Quantum Gravity **39**, 235012 (2022).

[14] A. Mariani, S. Pradhan, and E. Ercolessi, Hamiltonians and gauge-invariant Hilbert space for lattice Yang-Mills-like theories with finite gauge group, Phys. Rev. D **107**, 114513 (2023).

[15] B. Durhuus, On the structure of gauge invariant classical observables in lattice gauge theories, Lett. Math. Phys. **4**, 515 (1980).

[16] R. P. Feynman, The qualitative behavior of Yang-Mills theory in $(2+1)$-dimensions, Nucl. Phys. **B188**, 479 (1981).

[17] J. Greensite, Calculation of the Yang-Mills vacuum wave functional, Nucl. Phys. **B158**, 469 (1979).

[18] J. Greensite, Large-scale vacuum structure and new calculational techniques in lattice SU(N) gauge theory, Nucl. Phys. **B166**, 113 (1980).

[19] M. Halpern, Field strength and dual variable formulations of gauge theory, Phys. Rev. D **19**, 517 (1979).

[20] J.-P. Serre, *Linear Representations of Finite Groups*, Graduate Texts in Mathematics (Springer, New York, NY, 1967).

[21] A. W. Knapp, *Lie Groups: Beyond an Introduction*, Progress in Mathematics (Birkhäuser, Boston, 1996).

[22] A. Milsted and T. J. Osborne, Quantum Yang-Mills theory: An overview of a program, Phys. Rev. D **98** (2018).

[23] D. Tong, Gauge Theory (2018), http://www.damtp.cam.ac.uk/user/tong/gaugetheory.html.

[24] P. Orland, Duality for non-Abelian lattice fields, Nucl. Phys. **B163**, 275 (1980).

[25] S. Caspar, D. Mesterházy, T. Olesen, N. Vlasii, and U.-J. Wiese, Doubled lattice Chern–Simons–Yang–Mills theories with discrete gauge group, Ann. Phys. (Amsterdam) **374**, 255 (2016).

[26] J. C. Baez, Spin networks in gauge theory, Adv. Math. **117**, 253 (1996).

[27] G. Burgio, R. D. Pietri, H. Morales-Técotl, L. Urrutia, and J. Vergara, The basis of the physical Hilbert space of lattice gauge theories, Nucl. Phys. **B566**, 547 (2000).

[28] M. Creutz, Gauge fixing, the transfer matrix, and confinement on a lattice, Phys. Rev. D **15**, 1128 (1977).

[29] L. R. Foulds, *Graph Theory Applications* (Springer, New York, 1992).

[30] S. X. Cui, D. Ding, X. Han, G. Penington, D. Ranard, B. C. Rayhaun, and Z. Shangnan, Kitaev's quantum double model as an error correcting code, Quantum **4**, 331 (2020).

[31] A. Mariani, Finite-group Yang-Mills lattice gauge theories in the Hamiltonian formalism (2020), http://amslaurea.unibo.it/21183/.

[32] N. Watson, Solution of the SU(2) Mandelstam constraints, Phys. Lett. B **323**, 385 (1994).

[33] A. Sengupta, Gauge invariant functions of connections, Proc. Am. Math. Soc. **121**, 897 (1994).

[34] T. Lévy, Wilson loops in the light of spin networks, J. Geom. Phys. **52**, 382 (2004).

[35] C.-H. Sah, Automorphisms of finite groups, J. Algebra **10**, 47 (1968).

[36] P. Brooksbank and M. Mizuhara, On groups with a class-preserving outer automorphism, Involve J. Math. **7**, 171 (2013).

[37] M. Nguyen, Y. Tanizaki, and M. Ünsal, Semi-Abelian gauge theories, non-invertible symmetries, and string tensions beyond N-ality, J. High Energy Phys. 03 (2021) 238.

[38] B. Heidenreich, J. McNamara, M. Montero, M. Reece, T. Rudelius, and I. Valenzuela, Non-invertible global symmetries and completeness of the spectrum, J. High Energy Phys. 09 (2021) 203.

[39] A. Kitaev, Fault-tolerant quantum computation by anyons, Ann. Phys. (Amsterdam) **303**, 2 (2003).

[40] S. Pradhan, A. Maroncelli, and E. Ercolessi, Discrete Abelian lattice gauge theories on a ladder and their dualities with quantum clock models, Phys. Rev. B **109**, 064410 (2024).

