Solvable lattice Hamiltonians with fractional Hall conductivity

Zhaoyu Han[1] and Jing-Yuan Chen [2]

[1]*Department of Physics, Stanford University, Stanford, California 94305, USA*
[2]*Institute for Advanced Study, Tsinghua University, Beijing 100084, China*

We construct a class of lattice Hamiltonians that exhibit fractional Hall conductivity. These Hamiltonians, while not being exactly solvable, can be controllably solved in their low-energy sectors, through a combination of perturbative and exact techniques. Our construction demonstrates a systematic way to circumvent the Kapustin-Fidkowski no-go theorem and is generalizable.

I. INTRODUCTION

The topological states of matter have been an important theme in condensed matter physics since the discovery of the quantum Hall effect four decades ago [1–3]. Over time, as people's knowledge in this field broadened and understandings deepened, the construction of exactly solvable lattice models emerged as a promising approach to study a large variety of topological states as well as their interplays with global symmetries [4–6]. The success of the exactly solvable lattice model approach is at least threefold: (1) these models provide unambiguous microscopic completions for the associated topological phases, many of which hypothesized on theoretical grounds, and demonstrate that they can be realized in solid state systems at least in principle; (2) all the interesting topological properties can be exactly solved for and thereby apprehended in an explicit manner; (3) these models are constructed not out of fortuity but out of systematic considerations under certain principles, and the physical and mathematical origin of those principles is profound [5–9].

In condensed matter physics, one most important global symmetry a topological state may interplay ("be enriched") with is the electromagnetic U(1). Indeed, the quantum Hall effect, whose very advent opened the entire field, was primarily characterized by the integrally or fractionally quantized Hall conductivity, featuring such an interplay. A rather curious situation occurs, however, if we attempt to construct exactly solvable lattice Hamiltonians, applying the usual ingredients and wisdom that have generated so many successes, for the topological states that exhibit nontrivial Hall conductivity: such efforts never came to fruition. The fundamental reason behind this apparently disappointing situation was finally addressed by Kapustin and Fidkowski, stated by them in a no-go theorem [10]: *A nontrivial Hall conductivity is impossible in a gapped local commuting projector Hamiltonian with finite-dimensional local Hilbert space.*

The purpose of this work is to go beyond the usual approach of exactly solvable lattice models, and in particular beyond the limiting conditions stated in the Kapustin-Fidkowski no-go theorem, in order to construct lattice Hamiltonians that

can be solved, at least in the low-energy sectors, to find nontrivial Hall conductivity.

In Ref. [11] one of us studied the electromagnetic U(1) enrichment for a large class of Abelian topological phases [those which admit gapped boundary conditions were the electromagnetic U(1) absent [12,13]], including the cases with Hall conductivity, and found exactly solvable Lagrangians for them on *effective* space-time lattices (i.e., coarse-grained space-time manifolds). There it was also explained how and why, if we proceed with the usual wisdom [5,8] in an attempt to reduce an exactly solvable Lagrangian on a coarse-grained space-time to an exactly solvable toy model Hamiltonian on an actual spatial lattice, we would run into problems when the Hall conductivity is nontrivial. The occurrence of the problems is an embodiment of the Kapustin-Fidkowski no-go theorem.

Through the study in Ref. [11], however, a path was sketched towards constructing controllably solvable, despite not being exactly solvable, lattice Hamiltonians with nontrivial Hall conductivity. In this paper we elaborate on the simplest such cases and show that the Hamiltonians can indeed be solved, exhibiting the desired values of Hall conductivity which are fractional in general. The solution is obtained by a combination of perturbation theory and the exact techniques familiar in solving local commuting projector Hamiltonians. (We remark that a well-known example of Hamiltonian solved by a combination of perturbative and exact techniques is the celebrated Kitaev honeycomb model [14] which is a non-Abelian spin liquid, although the detailed procedure is very different from our present work.) Our work is conceptually straightforward and can be generalized to a larger class [11] of topological phases coupled to electromagnetic U(1), including the fermionic phases; we will elaborate on those generalizations in subsequent works.

This paper is organized as the following. In Sec. II we motivate the lattice Hamiltonians following the idea outlined in [11]; we also explain the relation between our present work and the previous literature, in particular Refs. [15,16] and [17]. In Sec. III A we use perturbation theory to obtain the low-energy subspace within each local Hilbert space.

In Sec. III B we solve for the many-body ground state(s) and anyon excitation states using steps similar to those in solving local commuting projector Hamiltonians. In Sec. IV we compute the Hall conductivity, and show it is precisely quantized to the values that we designated, despite that we have apparently used perturbation theory in the derivation. In Sec. V we make concluding remarks.

II. HAMILTONIAN

In this paper we focus on the simplest cases among the bosonic topological orders enriched by electromagnetic U(1) considered in [11]. In the continuum they are described by a class of doubled Chern-Simons (BF) theories

$$S = \int d^3x \left[\frac{n}{2\pi} a db - \frac{q}{2\pi} A da - \frac{p}{2\pi} A db \right] \quad (1)$$

with n, p, q being integers. Here a, b are dynamical U(1) gauge fields and A is the electromagnetic U(1) background. When the coupling to A is absent, the theory can be reduced to the $\mathbb{Z}_n$ generalization of the toric code (see below), therefore, n determines the intrinsic topological order, with $n = 1$ topologically trivial [18]. On the other hand, p, q determine the U(1) global symmetry enrichment. [In principle, a, b can couple to two different U(1) global symmetry backgrounds A, B, but in this work we will identify $B = A$.] This model can be viewed as certain "double-layered" bosonic quantum Hall. Its Hall conductivity is $-2pq/n$, and it supports two types of anyons, one coupled to a and the other to b, that have trivial self-statistics and $-2\pi/n$ mutual statistics, and carry electric charges p/n and q/n, respectively [19].

In [11] it was shown that such topological orders along with the electromagnetic U(1) enrichment admit effective Lagrangians on space-time lattice, that retain all the formal properties and are exactly solvable. By viewing one direction of the space-time lattice as the discretized time, through standard procedure, a space-time lattice Lagrangian gives rise to a set of commutation relations, as well as a spatial lattice Hamiltonian were it in the usual cases; in our case, however, related to the fact the Lagrangian is exactly solvable, instead of a Hamiltonian it turns out we actually get a set of strict constraints on the Hilbert space. While a set of constraints on the Hilbert space is not what we hope for, we can use them to motivate a lattice Hamiltonian, whose low-energy sector reproduces and therefore replaces those constraints. The Hamiltonian thereby motivated may not be exactly solvable any more, but as long as its low-energy sector can be controllably solved for, we have achieved our goal. The motivating procedure is reviewed in details in Appendix A. Here we will present the results: the lattice Hilbert space, the commutation relations, and the lattice Hamiltonians for general integers n, p, q.

We suppose the spatial lattice is a square lattice for simplicity (we can generalize all of our results to a triangulation with branching structure of an arbitrary two-dimensional spatial manifold). The local Hilbert spaces and their endowed operators are as follows (see Fig. 1 for illustration):

(i) On each link **l**, which is equivalent to a dual lattice link **l*** (with direction 90° counterclockwise to the direction of **l**), there is a local Hilbert space endowed with a conjugate pair
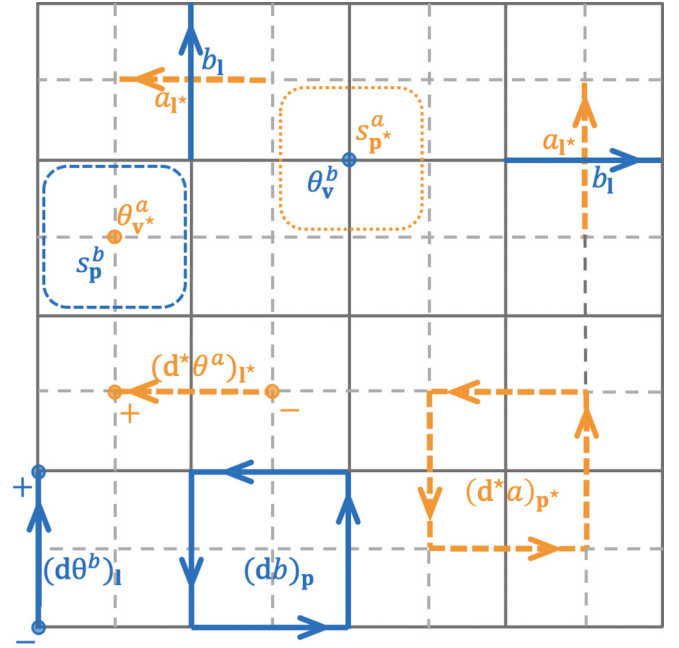

FIG. 1. Illustration for the lattice (solid line) and the dual lattice (dashed line), the conjugate pairs of operators, and the lattice and dual lattice "exterior derivatives." Note we have picked the natural directions of the lattice links to be $+\hat{\mathbf{x}}, +\hat{\mathbf{y}}$, and the natural directions of the dual lattice links to be 90° counterclockwise to the associated lattice link directions. When the direction label on a link (or a dual link) reverses, the associated operator picks a negative sign.

of real-valued operators satisfying

$$[b_{\mathbf{l}}, a_{\mathbf{l}^\star}] = i \frac{2\pi}{n}. \quad (2)$$

(ii) On each vertex **v**, which is equivalent to a dual lattice plaquette **p***, there is a local Hilbert space endowed with a conjugate pair of integer/U(1)-valued operators satisfying

$$\left[s_{\mathbf{p}^\star}^a, e^{i\theta_{\mathbf{v}}^b} \right] = e^{i\theta_{\mathbf{v}}^b}. \quad (3)$$

(iii) On each plaquette **p**, which is equivalent to a dual lattice vertex **v***, there is a local Hilbert space endowed with a conjugate pair of integer/U(1)-valued operators satisfying

$$\left[s_{\mathbf{p}}^b, e^{i\theta_{\mathbf{v}^\star}^a} \right] = e^{i\theta_{\mathbf{v}^\star}^a}. \quad (4)$$

To have an intuitive picture in mind (which will appear to be an emergent picture that is not exact, after we introduce our Hamiltonian later), one may think of a, b as dynamical $\mathbb{R}$ gauge fields on the lattice, such that they are effectively reduced to dynamical U(1) gauge fields upon the introduction of the associated dynamical Dirac string variables s^a and s^b. The "reduction to U(1)" can be understood in the following sense. The effective U(1) fluxes associated with the dynamical gauge fields a, b are

$$f_{\mathbf{p}}^b \equiv (db - 2\pi s^b)_{\mathbf{p}}, \quad f_{\mathbf{p}^\star}^a \equiv (d^\star a - 2\pi s^a)_{\mathbf{p}^\star}, \quad (5)$$

where the lattice curl db and the dual lattice curl $d^\star a$ are illustrated in Fig. 1. Thanks to the Dirac string variables, if we sum $f_{\mathbf{p}}^b$ or $f_{\mathbf{p}^\star}^a$ over all the plaquettes **p** or **p*** on a closed space (consider a square lattice plane with periodic boundary conditions, forming a torus), we can have arbitrary

$2\pi\mathbb{Z}$ values, reproducing the Dirac quantization condition for U(1) gauge fluxes. Moreover, we can see that the f^a, f^b fluxes are invariant under

$$b_{\mathbf{l}} \rightarrow b_{\mathbf{l}} + 2\pi z^b_{\mathbf{l}}, \quad s^b_{\mathbf{p}} \rightarrow s^b_{\mathbf{p}} + (dz^b)_{\mathbf{p}},$$
$$a_{\mathbf{l}^\star} \rightarrow a_{\mathbf{l}^\star} + 2\pi z^a_{\mathbf{l}^\star}, \quad s^a_{\mathbf{p}^\star} \rightarrow s^a_{\mathbf{p}^\star} + (d^\star z^a)_{\mathbf{p}^\star}, \quad (6)$$

where $z^b_{\mathbf{l}}$, $z^a_{\mathbf{l}^\star}$ are arbitrary $\mathbb{Z}$-valued transformations on the links and dual links, respectively. Such invariances manifest the fact that a, b are effectively reduced from $\mathbb{R}$ to $\mathbb{R}/2\pi\mathbb{Z} =$ U(1) gauge fields; they are known as 1-form $\mathbb{Z}$ gauge invariances {see [11] and Appendix A) (a general introduction to higher form symmetries can be found in [20]). Let us define the generators

$$g^a_{\mathbf{l}^\star} \equiv e^{i(d^\star\theta^a - na)_{\mathbf{l}^\star}}, \quad g^b_{\mathbf{l}} \equiv e^{i(d\theta^b - nb)_{\mathbf{l}}}, \quad (7)$$

where the notions of $d\theta^b$ and $d^\star\theta^a$ are illustrated in Fig. 1. Then the transformation $z^b_{\mathbf{l}}$ in (6) is generated by conjugating $b_{\mathbf{l}}$ with $g^a_{\mathbf{l}^\star}$ whose $\mathbf{l}^\star$ is dual to $\mathbf{l}$ is the aforementioned manner (actually $\mathbf{l}^\star$ is right on top of $\mathbf{l}$, so we may therefore say $\mathbf{l}^\star = \mathbf{l}$); likewise, the transformation $z^a_{\mathbf{l}^\star}$ is generated by conjugating $a_{\mathbf{l}^\star}$ with $g^b_{\mathbf{l}}$. The expressions of g^a, g^b invite us to think of θ^a, θ^b as the "superconducting phases" (with charge n) associated with the effectively U(1) dynamical gauge fields a, b, respectively. Indeed, the expressions of g^a, g^b are invariant under the ordinary (0-form) gauge transformations

$$b_{\mathbf{l}} \rightarrow b_{\mathbf{l}} + (d\varphi^b)_{\mathbf{l}}, \quad \theta^b_{\mathbf{v}} \rightarrow \theta^b_{\mathbf{v}} + n\varphi^b_{\mathbf{v}},$$
$$a_{\mathbf{l}^\star} \rightarrow a_{\mathbf{l}^\star} + (d^\star\varphi^a)_{\mathbf{l}^\star}, \quad \theta^a_{\mathbf{v}^\star} \rightarrow \theta^a_{\mathbf{v}^\star} + n\varphi^a_{\mathbf{v}^\star}, \quad (8)$$

where $\varphi^b_{\mathbf{v}}$, $\varphi^a_{\mathbf{v}^\star}$ are arbitrary U(1) transformations on the vertices and dual vertices, respectively; they are effectively U(1) instead of $\mathbb{R}$ because any $2\pi\mathbb{Z}$ part of them can be completely absorbed into (6). The two transformations by $\varphi^b_{\mathbf{v}}$, $\varphi^a_{\mathbf{v}^\star}$ are, in turn, generated by the commutators with $f^a_{\mathbf{p}^\star=\mathbf{v}}$ and $f^b_{\mathbf{p}=\mathbf{v}^\star}$ respectively. In fact, it is easy to check that the generators g^a, g^b, f^a, f^b for the gauge transformations (6) and (8) all commute with each other [so in particular their own expressions are invariant under (6) and (8)]; if a state is a simultaneous eigenstate of all of g^a, g^b, f^a, f^b, we may view the state as respecting the respective Gauss's law constraints with the "gauge charges" given by the simultaneous eigenvalues.

Having introduced the local Hilbert space and operators, we now introduce the lattice Hamiltonian. Before we present our final Hamiltonian, it turns out to be helpful to first consider a simpler "prototype" Hamiltonian, and understand its properties and problems. Through the aforementioned effective Lagrangian and treatments that are detailed in Appendix A, we are led to consider the "prototype" Hamiltonian

$$\widetilde{H} = \frac{U_b}{2}\sum_{\mathbf{l}}\left|1 - e^{iqA_{\mathbf{l}}}g^b_{\mathbf{l}}\right|^2 + \frac{U_a}{2}\sum_{\mathbf{l}^\star}\left|1 - e^{ipA_{\mathbf{l}^\star}^\star}g^a_{\mathbf{l}^\star}\right|^2$$
$$+ \frac{V_b}{2}\sum_{\mathbf{p}}\left(f^b_{\mathbf{p}}\right)^2 + \frac{V_a}{2}\sum_{\mathbf{p}^\star}\left(f^a_{\mathbf{p}^\star}\right)^2, \quad (9)$$

where $A_{\mathbf{l}}$ is the electromagnetic U(1) background field living on the link $\mathbf{l}$, while $A_{\mathbf{l}^\star}^\star$ on the dual link $\mathbf{l}^\star$ is identified with a nearby $A_{\mathbf{l}}$ pointing in the same direction (see Fig. 2 for a choice of identification). The coupling of the electromagnetic
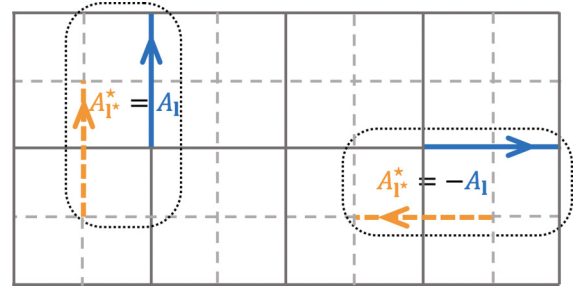

FIG. 2. A choice of identification of $A_{\mathbf{l}^\star}^\star$ with $A_{\mathbf{l}}$, where $\mathbf{l}$ is chosen to locate at $\hat{\mathbf{x}}/2 + \hat{\mathbf{y}}/2$ away from $\mathbf{l}^\star$.

field A into the system can be intuitively understood as the following. If we think of $s^b_{\mathbf{p}^\star}$ as a boson number operator on the vertex $\mathbf{v} = \mathbf{p}^\star$, then $e^{\pm i\theta^b_{\mathbf{v}}}$ is the creation/annihilation operator of the boson, and hence $g^b_{\mathbf{l}}$ involves the hopping of such a boson across the link $\mathbf{l}$; if the boson carries electric charge q, its hopping will indeed couple to A through the factor $e^{iqA_{\mathbf{l}}}$, likewise for the term with $A_{\mathbf{l}^\star}^\star$. Therefore, the local electric charge operator on a vertex $\mathbf{v}$ is given by

$$\rho_{\mathbf{v}} \equiv q\,s^a_{\mathbf{p}^\star=\mathbf{v}} + p\,s^b_{\mathbf{p}=\mathbf{v}-\hat{\mathbf{x}}/2-\hat{\mathbf{y}}/2}, \quad (10)$$

where the second term is due to the said identification between $A_{\mathbf{l}^\star}^\star$ and $A_{\mathbf{l}}$.

Let us examine the properties of $\widetilde{H}$. Since g^a, g^b, f^a, f^b commute with each other, all of the four terms in $\widetilde{H}$ commute with each other and can be simultaneously diagonalized; as mentioned before, the simultaneous diagonalization may be viewed as imposing the Gauss's laws for their respective gauge transformations. But, the four terms cannot be simultaneously minimized in general. To minimize the U_b term, we need an eigenstate such that

$$nb_{\mathbf{l}} = qA_{\mathbf{l}} + (d\theta^b)_{\mathbf{l}} \mod 2\pi, \quad (11)$$

but then in the V_b term

$$f^b_{\mathbf{p}} = \frac{q}{n}(dA)_{\mathbf{p}} \mod \frac{2\pi}{n} \quad (12)$$

which is nonzero in general, and hence the V_b term is not minimized; in particular, for small magnetic field dA, if the U term is already minimized by the state, $f^b_{\mathbf{p}} = (q/n)(dA)_{\mathbf{p}}$ is the choice that minimizes the V term the best. Likewise for the U_a term and the V_a term, with $q \rightarrow p$ and $A \rightarrow A^\star$. Suppose U_b/V_b, $U_a/V_a \rightarrow \infty$ so that the U_b, U_a terms are minimized first, and suppose the magnetic field dA is indeed small. Then, the ground state may be viewed as considering all the eigenstates in the b, θ^b, s_b basis that first minimizes the U_b term and then the V_b term, and then taking a suitable linear superposition of all such states according to minimizing the U_a term and then the V_a term, in a spirit similar to solving for the toric code ground state(s) [4]; clearly, one can also exchange the views between b and a.

Some connection to a $-2pq/n$ Hall conductivity can be readily noted. From the definitions of f^b, f^a and ρ, we find

the ground-state expectation value of the local charge:

$$
\begin{aligned}
\langle \rho_{\mathbf{v}} \rangle &= -\frac{q}{2\pi} \langle f^a - d^{\star} a \rangle_{\mathbf{p}^{\star}} - \frac{p}{2\pi} \langle f^b - db \rangle_{\mathbf{p}} \\
&= -\frac{pq}{2\pi n} [(dA)_{\mathbf{p}'} + (dA)_{\mathbf{p}}] \\
&\quad + \frac{q}{2\pi} \langle (d^{\star} a)_{\mathbf{p}^{\star}} \rangle + \frac{p}{2\pi} \langle (db)_{\mathbf{p}} \rangle,
\end{aligned}
\tag{13}
$$

where $\mathbf{p}^{\star} = \mathbf{v}$, $\mathbf{p} = \mathbf{v} - \hat{\mathbf{x}}/2 - \hat{\mathbf{y}}/2$, $\mathbf{p}' = \mathbf{v} + \hat{\mathbf{x}}/2 + \hat{\mathbf{y}}/2$. Since $(dA)_{\mathbf{p}}$ and $(dA)_{p'}$ are nearby magnetic fields, this result already has the flavor of $\langle \delta \rho \rangle = \sigma_H dA$ with the Hall conductivity $2\pi \sigma_H = -2pq/n$, if we can show that $\langle (d^{\star} a)_{\mathbf{p}^{\star}} \rangle$ and $\langle (db)_{\mathbf{p}} \rangle$ at ground state are somehow independent of the external magnetic field.

At this point we are ready to see the problem with our "prototype" Hamiltonian $\tilde{H}$. The problem manifests itself as twofolded:

(i) Since g^a, g^b, f^a, f^b commute with each other and their eigenvalues are all continuous, the Hamiltonian is gapless, rather than a gapped one that we want; in fact, the Hamiltonian is locally gapless, not just becoming gapless only after taking the limit of large system size. Related to this, the ground state and hence the response to the magnetic field are sensitive to the ratios U_b/V_b and U_a/V_a (which were taken towards infinite in the above), rather than being robustly fractionally quantized.

(ii) To compute the Hall conductivity, we need to show that the ground-state local expectation values $\langle (d^{\star} a)_{\mathbf{p}^{\star}} \rangle$ and $\langle (db)_{\mathbf{p}} \rangle$ are somehow independent of the external magnetic field. But, with a little extra effort it is easy to see these expectation values are ambiguous, somewhat like asking for the expectation $\langle x \rangle$ for a Bloch wave. The ambiguity arises from the infinite ranges of the local values of a, b involved in the superposition.

These two issues actually represent the same problem: since a, b are canonical conjugates, the local gaplessness of a is related to the local unboundedness of b, and vice versa.

Our task is therefore to resolve this problem of $\tilde{H}$, making a modified Hamiltonian gapped and its Hall conductivity unambiguous. This is the point at which the Kapustin-Fidkowski no-go theorem [10] makes its manifestation. When the Hall conductivity vanishes, i.e., when $p = 0$ or $q = 0$, there is a known resolution that maintains the local commutativity of the terms in the Hamiltonian, making it exactly solvable; the model was introduced in [21] (as part of a larger Hamiltonian), and revisited in our present framework in [11] (see Appendix A). Roughly speaking, the resolution is to note that when $q = 0$ (the case when $p = 0$ is analogous), the condition (11) is independent of the external field, and therefore, rather than viewing it as an energetic condition, we may instead, from the very beginning, let the full local Hilbert space on the link to take a finite set of discrete values $\tilde{b}_{\mathbf{l}} = \{0, 1, \ldots, n - 1\}$ [which may be viewed as a reminiscence of $(nb - d\theta^b)_{\mathbf{l}}/2\pi \bmod n$ under the condition (11)]; now that the local Hilbert space on a link is discrete and bounded, one can construct gapped local commuting terms that essentially play the roles of the V_b, U_a, V_b terms in $\tilde{H}$, and the commutativity leads to solvability. If both $p, q = 0$, the model can be further reduced to the $\mathbb{Z}_n$ generalization of the toric code [4]. Such a model becomes unavailable when the Hall conductivity is

$-2pq/n \neq 0$, i.e., when $p, q \neq 0$, since the condition (11) and its $a_{\mathbf{l}^{\star}}$ analog both depend on the external field and can no longer be realized by any fixed discrete local Hilbert space to begin with. Thus, for general values of p, q, we will resort to our resolution below, which makes the Hamiltonian not exactly solvable, but fortunately perturbatively solvable at its lower-energy sectors.

Our resolution is to add simple, noncommuting terms $\epsilon_a a_{\mathbf{l}^{\star}}^2/2$ and $\epsilon_b b_{\mathbf{l}}^2/2$ to the "prototype" Hamiltonian $\tilde{H}$, to open the local gaps and also to softly bound the ranges of the local Hilbert spaces. The Hamiltonian reads as

$$
\begin{aligned}
H = \sum_{\mathbf{l}=\mathbf{l}^{\star}} & \left[\frac{\epsilon_a}{2} a_{\mathbf{l}^{\star}}^2 + \frac{\epsilon_b}{2} b_{\mathbf{l}}^2 + U_b[1 - \cos(d\theta^b - nb + qA)_{\mathbf{l}}] \right. \\
& \left. + U_a[1 - \cos(d^{\star}\theta^a - na + pA^{\star})_{\mathbf{l}^{\star}}] \right] \\
& + \sum_{\mathbf{p}=\mathbf{v}^{\star}} V_b(db - 2\pi s^b)_{\mathbf{p}}^2 + \sum_{\mathbf{v}=\mathbf{p}^{\star}} V_a(d^{\star}a - 2\pi s^a)_{\mathbf{p}^{\star}}^2. \quad (14)
\end{aligned}
$$

Since the ϵ_a, ϵ_b terms do not commute with some of the remaining terms (nor with each other), a gap is opened up; the fluctuations of a, b on each link are also softly bounded by these terms. As the ϵ_a, ϵ_b terms violate the gauge invariances (6) and (8), the gauge field picture of the a, b variables is no longer exact, but emergent at best. Indeed, in the remaining sections of this paper, we will show such a gauge field picture emerges at the low-energy sector on each individual link when ϵ_a, ϵ_b are small compared to U_a, U_b, and we are able to solve the low-energy sectors of the full Hamiltonian, and show the system exhibits the desired Hall conductivity $-2pq/n$.

Before we move on towards the solution, we would like to comment on the relation between our work and the relevant literature. We first remark that the Hamiltonian H has appeared before as certain cases in [15,16], which numerically studied lattice models with Hall conductivity. The proper relation of the lattice models to Chern-Simons theory and the subsequent path towards analytically solving the models, however, were not addressed in [15,16]. In Appendix F we provide a detailed explanation of how to properly identify the nature of the topological orders for all of the models in [15,16], so to connect those models to Chern-Simons theory and to our work. In our work, we motivated the lattice Hamiltonian H starting from a class of well-established Chern-Simons theories, from which the solvability follows as a natural consequence. Our approach is systematic and can be straightforwardly applied to the more general twisted bosonic and fermionic topological orders [11]. We will elaborate on those more general cases in upcoming works; this work serves to demonstrate our approach through the most basic examples.