[41] H. Arisue, M. Kato, and T. Fujiwara, Variational study of vacuum wave function for lattice gauge theory in $2+1$ dimension, Prog. Theor. Phys. **70**, 229 (1983).

[42] H. Arisue, Variational investigation of the mass spectrum in $2+1$ dimensional SU(2) lattice gauge theory, Prog. Theor. Phys. **84**, 951 (1990).

[43] E. Dagotto and A. Moreo, Hamiltonian variational study of SU(2) lattice gauge theory, Phys. Rev. D **29**, 2350 (1984).

[44] A. Dabringhaus, M. L. Ristig, and J. W. Clark, Vacuum ground and excited states of the U(1) lattice gauge Hamiltonian, Phys. Rev. D **43**, 1978 (1991).

[45] S.-H. Guo, Q.-Z. Chen, and L. Li, Analytic calculation of the vacuum wave function for $(2+1)$-dimensional SU(2) lattice gauge theory, Phys. Rev. D **49**, 507 (1994).

[46] J. McIntosh and L. Hollenberg, Reaching the continuum limit in lattice gauge theory—without a computer, Phys. Lett. B **538**, 207 (2002).

[47] J. A. McIntosh and L. C. Hollenberg, Analytic calculation of the mass gap in $U(1)$ $2+1$ lattice gauge theory., Nucl. Phys. B, Proc. Suppl. **106–107**, 257 (2002).

[48] J. Carlsson, Improvement and analytic techniques in Hamiltonian lattice gauge theory, Ph.D. thesis, University of Melbourne, 2003.

[49] J. Greensite, H. Matevosyan, S. Olejník, M. Quandt, H. Reinhardt, and A. P. Szczepaniak, Testing proposals for the Yang-Mills vacuum wavefunctional by measurement of the vacuum, Phys. Rev. D **83**, 114509 (2011).

[50] J. Greensite, Computer measurement of the Yang-Mills vacuum wavefunctional in three dimensions, Phys. Lett. B **191**, 431 (1987).

[51] J. Greensite and J. Iwasaki, Monte Carlo study of the Yang-Mills vacuum wavefunctional in $D=4$ dimensions, Phys. Lett. B **223**, 207 (1989).

[52] H. Arisue, Monte Carlo measurement of the vacuum wavefunction for non-Abelian gauge theory in $D=3$ dimensions, Phys. Lett. B **280**, 85 (1992).

[53] S. K. Lam, A. Pitrou, and S. Seibert, NUMBA: A llvm-based PYTHON jit compiler, in *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC* (2015), pp. 1–6, 10.1145/2833157.2833162.

[54] P. Virtanen *et al.* (SciPy 1.0 Contributors), SciPy1.0: Fundamental Algorithms for Scientific Computing in Python, Nat. Methods **17**, 261 (2020).

[55] C. Rebbi, Phase structure of non-Abelian lattice gauge theories, Phys. Rev. D **21**, 3350 (1980).

[56] G. Bhanot and M. Creutz, Phase diagram of $Z(N)$ and $U(1)$ gauge theories in three dimensions, Phys. Rev. D **21**, 2892 (1980).

[57] G. Bhanot and C. Rebbi, Monte Carlo simulations of lattice models with finite subgroups of SU(3) as gauge groups, Phys. Rev. D **24**, 3319 (1981).

[58] D. Petcher and D. H. Weingarten, Monte Carlo calculations and a model of the phase structure for gauge theories on discrete subgroups of SU(2), Phys. Rev. D **22**, 2465 (1980).

[59] M. Creutz and M. Okawa, Generalized actions in $Z_p$ lattice gauge theory, Nucl. Phys. **B220**, 149 (1983).

[60] D. Heys and D. Stump, Evidence of the phase transition of the compact U(1) lattice gauge theory from a variational calculation, Nucl. Phys. **B257**, 19 (1985).

[61] E. Witten, An SU(2) anomaly, Phys. Lett. **117B**, 324 (1982).

[62] G. G. Batrouni, Plaquette formulation and the Bianchi identity for lattice gauge theories, Nucl. Phys. **B208**, 467 (1982).

[63] N. Ligterink, N. Walet, and R. Bishop, Toward a many-body treatment of Hamiltonian lattice SU(N) gauge theory, Ann. Phys. (N.Y.) **284**, 215 (2000).