In [16] the relation between some of the models and the $\mathbb{Z}_n$ toric code was discussed in the absence of coupling to the electromagnetic background (see Appendix F). As we will see in the following, interestingly, when we solve the Hamiltonian in the presence of electromagnetic coupling, crucial features appear: the low-energy subspace emerges to be an effective $\mathbb{Z}_n$-gauge-like theory, but with the $\mathbb{Z}_n$ in some ways "shifted by" and "projective under" the electromagnetic background; such modification is essential for a nontrivial Hall conductivity. Such background-dependent emergent field is beyond

the celebrated paradigm of exactly solvable lattice models established since the seminal works [4–6], and may shed light on the studies of the more general symmetry-enriched topological orders that are beyond this celebrated paradigm.

More recently [17] investigated exactly solvable local commuting projector Hamiltonians with integer values of Hall conductivity. There are two differences with our present work. First, [17] focused on integer Hall conductivity, while the Hall conductivity in our present work takes fractional values in general. Second, the spirit of their lattice model bears a major difference with ours. The model in [17] involved U(1)-valued local Hilbert space, and for their Hamiltonians to be made of local commuting projectors, the matrix elements must not be continuous functions of both the U(1) dynamical variables and the electromagnetic U(1) background gauge field. On the other hand, our Hamiltonians' matrix elements are regular functions of the continuous dynamical variables; the electromagnetic U(1) background gauge field also couples to the system in the usual form. Our models hold on to these important physical requirements [10] at the cost of exact solvability. It is interesting to note, however, that the discontinuities in the models of [17] may be interpreted as a reminiscence of the emergent 1-form $\mathbb{Z}$-gauge invariance (6); see the discussion below (29) for the origin of the emergent "discontinuity" in our continuous model. Since it is well known that discontinuities appear when the group cohomology machinery [6] is applied to continuous global symmetry groups [electromagnetic U(1) here], our resolution to the discontinuity problem is another reflection that our construction is reaching beyond the currently established paradigm.

III. SOLVING THE HAMILTONIAN

Now we proceed to solve the Hamiltonian H at its low-energy sector and find the Hall conductivity. The sketch is the following.

(i) First we solve the link terms in a solvable limit, i.e., perturbatively find the low-energy subspace determined by the $\epsilon_a, \epsilon_b, U_a, U_b$ terms on each individual link, which turn out to be an emergent $\mathbb{Z}_n$ subspace but with operator values shifted by the electromagnetic field. The approximate low-energy wave functions are illustrated in Fig. 3.

(ii) Then, with this emergent $\mathbb{Z}_n$ low-energy subspace on each link, we can solve the V_a, V_b terms in a manner similar to solving the toric code [4], and find fractionally charged anyons and the fractional Hall conductivity.

(iii) Finally, since we have used perturbation theory, we need to confirm that the Hall conductivity is precisely the designated fraction $-2pq/n$, rather than some nearby number blurred by the perturbative error. This is done by narrowing down our perturbative error without closing the gap, and implementing the usual derivation for the robustness of fractional Hall conductivity in gapped systems [22].

In this and the following section we carry out these steps. We remark that our solving procedure resembles what one does to study a real solid-state or cold-atom system: one first identifies the local low-energy subspace (the relevant electronic orbits on an atom, or the relevant atomic orbits in a potential trap) out of an infinite-dimensional local full Hilbert
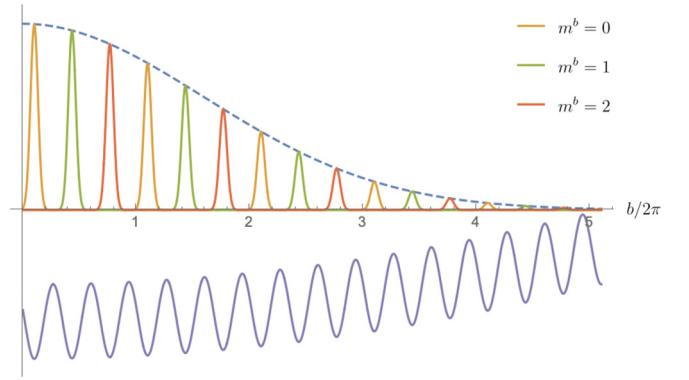

FIG. 3. The trial ground-state wave functions $|\Psi_{m^b}^{(0,0)}(b)|$ for the local Hamiltonian on each link, in the b basis. The "potential" for the b variable, $(\epsilon/2)b^2 - U \cos n(b - b_0)$, is illustrated below with the purple line. The parameters used are $n = 3$, $\epsilon_b = \epsilon_a = \epsilon$, $U_a = U_b = U$, $\epsilon/U = 1/4^4$, and $b_0 = 0.2\pi$. In the b basis, the wide envelope represented by the dashed line is a Gaussian centered at 0, with width $2\pi/nW_a$ where $W_a \sim (\epsilon_b/U_a)^{1/4}$; each narrow peak is a Gaussian centered at $\bar{b}_j = (2\pi/n)j + \bar{b}_0$, with width $W_b \sim (\epsilon_a/U_b)^{1/4}$. The errors of these trial wave functions to the true ones are $O(W^2)$.

space; then one projects the many-body interactions onto the low-energy subspace to study the many-body problem.

A. Local low-energy subspace

In this section, we solve the local low-energy subspace on each link in a solvable limit. The local Hamiltonian can be expressed as (the link index is omitted in this section)

$$H_{\text{link}} = \frac{1}{2}(\epsilon_b b^2 + \epsilon_a a^2)$$
$$- U_b \cos[n(b - \bar{b}_0)] - U_a \cos[n(a - \bar{a}_0)], \quad (15)$$

where $\bar{a}_0$ and $\bar{b}_0$ include terms that commute with the b and a operators. Our task is to perturbatively find the low-energy subspace of H_{link}, under the assumption

$$\epsilon \ll U, \quad (16)$$

where ϵ is the scale of ϵ_a and ϵ_b, and U the scale of U_a and U_b. We will find trial wave functions for n nearly degenerate low-energy states that are well separated for other states, forming an emergent $\mathbb{Z}_n$ low-energy subspace; the actual low-energy wave functions differ from our trial ones by $O(\sqrt{\epsilon/U})$ controlled errors, and their energy split is $O(\sqrt{\epsilon/U})$ smaller than their gap to the higher-energy states. The full details of the calculation and error control in this section can be found in Appendix B.

To motivate our trial wave functions, note that the leading noncommuting pairs of terms are the ϵ_a and U_b pair, as well as the ϵ_b and U_a pair. Let us first consider the ϵ_a and U_b terms

$$H_{\text{link},b} = \frac{1}{2}\epsilon_a a^2 - U_b \cos[n(b - \bar{b}_0)] \quad (17)$$

which may be viewed as a particle in a one-dimensional sinusoidal potential. Alternatively, if we are to view b as some gauge field as motivated in the previous section, then the U_b term is the Higgs potential and the ϵ_a term is the kinetic energy that opens the Higgs gap near the potential minima.

As long as the potential is deep enough, the neighborhood around each minimum, located at $\bar{b}_j = (2\pi/n)j + \bar{b}_0$ with $j \in \mathbb{Z}$, can be effectively expanded to the quadratic order. Therefore, the low-energy eigenfunctions are approximately those of infinitely many harmonic oscillators, one at each potential minimum $\bar{b}_j$. The eigenstates $|\phi_j^{(N_b)}\rangle$ are those of the harmonic oscillators at the potential minima $\bar{b}_j$ and with energy level N_b. The excitation energy of which is

$$\omega_b = 2\pi\sqrt{U_b \epsilon_a}. \tag{18}$$

In particular, the low-energy $|\phi_j^{(N_b)}\rangle$ are localized Gaussians of width

$$W_b = \frac{\sqrt{2\pi}}{n}\left(\frac{\epsilon_a}{U_b}\right)^{\frac{1}{4}} \tag{19}$$

centered at $\bar{b}_j$. As long as N_b is not too large, the tunneling overlap between different j's is exponentially small in $-1/W_b^2$, hence the N_b states, at least for small N_b can be treated as a nearly orthonormal and nearly degenerate under $H_{\text{link},b}$, analogous to a "tight-binding" electron in a sinusoidal potential. We need to consider $N_b = 0$ for low-energy subspace, and small $N_b > 0$ for error control. The higher powers in the expansion of the cos minima mix $N_b = 0$ with small N_b and modify the wave functions by a controlled error $O(W_b^2)$.

Now we have a new set of basis states describing the low-energy physics of the system, spanned by infinitely many highly localized orbits $|\phi_j^{(N_b)}\rangle$. Our next step would be to solve the remaining terms

$$H_{\text{link},a} = \frac{1}{2}\epsilon_b b^2 - U_a \cos[n(a - \bar{a}_0)] \tag{20}$$

in this basis. While the U_a term, the "hopping term of the tight-binding model," remains exact after the projection into this basis (as this basis spans an invariant subspace under the U_a

term), the "potential well" ϵ_b term does not. We approximate

$$b^2|\phi_j^{(N_b)}\rangle \approx \bar{b}^2|\phi_j^{(N_b)}\rangle \tag{21}$$

which we must justify later (see the end of this subsection). Here $\bar{b}$ is the operator identifying the center coordinates $(2\pi/n)j + \bar{b}_0$ of the orbits, which is well defined at least for N_b not too large. This means that at low energies we no longer need to consider the continuous b variable but can concentrate on a discrete $\bar{b}$ space for each level N_b. The basis of the conjugate variable of $\bar{b}$ operator, $\underline{a}$ (not to be confused with $\bar{a}$), is then confined onto the "first Brillioun zone" in the reciprocal space of the $\bar{b}_j$ lattice, i.e., $\underline{a} \in [0, 2\pi)$. Since the roles of $\underline{a}$ and $\bar{b}$ are conjugate, we can switch the perspective and now interpret $\underline{a}$ as a coordinate on a ring. Thus, we can regard $-U_a \cos[n(a - \bar{a}_0)] = -U_a \cos[n(\underline{a} - \bar{a}_0)]$ term as a sinusoidal potential, subjected to the periodic boundary condition identifying $\underline{a} = 0$ and 2π. This potential has n minima located at $\bar{a}_{m^a} = (2\pi/n)m^a + \bar{a}_0$, $m^a \in \mathbb{Z}_n$. We can solve the Hamiltonian around each minimum, again treating the potential in the neighborhoods of the minima as quadratic. The corresponding excitation energy is

$$\omega_a = 2\pi\sqrt{\epsilon_b U_a} \tag{22}$$

and the low-energy states $\tilde{\Psi}_{m^a}^{(N_b, N_a)}$ can be solved. They are n-fold degenerate for each level, and again localized in $\underline{a}$ basis by Gaussian width

$$W_a = \frac{\sqrt{2\pi}}{n}\left(\frac{\epsilon_b}{U_a}\right)^{\frac{1}{4}}. \tag{23}$$

Similar to the discussion in the previous step, the errors are $O(W_a^2)$.

Piecing up the above, we conclude that the link has an emergent $\mathbb{Z}_n$-like low-energy subspace, with trial wave functions

$$\left|\tilde{\Psi}_{m^a}^{(0,0)}\right\rangle = \sqrt{\frac{W_b}{n\pi W_a}} \int da \, e^{-i\frac{n}{2\pi}\bar{b}_0(a - \bar{a}_{m^a})} e^{-W_b^2(n/2\pi)^2 a^2/2}\left(\sum_{z^a \in \mathbb{Z}} e^{-[a - \bar{a}_{m^a} + nz^a]^2/2W_a^2}\right)|a\rangle \tag{24}$$

or their linear combinations

$$\left|\Psi_{m^b}^{(0,0)}\right\rangle \equiv \frac{1}{\sqrt{n}}\sum_{m^a} e^{-i\frac{n}{2\pi}\bar{a}_{m^a}\bar{b}_{m^b}}\left|\Psi_{m^a}^{(0,0)}\right\rangle$$

$$= \sqrt{\frac{W_a}{n\pi W_b}} \int db \left(\sum_{z^b \in \mathbb{Z}} e^{in\bar{a}_0 z^b} e^{-(n/2\pi)^2 W_a^2 \bar{b}_{m^b + nz^b}^2/2} e^{-[b - \bar{b}_{m^b + nz^b}]^2/2W_b^2}\right)|b\rangle. \tag{25}$$

Here m^a and m^b take values in $\{0, 1, \ldots, n - 1\}$; if we shift either by n, the corresponding wavefunction will return to itself but with an overall phase that depends on $\bar{b}_0$ or $\bar{a}_0$, an important feature different from an actual $\mathbb{Z}_n$, as we will see later. An illustration of $\Psi_{m^b}^{(0,0)}$ in b basis can be found in Fig. 3; it is clear that the solutions to $H_{\text{link},b}$ give the narrow peaks, with width W_b, while the solution to $H_{\text{link},a}$ determines the broad envelope, with width $2\pi/nW_a$.

One may notice the two expressions above are slightly asymmetric between a and b. This is because we treated $H_{\text{link},b}$ first. There is no contradiction to the apparent

symmetry between a and b in the original problem (if we have set $\epsilon_a = \epsilon_b = \epsilon$ and $U_a = U_b = U$) because these are trial wave functions that have $O(\sqrt{\epsilon/U})$ errors with the actual low-energy states anyways, and the said asymmetry is indeed of this order.

At this point we shall return and justify the approximation in (21). At first sight it seems this cannot be justified because no matter how small ϵ_b/U_b is, the b^2 potential will eventually be large enough to overcome the cos potential which determines the $\bar{b}$ minima; in other words, it is not obviously controlled to find the low-energy states of $H_{\text{link},b}$ first and then

project $H_{\text{link},a}$ into them. Our justification, in intuitive terms, is to note that the trial wave functions $|\Psi_{m^b}^{(N_a,N_b)}\rangle$ (with small values of N_a, N_b of interest) already have small amplitudes when b^2 is large enough to overcome ϵ_b/U_b. In particular, the typical spread of b in these wave functions is determined by the broad envelope of width $\sim 1/W_a \sim (\epsilon/U)^{-1/4}$, therefore, $\epsilon_b b^2 \sim \epsilon\sqrt{U/\epsilon}$ is still smaller than U_b by a factor of $O(\sqrt{\epsilon/U})$. A more rigorous error control is given in Appendix B. This error may introduce splitting among the n actual low-energy states, by an amount $O(\epsilon)$ which is $O(\sqrt{\epsilon/U})$ smaller than the gap $O(\sqrt{\epsilon U})$ with the higher-energy states.

B. Many-body ground states and excitations

Now that we have found the link low-energy subspace to be an effective $\mathbb{Z}_n$-like field [but bears important difference with a literal $\mathbb{Z}_n$ field, see discussion below (29)], one can envision that the remaining solution to the many-body problem is similar to that of a $\mathbb{Z}_n$ toric code, with suitable modifications by the electromagnetic background. We will see this is indeed the case. To control the error, the assumptions to be made are

$$\epsilon \ll V \ll \sqrt{\epsilon U}, \qquad (26)$$

where V is the scale of V_a, V_b. The second separation of scales is to ensure the applicability of the usual many-body perturbation theory, where we can neglect the higher-energy states in each local Hilbert space when considering the interaction couplings between local systems; this is justified in Appendix C. The first separation of scales is to ensure that the energy split between the states within each local low-energy subspace can be treated as a local perturbation on top of the toric-code-like physics, so that the robustness of the toric-code-like physics [4] is applicable.

We first note that for each link low-energy trial state, we shall make the substitutions

$$b_{0,\mathbf{l}} \rightarrow (d\theta^b + qA)_{\mathbf{l}}/n, \quad a_{0,\mathbf{l}^\star} \rightarrow (d^\star\theta^a + pA^\star)_{\mathbf{l}^\star}/n$$

so that the link state $|\Psi_{m^b,\mathbf{l}}^{(0,0)}\rangle$ depends on the neighboring vertex states $|\theta_{\mathbf{v}}^b\rangle$ and plaquette states $|\theta_{\mathbf{v}^\star}^a\rangle$, and also on the electromagnetic background. So, any given sets of $\{m_{\mathbf{l}}^b\}$, $\{\theta_{\mathbf{v}^\star}^a\}$, $\{\theta_{\mathbf{v}}^b\}$ specify a basis state for the many-body low-energy basis

$$|\{m_{\mathbf{l}}^b\}, \{\theta_{\mathbf{v}^\star}^a\}, \{\theta_{\mathbf{v}}^b\}\rangle \equiv \bigotimes_{\mathbf{v}^\star,\mathbf{v},\mathbf{l}} |\theta_{\mathbf{v}^\star}^a\rangle|\theta_{\mathbf{v}}^b\rangle|\Psi_{m_{\mathbf{l}}^b,\mathbf{l}}^{(0,0)}\rangle. \qquad (27)$$

The Fourier transforms of $\{\theta_{\mathbf{v}^\star}^a\}$ can also serve as a well-defined basis:

$$|\{m_{\mathbf{l}}^b\}, \{s_{\mathbf{p}}^b\}, \{\theta_{\mathbf{v}}^b\}\rangle \equiv \int_{\{\theta_{\mathbf{v}^\star}^a\}} e^{i\sum_{\mathbf{v}^\star} s_{\mathbf{p}}^b \cdot \theta_{\mathbf{v}^\star}^a} |\{m_{\mathbf{l}}^b\}, \{\theta_{\mathbf{v}^\star}^a\}, \{\theta_{\mathbf{v}}^b\}\rangle. \qquad (28)$$

The basis vectors can be expanded in the original basis (for simplicity, we omit the possible normalization factors in this subsection)

$$|\{m_{\mathbf{l}}^b\}, \{s_{\mathbf{p}}^b\}, \{\theta_{\mathbf{v}}^b\}\rangle = \sum_{\{z_{\mathbf{l}}^b\}} \int_{\{b_{\mathbf{l}}\}} \left(\prod_{\mathbf{l}} e^{iz_{\mathbf{l}}^b \cdot (pA^\star)_{\mathbf{l}^\star}} e^{-\bar{b}_{m_{\mathbf{l}}^b,z_{\mathbf{l}}^b}^2 (n/2\pi)^2 W_a^2/2} e^{-(b_{\mathbf{l}} - \bar{b}_{m_{\mathbf{l}}^b,z_{\mathbf{l}}^b})^2/2W_b^2} \right) \bigotimes_{\mathbf{l},\mathbf{p},\mathbf{v}} |b_{\mathbf{l}}\rangle |(s^b + dz^b)_{\mathbf{p}}\rangle |\theta_{\mathbf{v}}^b\rangle, \qquad (29)$$

where the minima positions

$$\bar{b}_{m_{\mathbf{l}}^b,z_{\mathbf{l}}^b} \equiv \frac{2\pi}{n}\left(m^b + nz^b + \frac{d\theta^b + qA}{2\pi}\right)_{\mathbf{l}}. \qquad (30)$$

We make a few remarks about this set of many-body low-energy basis states.

(i) The local minima in the b basis, specified by $\bar{b}_{m^b,z^b}$, may be viewed as an emergent realization of the Gauss's constraint $g_{\mathbf{l}}^b = e^{-iqA_{\mathbf{l}}}$ [see also (11)]. The positions of the minima are shifted by the background electromagnetic field $A_{\mathbf{l}}$.

(ii) For definiteness of the labels, we need to fix the emergent $\mathbb{Z}_n$ label $m_{\mathbf{l}}^b$ to take values in the range, say, $\{0, 1, \ldots, n-1\}$. If an $m_{\mathbf{l}}^b$ shifts by n, the effect can be absorbed into a shift of the nearby $s_{\mathbf{p}}^b$ through a shift of the summation variable $z_{\mathbf{l}}^b$; moreover, an overall phase that depends on $A_{\mathbf{l}^\star}^\star$ is generated. In other words, for $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}$, the mod out of $n\mathbb{Z}$ is now "projective" under the electromagnetic field $A_{\mathbf{l}^\star}^\star$. It is easy to see that the summation over $z_{\mathbf{l}}^b$, entangling the link Hilbert space and the nearby plaquette Hilbert spaces, is an emergent realization of the Gauss's constraint $g_{\mathbf{l}^\star}^a = e^{-ipA_{\mathbf{l}^\star}^\star}$ which generates the first line of the invariance (6).

(iii) While any final physical result must not depend on any 2π shift of $\theta_{\mathbf{v}}^b$ or $A_{\mathbf{l}}$ because the original Hamiltonian does not, the labeling of the basis states does. If such a shift is made, through the expression of $\bar{b}_{m_{\mathbf{l}}^b,z_{\mathbf{l}}^b}$ we can see $m_{\mathbf{l}}^b$ must be relabeled to represent the same physical state; furthermore, if $m_{\mathbf{l}}^b$ is relabeled out of the $\{0, 1, \ldots, n-1\}$ range, to relabel it back in, the nearby $s_{\mathbf{p}}^b$ must also be relabeled, and an overall phase that depends on ("projective under") $A_{\mathbf{l}^\star}^\star$ is generated. One may avoid this ambiguity by fixing $\theta_{\mathbf{v}}^b$ and $A_{\mathbf{l}}$ to belong to $(-\pi, \pi]$, but we do not have to, as long as the said relabeling is understood.

All these are crucial features of our emergent $\mathbb{Z}_n$-like theory, compared to a usual lattice $\mathbb{Z}_n$ theory that cannot couple to electromagnetic background. In particular, the last point above is closely related to the *effective* indistinguishability (A6), which is in turn crucial to the embodiment of the Kapustin-Fidkowski no-go theorem (see Appendix A). More broadly, the fact that the emergent $\mathbb{Z}_n$-like field is "shifted by" and "projective under" the electromagnetic background is a feature that cannot be achieved in the standard paradigm of constructing exactly solvable lattice models [4–6] which would have started with a literal $\mathbb{Z}_n$ field; in particular, the relabeling of $m_{\mathbf{l}}^b, s_{\mathbf{p}}^b$ needed when $A_{\mathbf{l}}$ gradually increases from 0 to 2π may be interpreted as the "discontinuity" [17] in the effective description, and here we see such "discontinuity" emerges from a lattice model that is in fact continuous in all variables. This may shed light on the study of more general symmetry-enriched topological orders.

It is useful to also introduce the dual basis states $|\{m_{\mathbf{l}^\star}^a\}, \{\theta_{\mathbf{v}^\star}^a\}, \{s_{\mathbf{p}^\star}^a\}\rangle$ in a symmetric manner to (29), with minima

$$\bar{a}_{m_{\mathbf{l}^\star}^a, z_{\mathbf{l}^\star}^a} \equiv \frac{2\pi}{n}\left(m^a + nz^a + \frac{d^\star\theta^a + pA^\star}{2\pi}\right)_{\mathbf{l}^\star}. \quad (31)$$

[As commented in the paragraph after (25), within our controlled error we can safely neglect the slight asymmetry between the $\langle a|\tilde{\Psi}_{m^a}^{(0,0)}\rangle$ and the $\langle b|\Psi_{m^b}^{(0,0)}\rangle$ expressions.] The formal remarks are analogous to the above and will not be repeated.

The many-body part of the Hamiltonian is $H_b + H_a$, where $H_b \equiv V_b \sum_{\mathbf{p}}(\frac{db}{n} - 2\pi s^b)_{\mathbf{p}}^2$ and $H_a \equiv V_a \sum_{\mathbf{v}}(\frac{d^\star a}{n} - 2\pi s^a)_{\mathbf{p}^\star}^2$. In Appendix C we justify that it suffices to focus on their projections into the local low-energy trial subspace we found above. As in the usual toric code these two terms commute and can be simultaneously diagonalized, even so after the projection into the local low-energy trial subspace, up to errors exponentially small in $1/W^2$. Moreover, up to the same error, it is straightforward to show that H_b and H_a are, respectively, diagonal in the two sets of trial basis states introduced above, with eigenvalues

$$V_b \sum_{\mathbf{p}}\left(\frac{2\pi dm^b + qdA}{n} - 2\pi s^b\right)_{\mathbf{p}}^2, \quad (32)$$

$$V_a \sum_{\mathbf{p}^\star}\left(\frac{2\pi d^\star m^a + pd^\star A^\star}{n} - 2\pi s^a\right)_{\mathbf{p}^\star}^2, \quad (33)$$

respectively. Therefore, we can denote the local excitation numbers by

$$v_{\mathbf{p}}^b \equiv -\left(dm^b - ns^b + \left[\frac{qdA}{2\pi}\right]\right)_{\mathbf{p}}, \quad (34)$$

$$v_{\mathbf{p}^\star}^a \equiv -\left(d^\star m^a - ns^a + \left[\frac{pd^\star A^\star}{2\pi}\right]\right)_{\mathbf{p}^\star}, \quad (35)$$

where $[x]$ denotes the nearest integer to a real number x. H_b and H_a are minimized when v^b and v^a vanish, respectively.

To see $v_{\mathbf{p}}^b$ describes anyon excitations, we use the Wilson loop operator

$$L_\ell^b \equiv \exp\left[i\sum_{\mathbf{l}\in\ell} b_{\mathbf{l}}\right] \rightarrow \exp\left[i\frac{2\pi}{n}\sum_{\mathbf{l}\in\ell}\left(m^b + q\frac{A}{2\pi}\right)_{\mathbf{l}}\right] \quad (36)$$

with ℓ a lattice loop encircling $\mathbf{p}$; the expectation would be a phase $e^{-i2\pi v_{\mathbf{p}}^b/n}$ in the absence of background field. Moreover, it is easy to see that if ℓ is a line with two open ends, the Wilson line L_ℓ^b will create a ± 1 pair of $v_{\mathbf{p}^\star}^a$ excitations at its ends. This is why $-v^b$ and v^a are usually called the "flux" and

"charge" anyons in the $\mathbb{Z}_n$ toric code, named in the perspective of b being a gauge field. The $-2\pi/n$ phase in the expectation value is the mutual braiding statistics between the charge and the flux anyons. Similarly,

$$L_{\ell^\star}^a \equiv \exp\left[i\sum_{\mathbf{l}^\star\in\ell^\star} a_{\mathbf{l}^\star}\right] \rightarrow \exp\left[i\frac{2\pi}{n}\sum_{\mathbf{l}^\star\in\ell^\star}\left(m^a + p\frac{A^\star}{2\pi}\right)_{\mathbf{l}^\star}\right] \quad (37)$$

detects $v_{\mathbf{p}^\star}^a$ if $\ell^\star$ is a dual lattice loop encircling $\mathbf{p}^\star$, and creates a ± 1 pair of $v_{\mathbf{p}}^b$ if $\ell^\star$ has open ends. We will leave the discussions of the electromagnetic responses to the next section.

Consider the condition (34). Different m^b, s^b configurations may have the same values of $(dm^b - ns^b)_{\mathbf{p}}$. Locally, such equivalent configurations are generated by the $\mathbb{Z}/n\mathbb{Z}$ reminiscence of the 0-form $\mathbb{R}/2\pi\mathbb{Z}$ gauge invariance, the first line of (8) rescaled by $n/2\pi$. The mod out of the $n\mathbb{Z}$ is embodied such that, if a 0-form gauge transformation brings some $m_{\mathbf{l}}^b$ out of the specified range $\{0, 1, \ldots, n-1\}$, we must bring it back into the range by the 1-form $\mathbb{Z}$ gauge transformation mentioned before, which thereby changes the nearby $s_{\mathbf{p}}^b$ configurations [21]. The explicit form of such local transformation is given in (D1) in Appendix D.

Globally, there are also m^b, s^b configurations that give the same values of (34) everywhere, but nonetheless belong to different topological classes C. Consider the square lattice with periodic boundary conditions, forming a torus. Each topological class C is characterized by the expectation values of two Wilson loop operators, $L_{\ell_x}^b$ and $L_{\ell_y}^b$, where ℓ_i is a noncontractible loop that runs around the i direction. The particular path that ℓ_i runs through does not matter if $A_{\mathbf{l}} = 0$ and $v_{\mathbf{p}}^b = 0$ everywhere, but otherwise we need to fix the particular path to make comparisons. Both $L_{\ell_x}^b$ and $L_{\ell_y}^b$ can take n different expectation values differing from each other by a $e^{i2\pi/n}$ phase. Therefore, for fixed A and v^b, the total number of topological classes C on a torus is n^2 [4].

Similar discussions apply if we consider the condition (35). The local transformation is given in (D2). The topological classes, characterized by $L_{\ell_x^\star}^a$ and $L_{\ell_y^\star}^a$, are labeled by $C^\star$.

So, far, we have considered H_a and H_b separately under the two sets of bases $\{|\{m_{\mathbf{l}}^b\}, \{s_{\mathbf{p}}^b\}, \{\theta_{\mathbf{v}}^b\}\rangle\}$ and $\{|\{m_{\mathbf{l}^\star}^a\}, \{s_{\mathbf{p}^\star}^a\}, \{\theta_{\mathbf{v}}^b\}\rangle\}$, respectively. However, in order to find the eigenstates for $H_a + H_b$, where the two terms commute and can be simultaneously diagonalized (even so after projected into the links' low-energy subspaces in the small-W limit), we must take proper superpositions within either set of basis states [4,5]. In particular, we need to sum over all the aforementioned gauge-equivalent states that share the same $\{v_{\mathbf{p}^\star}^a\}, \{v_{\mathbf{p}}^b\}$ excitation configurations and belong to the same topological class C or $C^\star$, with suitable phase coefficients. We find the proper combinations of all gauge-equivalent states to be

$$|C\rangle \equiv \sum_{\{t_{\mathbf{v}}^b\}}\int_{\{\theta_{\mathbf{v}}^b\}} \exp\left[-\sum_{\mathbf{v}}\frac{i}{n}(2\pi t^b + \theta^b)_{\mathbf{v}}\left(\left[:\frac{pd^\star A^\star}{2\pi}:\right] - v^a\right)_{\mathbf{p}^\star = \mathbf{v}}\right] \mathsf{T}(\{t_{\mathbf{v}}^b\})\,|\{m_{\mathbf{l}}^{b,C\,rep}\}, \{s_{\mathbf{p}}^{b,C\,rep}\}, \{\theta_{\mathbf{v}}^b\}\rangle, \quad (38)$$

$$|C^\star\rangle \equiv \sum_{\{t_{\mathbf{v}^\star}^a\}}\int_{\{\theta_{\mathbf{v}^\star}^a\}} \exp\left[-\sum_{\mathbf{v}^\star}\frac{i}{n}(2\pi t^a + \theta^a)_{\mathbf{v}^\star}\left(\left[:\frac{qdA}{2\pi}:\right] - v^b\right)_{\mathbf{p} = \mathbf{v}^\star}\right] \mathsf{T}^\star(\{t_{\mathbf{v}^\star}^a\})\,|\{m_{\mathbf{l}^\star}^{a,C^\star\,rep}\}, \{\theta_{\mathbf{v}^\star}^a\}, \{s_{\mathbf{p}^\star}^{a,C^\star\,rep}\}\rangle, \quad (39)$$

where we introduced $[: x :] \equiv x - [x]$. The state $|\{m_{\mathbf{l}}^{b,C\,rep}\}, \{s_{\mathbf{p}}^{b,C\,rep}\}, \{\theta_{\mathbf{v}}^{b}\}\rangle$ is a representative state in the class C and $|\{m_{\mathbf{l}^{\star}}^{a,C^{\star}\,rep}\}, \{\theta_{\mathbf{v}^{\star}}^{a}\}, \{s_{\mathbf{p}^{\star}}^{a,C^{\star}\,rep}\}\rangle$ in $C^{\star}$ (with some given v^{b}, v^{a} configurations of excitations), and $\mathsf{T}(\{t_{\mathbf{v}}^{b}\})$ and $\mathsf{T}^{\star}(\{t_{\mathbf{v}^{\star}}^{a}\})$ are the aforementioned 0-form $\mathbb{Z}_n$ gauge transformations specified by $\mathbb{Z}_n$ variables $\{t_{\mathbf{v}}^{b}\}$, $\{t_{\mathbf{v}^{\star}}^{a}\}$, the definition of which is given in details in (D1) and (D2). The set of states $\{|C\rangle\}$ are manifest eigenstates of H_b, while the set $\{|C^{\star}\rangle\}$ are manifest eigenstates of H_a, but we now show either set can actually serve as simultaneous eigenstates of both terms, i.e., eigenstates of the entire Hamiltonian. To see this, in Appendix D we show $|C\rangle$ is a linear combination of all $|C^{\star}\rangle$:

$$|C\rangle = \sum_{C^{\star}} \prod_{\mathbf{l}} e^{-i\frac{2\pi}{n}(m^{b,C\,rep} + \frac{qA}{2\pi})_{\mathbf{l}}(m^{a,C^{\star}\,rep} + \frac{pA^{\star}}{2\pi})_{\mathbf{l}^{\star}}} |C^{\star}\rangle \quad (40)$$

which means either the set $\{|C\rangle\}$ or the set $\{|C^{\star}\rangle\}$ is a choice of orthonormal basis for the n^2-dimensional eigensubspace, specified by the given anyon configurations $\{v_{\mathbf{p}^{\star}}^{a}\}$ and $\{v_{\mathbf{p}}^{b}\}$, of the full Hamiltonian.

We emphasize again that the under 2π shift of $A_{\mathbf{l}}$ (and the identified $A_{\mathbf{l}^{\star}}^{\star}$), the explicit expression for a given $|C\rangle$ or $|C^{\star}\rangle$ in the original physical basis of $|b_{\mathbf{l}}\rangle$ or $|a_{\mathbf{l}^{\star}}\rangle$ must remain unchanged, even though the representative labels $\{m_{\mathbf{l}}^{b,C\,rep}\}$, $\{s_{\mathbf{p}}^{b,C\,rep}\}$ or $\{m_{\mathbf{l}^{\star}}^{a,C^{\star}\,rep}\}$, $\{s_{\mathbf{p}^{\star}}^{a,C^{\star}\,rep}\}$ would change. Clearly, the physical excitation numbers $\{v_{\mathbf{p}}^{b}\}$, $\{v_{\mathbf{p}^{\star}}^{a}\}$ and the Wilson loops $L_{\ell_{i}}^{b}$ or $L_{\ell_{\mathbf{i}}^{\star}}^{a}$ that characterize $|C\rangle$ or $|C^{\star}\rangle$ remain unchanged.

We finally recall that all we have solved above was $H_a + H_b$ projected in the trial subspaces $\bigotimes_{\mathbf{l}} \mathcal{T}_{\mathbf{l}}$, while we really should have done it in the actual low-energy subspaces $\bigotimes_{\mathbf{l}} \mathcal{L}_{\mathbf{l}}$. Since the error between $\mathcal{L}_{\mathbf{l}}$ and $\mathcal{T}_{\mathbf{l}}$ is $O(\sqrt{\epsilon/U})$ for each $\mathbf{l}$, we may view any correction as a having an extra perturbation Hamiltonian on each link $\mathbf{l}$. But it is well known that in the thermodynamic limit such local perturbation terms do not alter the topological physics of the toric code model [4], and the same reasoning applies here. Moreover, to ensure that the H_a, H_b couplings do not induce large mix with the higher-energy subspace $\mathcal{T}_{\perp,\mathbf{l}}$ on each link, we need the H_a, H_b to be smaller than ω_a, ω_b. The conditions are indeed the $\epsilon \ll V \ll \sqrt{\epsilon U}$ that we mentioned at the beginning of this subsection.

IV. HALL CONDUCTIVITY AND FRACTIONALIZED ELECTRIC CHARGE

In the previous section we have controllably solved the Hamiltonian, with errors bounded by the orders $O(\sqrt{\epsilon/U}, \epsilon/V, V/\sqrt{\epsilon U})$. Now we investigate the electromagnetic responses, most importantly the Hall conductivity, in both the local and the global perspectives. The local perspective is more straightforward, while the global perspective settles the exactness of the quantization of the responses, which is necessary since we have errors from the use of perturbation theory.

First we compute the expectation of the local electric charge density (10). Carrying out the calculation in

Appendix E, we find, in general,

$$\langle p s_{\mathbf{p}}^{b} \rangle = -\frac{p}{n} \frac{q(dA)_{\mathbf{p}}}{2\pi} + \frac{p}{n}\left(v^{b} + \left[\frac{qdA}{2\pi}\right]\right)_{\mathbf{p}},$$

$$\langle q s_{\mathbf{p}^{\star}}^{a} \rangle = -\frac{q}{n} \frac{p(d^{\star}A^{\star})_{\mathbf{p}^{\star}}}{2\pi} + \frac{q}{n}\left(v^{a} + \left[\frac{pd^{\star}A^{\star}}{2\pi}\right]\right)_{\mathbf{p}^{\star}} \quad (41)$$

which are independent of the class C.

When the background magnetic field dA is small (and so for $d^{\star}A^{\star}$, given the identification between $A^{\star}$ and A in Fig. 2) and when excitations v^{b}, v^{a} are absent, we find the local charge density to be

$$\langle \rho_{\mathbf{v}} \rangle = -\frac{pq}{2\pi n}[(dA)_{\mathbf{p}'=\mathbf{v}+\hat{\mathbf{x}}/2+\hat{\mathbf{y}}/2} + (dA)_{\mathbf{p}=\mathbf{v}-\hat{\mathbf{x}}/2-\hat{\mathbf{y}}/2}]. \quad (42)$$

Since the Hall conductivity $2\pi\sigma_H = \delta\langle\rho\rangle/\delta(dA)$, this means the Hall conductivity equals $-2pq/n$ as desired, at least approximately to within our perturbative error. We remark that the result applies to any geometry of the lattice, even for the lattice with boundaries. This suggests that upon adiabatically applying a uniform A, charge accumulates onto the boundary, another well-known manifestation of the Hall conductivity.

On the other hand, when the electromagnetic background is absent but v^{b}, v^{a} excitations present, we have

$$\langle \rho_{\mathbf{v}} \rangle = \frac{q}{n} v_{\mathbf{p}^{\star}=\mathbf{v}}^{a} + \frac{p}{n} v_{\mathbf{p}=\mathbf{v}-\hat{\mathbf{x}}/2-\hat{\mathbf{y}}/2}^{b}, \quad (43)$$

which means the v^{a} and v^{b} anyons carry fractionalized electric charges q/n and p/n, respectively. Another manifestation of the anyons' electric charges is the coefficients of the A dependence in the L^{b} and L^{a} Wilson lines, (36) and (37), which are the world line operators that create the v^{a} and v^{b} anyons, respectively.

Now we consider $(dA)_{\mathbf{p}}$ being not so small, so that it is closer to an integer $w_{\mathbf{p}}^{b}$ multiple of $-2\pi/q$. Then we can see from (32) that the ground states with $v^{b} = 0$ already energetically prefer m^{b} configurations that have $w_{\mathbf{p}}^{b}$ fluxes at $\mathbf{p}$. Then the Wilson loop operator L_{ℓ}^{b} with ℓ enclosing $\mathbf{p}$ indeed picks a phase $e^{i\frac{2\pi}{n}w_{\mathbf{p}}^{b}}$ aside from the regular part $e^{i\frac{q}{n}(dA)_{\mathbf{p}}}$, and the charge on $\mathbf{p}$ also receives a contribution $-\frac{p}{n}w_{\mathbf{p}}^{b}$. Therefore, these $\mathbb{Z}_n$ gauge fluxes should also be interpreted as flux anyons, but attached to the background fluxes. Likewise, a magnetic flux $(d^{\star}A^{\star})_{\mathbf{p}^{\star}}$ close to an integer $w_{\mathbf{p}^{\star}}^{a}$ multiple of $2\pi/p$ creates $w_{\mathbf{p}^{\star}}^{a}$ charge anyons at $\mathbf{p}^{\star}$ in the ground state. When some $(dA)_{\mathbf{p}}$ or $(d^{\star}A^{\star})_{\mathbf{p}^{\star}}$ take $(2\pi/q)(\mathbb{Z} + 1/2)$ or $(2\pi/p)(\mathbb{Z} + 1/2)$ value, the Hamiltonian becomes gapless in the vicinity of $\mathbf{p}$ or $\mathbf{p}^{\star}$, undergoing a transition to a new ground state with different anyon numbers.

The results above are all computed using the trial wave functions, which differ from the actual wave functions with errors controlled by $O(\sqrt{\epsilon/U}, \epsilon/V, V/\sqrt{\epsilon U})$. It is well known that the $\mathbb{Z}_n$ toric code topological physics discussed in the previous section are robust against such local errors [4]. One can envision that the topological electromagnetic responses, the Hall conductivity and the anyons' electric charges, are robust for similar reasons. In the following we focus on the Hall conductivity. We present a global definition of Hall conductivity according to [22] (which is also the setup employed to prove the Kapustin-Fidkowski no-go theorem [10]), which is robust in the thermodynamic limit, and show it must be precisely

quantized to our designated value $-2pq/n$ rather than any other nearby fraction blurred by the use of perturbation theory. The strategy is as follows:

(i) We employ the method of [22] to show that the globally defined Hall conductivity must be an integer multiple of the inverse of the ground-state degeneracy $1/n^2$.

(ii) We can compute that the value is close to $-2pqn/n^2$, controlled up to errors bounded by $O(\sqrt{\epsilon/U}, \epsilon/V, V/\sqrt{\epsilon U})$.

(iii) By adiabatically changing the parameters, we can make the errors much smaller than the spacing $1/n^2$ and eventually approach 0 arbitrarily closely, so we can conclude the Hall conductivity must be exactly $-2pqn/n^2$; since the gap is not closed when we decreased the error adiabatically, the Hall conductivity must take this value for some finite range of parameters away from the zero error limit.

Following [22], we consider a square lattice with periodic boundary conditions, forming a torus, and apply a uniform background, in which the $A_\mathbf{l}$ on all x-direction links take value α_x/N_x and the $A_\mathbf{l}$ on all y-direction links take value α_y/N_y, where N_i is the number of sites in the i direction (the identification of $A^\star$ with A is understood), so that there is no magnetic field, but only flat holonomies (α_x, α_y). Consider the two-parameter space where α_x and α_y vary from 0 to $2\pi n$. Suppose we fix α_y and vary α_x adiabatically (so there is a weak electric field in the y direction) from 0 to $2\pi n$. Each time α_x takes $2\pi\mathbb{Z}$ value, the background holonomies are gauge equivalent to the original $(\alpha_x = 0, \alpha_y)$. However, while the same background is revisited n times over this adiabatic process, each time the ground state [under the given background $(\alpha_x = 0, \alpha_y)$] may not be same one, until it must come back to the original one at the nth time. This can be seen by first performing a gauge transformation of A so that only one column of x-direction links has nonzero values, with $A_\mathbf{l} = \alpha_x$, and then carefully examining the definition of $|C\rangle$ and the definition of the noncontractible Wilson loops $L^b_{\ell_x}$, $L^b_{\ell_y}$ that characterize C. (In particular, one will find that, when p, q are both coprime with n, the adiabatically evolving state would reach n distinct ground states out of the total n^2, while for other values of p, q not both coprime with n, some of the visited ground states would already be revisited before the nth time.) Only when α_x reaches $2\pi n$ are we guaranteed to return to the original state; one may note, however, that there is an overall phase that depends on A, and this is in fact related to the Chern number below. The same reasoning applies if we fix α_x and vary α_y. Therefore, in the two-parameter space where α_x and α_y adiabatically vary from 0 to $2\pi n$, at the n^2 points where (α_x, α_y) are gauge equivalent to $(0, 0)$, the adiabatic state has visited the degenerate ground states under the $(0, 0)$ background holonomies for n^2 times (exhausting all the n^2 degenerate ground states if p, q are both coprime with n, but otherwise missing some and repeating some).

The globally defined Hall conductivity is given by the Chern number over the $\alpha_x, \alpha_y \in [0, 2\pi n)$ space of holonomies, averaged over n^2 visited states [22]:

$$2\pi\sigma_H = \frac{1}{n^2} \int_0^{2\pi n} d\alpha_x \int_0^{2\pi n} d\alpha_y \frac{\mathcal{B}}{2\pi},$$

$$\mathcal{B} \equiv -i\left(\left\langle\frac{\partial C_0}{\partial \alpha_x}\bigg|\frac{\partial C_0}{\partial \alpha_y}\right\rangle - \left\langle\frac{\partial C_0}{\partial \alpha_y}\bigg|\frac{\partial C_0}{\partial \alpha_x}\right\rangle\right). \qquad (44)$$

Here $\mathcal{B}$ is the Berry curvature in the space of holonomies, and importantly the integral of $\mathcal{B}/2\pi$ is the Chern number that must be an integer. Carrying out the calculation in Appendix E, we find the Chern number is indeed $-2pqn$, and hence the Hall conductivity is indeed $-2pq/n$ as expected. (In fact, in our particular model, the Berry curvature is constant over the space of holonomies, $\mathcal{B} = -pq/\pi n$.) In the calculation the largest error is still the difference between our trial ground states and the actual ground states, bounded by $O(\sqrt{\epsilon/U}, \epsilon/V, V/\sqrt{\epsilon U})$, hence our strategy above applies. The Hall conductivity must be the designated fraction within a finite range of parameters. This completes the proof of the exactness and robustness of the Hall conductivity.

V. CONCLUSION

In this paper we constructed a class of lattice Hamiltonians that can be solved at low energies and exhibit fractional Hall conductivity. The construction is systematic, motivated by the doubled Chern-Simons theory description of the associated bosonic topological orders enriched with electromagnetic U(1); the solution is obtained by a combination of perturbative and exact techniques. In forthcoming works we will elaborate on generalizations towards twisted bosonic topological orders as well as the fermionic ones.

This work is a solution to lattice Hamiltonians that exhibit fractional Hall conductivity. (The relation to the previous literature is explained at the end of Sec. II.) Our construction method is of theoretical significance, particularly in light of the Kapustin-Fidkowski no-go theorem [10] which forbids Hall conductivity in any local commuting projector Hamiltonian with finite-dimensional local Hilbert space. Since the previously existing methods to systematically construct exactly solvable lattice Hamiltonians for topological phases are mostly subjected to these limiting conditions, our work, going beyond these constraints, may shed light on the methodology in the more general studies of topological phases.

Let us further our discussion on the constraints in the Kapustin-Fidkowski no-go theorem. There are three nontrivial constraining assumptions in the statement of the theorem: the local Hilbert space is finite dimensional, the terms of the Hamiltonian commute, and the terms are projectors. Our construction breaks all three assumptions. It is interesting to ask whether one may break less of these constraining assumptions and still have Hall conductivity. Since commutativity is key to exact solvability, let us first consider the scenarios where commutativity is assumed:

(i) If the local Hilbert space remains finite dimensional, and the terms of the Hamiltonian are commuting, then relaxing the projector assumption does not help because the terms can be smoothly deformed to projectors without closing the gap.

(ii) At the end of [10], the possibility of local commuting projector Hamiltonian on infinite-dimensional local Hilbert space was mentioned, but such models are deemed not physical, as the Hamiltonian's matrix elements would cease to be continuous in the local dynamical and background variables on the lattice.

(iii) In the same discussion, the possibility of further relaxing the projector assumption was also mentioned. To us, this

possibility seems unlikely to help. In order for the terms in the Hamiltonian to be more physical than projectors, their matrix elements should be continuous functions of the local dynamical and background variables on the infinite-dimensional local Hilbert space. But these terms are also assumed commuting. Then it seems the Hamiltonian will always be gapless even in finite system sizes, as is in the case of our "prototype" Hamiltonian $\tilde{H}$.

Based on these arguments, we may conjecture the following:

(i) To have nontrivial Hall conductivity, it is necessary to give up the commutativity assumption. As a result, the correlations of local operators do not exactly vanish outside of any finite distance, but only decay exponentially.

(Note the conjecture is made given the aforementioned physical requirement that the Hamiltonian's matrix elements must be continuous in the dynamical and background variables on the lattice.) Such noncommuting Hamiltonians are not exactly solvable in general. If the arguments above can be turned into a rigorous proof to the conjecture, then we would have a theorem generalizing the theorem in [23] from noninteracting Chern insulators to interacting systems. A natural question to ask is then as follows:

(i) Can one construct a Hamiltonian, over finite-dimensional local Hilbert space and with noncommuting terms, that is controllably solvable at lower energies, gapped, and exhibits Hall conductivity?

(The projector assumption is no longer relevant if the terms are already noncommuting.) This is indeed a very interesting theoretical modeling problem to tackle. Moreover, limiting the local Hilbert space to be finite dimensional is usually seen as a desired "physical" feature because it means the Hamiltonian can be realized in generalized "spin" systems at least in principle. [For instance, recently [24] constructed exactly solvable lattice Hamiltonians for a large class of topological orders enriched by electromagnetic U(1) global symmetry but without Hall conductivity, and the Hilbert space being finite dimensional was emphasized, in comparison to the previous examples [21].] We would like to remind, however, that in the conceivable proposals to realize such generalized "spin" models in solid-state or cold-atom systems, that generalized "spin," the finite-dimensional local Hilbert space, is always the local lower-energy subspace of some infinite-dimensional full local Hilbert space; but this situation is indeed what happens in our models. Therefore, in this sense, a theoretical model with finite-dimensional local Hilbert space is not necessarily more "physical" than our ones with infinite-dimensional local Hilbert space.

APPENDIX A: FROM EFFECTIVE LAGRANGIAN TO TOY HAMILTONIAN

In [11] it was shown that the doubled Chern-Simons theory (1) has a Lagrangian description on *effective* space-time lattice, where by *effective* space-time lattice we mean a coarse-grained space-time manifold; the conceptual importance of this interpretation will become clear soon. For simplicity we will think of a three-dimensional cubic lattice, but the discussions below can be straightforwardly applied to any tetrahedral decomposition of a three-dimensional manifold, or a triangulation of a two-dimensional spatial manifold together with a discretization of time (hence a "prism" decomposition of the space-time). The action is

$$S = \frac{n}{2\pi} \sum_{plaq\ p} a_p (db)_p - n \sum_{plaq\ p} a_p s_p^b - n \sum_{link\ l} s_l^a b_l + \sum_{cube\ c} \theta_c^a (ds^b)_c + \sum_{vert\ v} \theta_v^b (\partial s^a)_v$$
$$- \frac{q}{2\pi} \sum_{link\ l} A_l (\partial a - 2\pi s^a)_l - \frac{p}{2\pi} \sum_{plaq\ p} A_p^\star (db - 2\pi s^b)_p + \sum_{plaq\ p} a_p L_p^a + \sum_{link\ l} b_l L_l^b. \tag{A1}$$

Here the dynamical variables include the following: a_p and b_l take real values on plaquette p and link l, respectively (one may view a_p as living on the dual lattice, with $l^\star = p$ in three dimensions, and likewise for other quantities with superscript ^{a}); s_l^a and s_p^b take integer values on link l and plaquette p, respectively; θ_c^a and θ_v^b take $U(1) = \mathbb{R}/2\pi\mathbb{Z}$ values on cube c and vertex v, respectively. The background fields A_l and $A_p^\star$ take real values on link l and plaquette p, respectively; they may either be independent, or identified in a certain way, say, $A_p^\star = A_{l=p+\hat{x}/2+\hat{y}/2+\hat{z}/2}$ (which can be generalized to other discretizations of the space-time as long as there is a branching structure). The Wilson loop observables L_p^a and L_l^b take integer values on plaquette p and link l, respectively. d is the lattice coboundary operator, i.e., lattice exterior derivative, and ∂ is the lattice boundary operator which is the inverse of d; one may check that

$$\sum_{cube\ c} \theta_c^a (ds^b)_c = \sum_{plaq\ p} (\partial \theta^a)_p s_p^b, \quad \sum_{plaq\ p} a_p (db)_p = \sum_{link\ l} (\partial a)_l b_l, \quad \sum_{link\ l} s_l^a (d\theta^b)_l = \sum_{vert\ v} (\partial s^a)_v \theta_v^b \tag{A2}$$

on a lattice without boundary. We require the Wilson loop observables to be closed loops: $(\partial L^b)_v = 0$, $(dL^a)_c = 0$. The action has the following gauge invariances mod 2π:

$$b_l \rightarrow b_l + 2\pi z_l^b + (d\varphi^b)_l, \quad s_p^b \rightarrow s_p^b + (dz^b)_p, \quad \theta_v^b \rightarrow \theta_v^b + n\varphi_v^b, \tag{A3}$$

$$a_p \rightarrow a_p + 2\pi z_p^a + (\partial \varphi^a)_p, \quad s_l^a \rightarrow s_l^a + (\partial z^a)_l, \quad \theta_c^a \rightarrow \theta_c^a + n\varphi_c^a, \tag{A4}$$

where z_p^a and z_l^b take integer values on plaquette p and link l, respectively, and φ_c^a and φ_v^b take $U(1) = \mathbb{R}/2\pi\mathbb{Z}$ values on cube c and vertex v, respectively; note the importance of the integrity of n. The z^a and z^b parametrize 1-form $\mathbb{Z}$ gauge invariances that reduce the a, b from real valued to effectively $U(1)$ valued, and subsequently the φ^a and φ^b parametrize the effectively $U(1)$ ordinary (0-form) gauge invariances. The action also has the following invariances under changes of background variables:

$$A_l \rightarrow A_l + 2\pi Z_l + (d\Phi)_l, \quad L_p^a \rightarrow L_p^a + q(dZ)_p, \quad \theta_v^b \rightarrow \theta_v^b - q\Phi_v, \tag{A5}$$

$$A_p^\star \rightarrow A_p^\star + 2\pi Z_p^\star + (\partial\Phi^\star)_p, \quad L_l^b \rightarrow L_l^b + p(\partial Z^\star)_l, \quad \theta_c^a \rightarrow \theta_c^a - p\Phi_c^\star, \tag{A6}$$

where Z_l and $Z_p^\star$ take integer values on link l and plaquette p, respectively, and Φ_v and $\Phi_c^\star$ take $U(1) = \mathbb{R}/2\pi\mathbb{Z}$ values on vertex v and cube c, respectively; note the importance of the integrity of p, q. The Z and $Z^\star$ parametrize 1-form $\mathbb{Z}$ gauge invariances that reduce the background $A, A^\star$ from real valued to effectively $U(1)$ valued, and subsequently the Φ and $\Phi^\star$ parametrize the effectively $U(1)$ ordinary (0-form) global symmetries. Apparently, if $A^\star$ is identified with A in the said way, then $Z^\star, \Phi^\star$ must be identified with Z, Φ in the associated manner (but the Wilson loop observables L^a and L^b are still independent).

An important feature of this *effective* Lagrangian theory is that the backgrounds A_l and $A_l + 2\pi Z_l$ are not identical outright; an associated transformation of the Wilson loop observables is required. In physical terms, this means in this theory, a narrow thread of 2π electromagnetic flux in the background is physically indistinguishable from inserting a Wilson loop observable which creates a certain anyon world loop. This is indeed a property of the original doubled Chern-Simons theory (1) in the continuum. A conceptually important point, however, is that this property is an effective, macroscopic one, in the sense that it only applies when the thread of the electromagnetic flux is "narrow" compared to the scales of interest but large compared to any microscopic scale. This is why the lattice Lagrangian theory above is only applicable to *effective* space-time lattice, by which we mean a coarse-grained space-time manifold. Were the "lattice" not a coarse grain but an actual *microscopic* lattice, the description is inapplicable because a 2π magnetic flux that is narrower than the microscopic lattice scale should be invisible outright, i.e., the backgrounds A_l and $A_l + 2\pi Z_l$ should be identical outright. It is this important difference in the physical requirements between a theory on an effective lattice (coarse grain) and one on an actually microscopic lattice that makes the usual wisdom [5,8] which generates toy model Hamiltonians on microscopic lattice from coarse-grained descriptions (fixed-point properties) not directly applicable to our present problem [11].

Let us nonetheless proceed and obtain the Hilbert space and operator contents from this effective Lagrangian. For the cubic lattice, we view two directions as the spatial lattice and one direction as the discretized time. (We may do the same for a prism decomposition of the space-time, where the space is triangulated and the time is discretized.) Then a three-dimensional vertex v is associated with a spatial vertex $\mathbf{v}$ and an integer time step t, denoted as $v = (\mathbf{v}, t)$, while (the center of) a three-dimensional cube $c = (\mathbf{p}, t + 1/2)$; on the other hand, a three-dimensional link l has two possibilities $l = (\mathbf{l}, t)$ or $l = (\mathbf{v}, t + 1/2)$, and a three-dimensional plaquette has two possibilities $p = (\mathbf{p}, t)$ or $p = (\mathbf{l}, t + 1/2)$. In these notations, our previous action reads as $S = \sum_t L_t$, with

$$\begin{aligned}
L_t &= \frac{n}{2\pi}\sum_{\mathbf{l}} a_{\mathbf{l},t+1/2}(b_{\mathbf{l},t+1} - b_{\mathbf{l},t}) - \sum_{\mathbf{p}}\theta_{\mathbf{p},t+1/2}^a(s_{\mathbf{p},t+1}^b - s_{\mathbf{p},t}^b) + \sum_{\mathbf{v}}s_{\mathbf{v},t+1/2}^a(\theta_{\mathbf{v},t+1}^b - \theta_{\mathbf{v},t}^b) \\
&\quad + \sum_{\mathbf{l}}s_{\mathbf{l},t}^a[d\theta^b - nb + qA]_{\mathbf{l},t} + \sum_{\mathbf{l}}s_{\mathbf{l},t+1/2}^b[\partial\theta^a - na + pA^\star]_{\mathbf{l},t+1/2} \\
&\quad + \sum_{\mathbf{p}}\frac{a_{\mathbf{p},t}}{2\pi}[n(db - 2\pi s^b) - qdA + 2\pi L^a]_{\mathbf{p},t} + \sum_{\mathbf{v}}\frac{b_{\mathbf{v},t+1/2}}{2\pi}[n(\partial a - 2\pi s^a) - pdA^\star + 2\pi L^b]_{\mathbf{v},t+1/2},
\end{aligned} \tag{A7}$$

where for simplicity we have assumed that $A, A^\star$ only have magnetic fields but no electric fields, and L^a, L^b only run along the time direction, i.e., the anyon insertions created by inserting these Wilson loops are held at fixed positions in the space. Only the three terms in the first line involve discretized time derivative; as usual, they give rise to the local Hilbert spaces endowed with the commutation relations (2), (3), and (4), as long as we introduce the spatial dual lattice notions as explained there. On the other hand, the four terms in the last two lines all involve Lagrange multipliers, so instead of Hamiltonian terms, they give rise to strict constraints on the Hilbert space. In particular, summing over the integer valued $s_{\mathbf{l},t}^a$ and $s_{\mathbf{l},t+1/2}^b$ imposes the Gauss's law constraints for the 1-form $\mathbb{Z}$ gauge transformations (6), generated by (7), while integrating over the real valued $a_{\mathbf{p},t}$ and $b_{\mathbf{v},t+1/2}$ imposes the

Gauss's law constraints for the ordinary (0-form) gauge transformations, generated by (5):

$$g_{\mathbf{l}}^b = e^{-iqA_{\mathbf{l}}}, \quad g_{\mathbf{l}^\star}^a = e^{-ipA_{\mathbf{l}^\star}^\star}; \quad f_{\mathbf{p}}^b = \frac{q}{n}(dA)_{\mathbf{p}} - \frac{2\pi}{n}L_{\mathbf{v}^\star=\mathbf{p}}^a,$$

$$f_{\mathbf{p}^\star}^a = \frac{p}{n}(d^\star A^\star)_{\mathbf{p}^\star} - \frac{2\pi}{n}L_{\mathbf{v}=\mathbf{p}^\star}^b. \tag{A8}$$

The commutation relations along with these Gauss's law constraints are the full content of the theory; there is no Hamiltonian on top of the subspace specified by these constraints.

While this theory is well defined, there are two undesired features, given that our goal is to construct microscopic lattice (toy) model Hamiltonians:

(i) Again, a 2π background magnetic flux $(dA)_{\mathbf{p}}$ through a single plaquette $\mathbf{p}$ is not invisible, but only equivalent to an

$L_{\mathbf{p}}^a$ insertion taking value $-q$. Therefore, the "spatial lattice" here must be viewed as a coarse grain rather than an actual microscopic spatial lattice. We are interested in theories (albeit toy theories) on microscopic spatial lattice, in which a 2π background magnetic flux $(dA)_{\mathbf{p}}$ through a single plaquette $\mathbf{p}$ is invisible outright.

(ii) The current theory is a lattice gauge theory with strict Gauss's law constraints imposed on the physical Hilbert space. We are interested in theories with no strict constraints on Hilbert space, and any appearance of gauge constraint should be emergent at low energies [4,5].

Therefore, the remaining task is to modify the theory to evade these two issues.

To resolve the first issue, partly motivated by [21], in L_t we may strip off the direct coupling of the backgrounds $A, A^\star$ to the real valued a, b, so that the backgrounds only couple to integer-valued variables s^a, s^b, and hence their $2\pi\mathbb{Z}$ parts indeed do not matter, and they are indeed U(1); as a result of this, the f^b, f^a constraints in (A8) can no longer be gauge constraints, but they can emerge as energetic conditions. Consider the lattice gauge theory with two 1-form $\mathbb{Z}$ Gauss's law constraints and a Hamiltonian:

$$g_{\mathbf{l}}^b = e^{-iqA_{\mathbf{l}}}, \qquad g_{\mathbf{l}^\star}^a = e^{-ipA_{\mathbf{l}^\star}^\star}$$

$$H_{gauge} = \frac{V_b}{2}\sum_{\mathbf{p}}\left(f_{\mathbf{p}}^b\right)^2 + \frac{V_a}{2}\sum_{\mathbf{v}}\left(f_{\mathbf{v}}^a\right)^2. \qquad (A9)$$

In this theory, the background fields $A_{\mathbf{l}}$ and $A_{\mathbf{l}^\star}^\star$ are indeed U(1) valued as desired. [In the above we omitted electric field for convenience; it can be shown that the theory (A9) stays the same even if we include electric field via the time dependence of A and $A^\star$.] Moreover, under the Gauss's law constraints, minimizing the Hamiltonian H_{gauge} will indeed lead to the f^a, f^b constraints in (A8), with $L_{\mathbf{v}^\star=\mathbf{p}}^a$ given by the integer closest to $q(dA)_{\mathbf{p}}$ and $L_{\mathbf{v}=\mathbf{p}^\star}^b$ given by the integer closest to $p(d^\star A^\star)_{\mathbf{p}^\star}$; this means large enough magnetic fluxes can create anyon insertions, and one reminiscence of this fact is the aforementioned macroscopic effective indistinguishably between a finite-size 2π background magnetic flux and a certain Wilson loop insertion.

The theory (A9) is, however, still a gauge theory. To resolve this second issue, we recall that in the previous studies of exactly solvable models, gauge constraints are energetically imposed [4,5]. This motivates us to view (A9) as the $U_a, U_b \to \infty$ limit of the "prototype" Hamiltonian $\tilde{H}$ introduced in Sec. II. This is how we motivate for $\tilde{H}$. Upon making U_a, U_b any finite values, however, the local gaplessness problem arises, as explained in Sec. II. This problem occurs here but not in [4,5] because the local operators here take continuous rather than discrete values, in order to accommodate for suitable couplings to the continuous background U(1) gauge field(s) A and $A^\star$. The resolution to this local gaplessness problem finally led us to our construction of the Hamiltonian H.

To further understand how the discussion so far is related to the Kapustin-Fidkowski no-go theorem, let us first review how, when either p or q vanishes (and hence the Hall conductivity vanishes), one may construct an exactly solvable lattice Hamiltonian [21] that resolves the two issues above. For concreteness we take $q = 0$. Then, on each link $\mathbf{l}$, instead of the

real-valued operator $b_{\mathbf{l}}$ subjected to the constraint $g_{\mathbf{l}}^b = 1$ in (A8), we can use a finite local Hilbert space endowed with

$$\tilde{b}_{\mathbf{l}} \in \{0, 1, \ldots, n-1\}, \qquad [\tilde{b}_{\mathbf{l}}, e^{-i\tilde{a}_{\mathbf{l}^\star}}] = e^{-i\tilde{a}_{\mathbf{l}^\star}} \quad \text{except}$$

$$e^{-i\tilde{a}_{\mathbf{l}^\star}}|\tilde{b}_{\mathbf{l}} = n-1\rangle = 0, \quad e^{i\tilde{a}_{\mathbf{l}^\star}}|\tilde{b}_{\mathbf{l}} = 0\rangle = 0. \qquad (A10)$$

We may view $\tilde{b}_{\mathbf{l}}$ as a reminiscence of $(nb - d\theta^b)_{\mathbf{l}}/2\pi$. Since the local Hilbert space is now finite, the other constraints in (A9) can be energetically imposed by the local commuting gapped Hamiltonian

$$H_{\text{comm}} = \frac{V}{2}\sum_{\mathbf{p}}(d\tilde{b} - ns)_{\mathbf{p}}^2 + U\sum_{\mathbf{v}=\mathbf{p}^\star}$$

$$\times\left[|1 - e^{i(d^\star\tilde{a})_{\mathbf{p}^\star}}|^2 \prod_{\mathbf{l}^\star\in\partial\mathbf{p}^\star}|1 - e^{i(d^\star\theta - n\tilde{a} + pA^\star)_{\mathbf{l}^\star}}|^2\right],$$

$$(A11)$$

where $s_{\mathbf{p}}, \theta_{\mathbf{v}^\star=\mathbf{p}}$ are reminiscent of $s_{\mathbf{p}}^b, \theta_{\mathbf{v}^\star=\mathbf{p}}^a$, while $s_{\mathbf{p}^\star}^a$ and $\theta_{\mathbf{v}=\mathbf{p}^\star}^b$ are gone; note that the $e_{\mathbf{l}^\star}^{\mp in\tilde{a}}$ in the last factor is nonvanishing only if it is multiplied to an $e^{\pm i\tilde{a}_{\mathbf{l}^\star}}$ from the previous $e^{\pm i(d^\star\tilde{a})_{\mathbf{p}^\star}}$ factor. When $p \neq 0$ and $n \neq 1$, we may solve for fractionally charged anyons [21]; when $p = 0$, we may further simplify the model to a $\mathbb{Z}_n$ toric code. Recently in [24] it was emphasized that the topological low-energy physics of this model does not involve states with large values of $|s_{\mathbf{p}}|$, so the local Hilbert space on $\mathbf{p}$ can also be trimmed to finite, making all the local Hilbert spaces finite dimensional (in fact [24] encompassed more general cases including twisted bosonic topological orders, as long as the Hall conductivity vanishes).

Apparently, such exactly solvable Hamiltonian becomes unavailable when both p, q are nonzero, i.e., when the Hall conductivity does not vanish, in echo with the Kapustin-Fidkowski no-go theorem [10]. In particular, the background fields must couple to the system only through the Hamiltonian, but not the eigenvalues of any local microscopic operator, so we may no longer define some $\tilde{b}_{\mathbf{l}}$ as a reminiscence of $(nb - d\theta^b)_{\mathbf{l}}/2\pi$ under the constraint $g_{\mathbf{l}}^b = e^{-iqA_{\mathbf{l}}}$ when $q \neq 0$. On the other hand, we may not define $\tilde{b}$ as a reminiscence of $(nb - d\theta^b - qA)_{\mathbf{l}}/2\pi$ either because in that case the first term of H_{comm} would involve $(d\tilde{b} - ns - qdA)^2$, violating the invisibility requirement of a 2π background flux; we may not use $\cos(d\tilde{b} - ns - qdA)$ either because then the s operator which importantly carries electric charge (when $p \neq 0$) would have dropped out. This is why we emphasize [11] that the usual procedure to obtain exactly solvable lattice Hamiltonian from an exactly solvable-coarse grained effective theory becomes unsuccessful here for general values of couplings p, q, due to the difference in the physical requirements satisfied by a 2π background flux on a coarse-grained "lattice" versus on an actual microscopic lattice.

APPENDIX B: DETAILS FOR LOCAL LOW-ENERGY SUBSPACE

The local Hamiltonian on each link can be expressed as (the link index is omitted in this section)

$$H_{\text{link}} = \frac{1}{2}(\epsilon_b b^2 + \epsilon_a a^2) - U_b\cos[n(b - \bar{b}_0)]$$

$$- U_a\cos[n(a - \bar{a}_0)], \qquad (B1)$$

where we remind the commutation relation is $[b, a] = i2\pi/n$, and $\bar{a}_0$, $\bar{b}_0$ are terms that commute with a and b; the electromagnetic background couples to the system through $\bar{a}_0, \bar{b}_0$. The task of this section is to perturbatively find the low-energy subspace of H_{link} under the assumption $\epsilon \ll U$ (where ϵ is the scale of ϵ_a and ϵ_b, and U the scale of U_a and U_b). The low-energy subspace turns out to be an emergent $\mathbb{Z}_n$ space with nearly degenerate energies. More particularly, we will do the following:

(i) construct n trial wave functions that are nearly orthogonal up to error exponentially small in $\sqrt{U/\epsilon}$, and denote the space spanned by them as $\mathcal{T}$;

(ii) show there is large energy gap $O(\sqrt{\epsilon U})$ between when H_{link} is projected into $\mathcal{T}$ and into the orthogonal subspace $\mathcal{T}_\perp$;

(iii) show H_{link} projected into $\mathcal{T}$ is nearly proportional to the identity matrix, up to error $O(\epsilon)$, which is smaller by $O(\sqrt{\epsilon/U})$ compared to the large gap $O(\sqrt{\epsilon U})$;

(iv) show H_{link} mixes $\mathcal{T}$ and $\mathcal{T}_\perp$ at order $O(\epsilon)$, again smaller by $O(\sqrt{\epsilon/U})$ compared to the large gap $O(\sqrt{\epsilon U})$.

This means H_{link} has an actual low-energy subspace $\mathcal{L}$ that is n dimensional, well separated in energy from $\mathcal{L}_\perp$, and the energy split within $\mathcal{L}$ is $O(\epsilon)$ which is $O(\sqrt{\epsilon/U})$ compared to the gap $O(\sqrt{\epsilon U})$; moreover, the error between the actual $\mathcal{L}$ and the trial $\mathcal{T}$ is $O(\sqrt{\epsilon/U})$.

To motivate our construction of the trial wave functions, first we only take ϵ_a and U_b terms into consideration,

$$H_{\text{link},b} = \frac{1}{2}\epsilon_a a^2 - U_b \cos[n(b - \bar{b}_0)], \quad (B2)$$

which analogously describe the one-dimensional, quantum mechanical motion of a particle in a sinusoidal potential. As long as the potential is deep enough, the neighborhood around each minimum, located at $\bar{b}_j = \frac{2\pi}{n}j + \bar{b}_0$ with $j \in \mathbb{Z}$, can be effectively described by the Taylor expansion to the quadratic order. Therefore, the low-energy eigenfunctions are approximately those of infinitely many harmonic oscillators at the potential minima, the approximate Hamiltonians of which can be written as

$$H_j \approx \frac{1}{2}\epsilon_a a^2 + \frac{1}{2}U_b n^2 (b - \bar{b}_j)^2 = \omega_b\left(c_j^\dagger c_j + \frac{1}{2}\right) \quad (B3)$$

with excitation energy

$$\omega_b = 2\pi\sqrt{U_b \epsilon_a} \quad (B4)$$

and standard construction of ladder operators c_j. The solution of the low-energy states is

$$|\phi_j^{(N_b)}\rangle \equiv \int db\, C^{(N_b)} H^{(N_b)}\left(\frac{b - \bar{b}_j}{W_b}\right) e^{-(b-\bar{b}_j)^2/2W_b^2}|b\rangle, \quad (B5)$$

where N_b represent the energy level, $C^{(N_b)} = \frac{1}{\pi^{1/4}\sqrt{W_b}\sqrt{2^{N_b}N_b!}}$ is a normalization factor, $H^{(N_b)}(x)$ is the N_bth order Hermite polynomial, and

$$W_b = \frac{\sqrt{2\pi}}{n}\left(\frac{\epsilon_a}{U_b}\right)^{\frac{1}{4}} \quad (B6)$$

is the half-width of each state. For finding the low-energy subspace we only need $N_b = 0$, and for controlling errors we will only include small values of N_b in later considerations.

To justify our negligence of the higher-order terms in the Taylor expansion of cosine function at each minimum, and our elimination of the overlap between different minimums' orbitals, we will need the width of each state to be much smaller than the distance between two minima of the potential. We thus reach our first "separation of length scale" condition of the solvable limit:

$$W_b \ll \frac{2\pi}{n} \Rightarrow \frac{\epsilon_a}{U_b} \ll (2\pi)^2. \quad (B7)$$

Near this limit, we can estimate the correction of a small but finite W_b to the energy gap. First, the higher order of the expansion of the potential $\{U_b(1 - \cos[n(b - \bar{b}_j)])\} \approx U_b[\frac{n^2(b-\bar{b}_j)^2}{2!} - \frac{n^4(b-\bar{b}_j)^4}{4!} + \cdots]$ lead to the correction of energy gap between the ground states and the first excited states:

$$\Delta_b \equiv \langle \phi_j^{(1)}|H_{\text{link},b}|\phi_j^{(1)}\rangle - \langle \phi_j^{(0)}|H_{\text{link},b}|\phi_j^{(0)}\rangle$$
$$\approx \omega_b\left\{1 - \frac{n^2}{6}W_b^2 + O(W_b^4)\right\} \quad (B8)$$

(same for all j). This correction decreases the gap while it indeed approaches zero as $W_b \to 0$. There is also an order $O(W_b^2)$ mixing between different N_b states. Second, the overlap integral between neighboring orbits

$$t_b \equiv \langle \phi_j^{(0)}|H_{\text{link},b}|\phi_{j+1}^{(0)}\rangle \approx \omega_b\frac{\pi^2}{n^2 W_b^2}e^{-\pi^2/n^2 W_b^2} \quad (B9)$$

can lead to an energy splitting up to $4t_b$ in the infinitely degenerate ground-state space. Since t_b is exponentially suppressed, compared to other quantities of order $O(1)$ or other errors of order $O(W_b^2)$, we can safely neglect the splitting in the said limit.

Now we have a new set of basis states describing the low-energy physics of the system, spanned by infinitely many highly localized orbits $|\phi_j^{(N_b)}\rangle$. Our next step would be solving the remaining terms,

$$H_{\text{link},a} = \frac{1}{2}\epsilon_b b^2 - U_a \cos[n(a - \bar{a}_0)] \quad (B10)$$

in this basis. Note that the U_a term commutes with $H_{\text{link},b}$ and only changes $|\phi_j^{(N_b)}\rangle$ to $|\phi_{j+1}^{(N_b)}\rangle$. It remains to consider the ϵ_b term in the $|\phi_j^{(N_b)}\rangle$ basis. We approximate

$$b^2|\phi_j^{(N_b)}\rangle = \bar{b}_j^2|\phi_j^{(N_b)}\rangle + (\dots), \quad (B11)$$

where the $\bar{b}_j \equiv (2\pi/n)\hat{j} + \bar{b}_0$ operator (with $\hat{j}$ measuring the value of j), well defined at least for states $|\phi_j^{(N_b)}\rangle$ with small values of N_b, measures the center coordinates. We claim that for small values of N_b the neglected terms $(\dots)$ are suppressed under suitable assumptions of parameters; we will justify the claim in details later. This means that at low energies we no longer need to consider the continuous b variable but can concentrate on a discrete $\bar{b}$ space for each level N_b (only small N_b are of interest). The basis of the conjugate variable of $\bar{b}$ operator, $\underline{a}$ (not to be confused with $\bar{a}$), is then defined by

$$|\underline{a}^{(N_b)}\rangle \equiv \sum_j e^{i\bar{b}_j\underline{a}}|\phi_j^{(N_b)}\rangle \quad (B12)$$

and is the projection of a. The value of $\underline{a}$ is confined onto the "first Brillioun zone" in the reciprocal space of the $\bar{b}_j$ lattice,

i.e., $\underline{a} \in [0, 2\pi)$. Since the roles of $\underline{a}$ and $\bar{b}$ are conjugate, we can switch the perspective and now interpret $\underline{a}$ as a coordinate on a ring. Thus, we can regard $-U_a \cos[n(\underline{a} - \bar{a}_0)]$ term as a sinusoidal potential subjected to the periodic boundary condition identifying $\underline{a} = 0$ and $\underline{a} = 2\pi$. This potential has n minima located at $\bar{a}_{m^a} = \frac{2\pi m^a}{n} + \bar{a}_0$, $m^a \in \mathbb{Z}_n$. We can solve the Hamiltonian around each minimum, again treating the potential in the neighbors of the minima as quadratic:

$$H_{m^a} \approx \frac{1}{2}\epsilon_b \bar{b}^2 + \frac{1}{2}U_a n^2 (\underline{a} - \bar{a}_{m^a})^2 = \omega_a\left(d_{m^a}^\dagger d_{m^a} + \frac{1}{2}\right)$$
(B13)

with the standard construction of ladder operators d_m, and excitation energy

$$\omega_a = 2\pi\sqrt{\epsilon_b U_a}.$$
(B14)

Then, the eigenstates under this approximation are

$$\left|\tilde{\Psi}_{m^a}^{(N_b, N_a)}\right\rangle = \int \frac{d\underline{a}}{\sqrt{2\pi}} C^{(N_a)} H_k\left(\frac{\underline{a} - \bar{a}_{m^a}}{W_a}\right) e^{-(\underline{a} - \bar{a}_{m^a})^2/2W_a^2} |\underline{a}^{(N_b)}\rangle$$
(B15)

with half-width W_a in the a space:

$$W_a = \frac{\sqrt{2\pi}}{n}\left(\frac{\epsilon_b}{U_a}\right)^{\frac{1}{4}}.$$
(B16)

Each level labeled by (N_b, N_a) then would be n-fold degenerate, with energy $E^{(N_b, N_a)} = \omega_b(N_b + \frac{1}{2}) + \omega_a(N_a + \frac{1}{2})$. We will refer to the n states with $N_b = N_a = 0$ and $m^a = 0, 1, \ldots, n-1$ as the trial ground states.

We again need the width of those states to be much smaller than the distance between two minima of the potential, and thus reach our second "separation of length scale" condition of the solvable limit:

$$W_a \ll \frac{2\pi}{n} \Rightarrow \frac{\epsilon_b}{U_a} \ll (2\pi)^2.$$
(B17)

We can also discuss the effect of a small but nonvanishing W_a. Similar to the previous discussion, the higher order of the expansion of the potential leads to the correction of energy gap between the ground states and the first excitation states:

$$\Delta_a = \omega_a\left\{1 - \frac{n^2}{6}W_a^2 + O(W_a^4)\right\}$$
(B18)

which is unimportant in the $W_a \to 0$ limit. Again, different N_a states have an order $O(W_a)^2$ mixing. And the overlap integral between two neighboring orbits

$$t_a \equiv \left\langle \tilde{\Psi}_{m^a}^{(0,0)} \middle| H_{\text{link},a} \middle| \tilde{\Psi}_{m^a+1}^{(0,0)} \right\rangle \approx \omega_a \frac{\pi^2}{n^2 W_a^2} e^{-\pi^2/n^2 W_a^2}$$
(B19)

can lead to an exponentially small energy splitting of up to $4t_a$ among the n-fold-degenerate ground states, which can be neglected in the $W_a \to 0$ limit.

Now we come back to the approximation $b^2 \to \bar{b}^2$ in (B11) that is yet to be justified. We show this can be justified in under yet another condition. Writing $b^2 = \bar{b}_j^2 + 2\bar{b}_j(b - \bar{b}_j) + (b - \bar{b}_j)^2$, the terms neglected in (B11) are

$$\frac{1}{2}\epsilon_b(b^2 - \bar{b}^2) = \omega_a \sum_{j,m^a} iW_b W_a \frac{n}{2\pi}(c_j^\dagger + c_j)(d_{m^a}^\dagger - d_{m^a})$$

$$+ \omega_a \sum_j \frac{W_a^2 W_b^2}{2}\left(\frac{n}{2\pi}\right)^2 (c_j^\dagger + c_j)^2, \quad \text{(B20)}$$

where, as said before, the $\bar{b}$ operator and the associated $d_{m^a}, d_{m^a}^\dagger$ are well defined for small values of N_b, and we have neglected the terms that are exponentially suppressed in $1/W_b^2$ since we are justifying the error being polynomial in W_b^2 or smaller. Note that although we used multiple ladder operators to express the original operator, it should be kept in mind that the problem we are considering here is just a single-body problem. It is thus remarkable that these terms cannot mix different states in the degenerate ground-state manifold spanned by $\tilde{\Psi}_{m^a}^{(N_b, N_a)}$, so they can at most modify the size of the excitation gaps and the form of the trial ground states. For N_b and N_a of order 1, the matrix elements of c, $c^\dagger$ and d, $d^\dagger$ are of order 1. There are several types of terms: First, the cc, cd, $c^\dagger d$, $cd^\dagger$ terms annihilate the trial ground states. Second, the $c^\dagger c$ terms only positively modify the excitation gap Δ_b (by the same amount for all j's), and hence unimportant. Third, the $c^\dagger d^\dagger$ mix the trial ground states $(N_b, N_a) = (0, 0)$ with $(1,1)$; but since W_a, W_b are small, the matrix element of order $\omega_a W_b W_a$ is indeed small compared to the energy difference $\omega_b + \omega_a$ between $(0, 0)$ and $(1, 1)$. Finally, there is a $c^\dagger c^\dagger$ term with matrix element $\omega_a W_a^2 W_b^2 (n/2\pi)^2/2$; to justify our approximation this term must be small compared to the energy difference $2\omega_b$ between $(0,0)$ and $(2,0)$. This leads to the last "separation of energy scale" condition of the solvable limit:

$$\omega_a \frac{W_a^2 W_b^2}{2}\left(\frac{n}{2\pi}\right)^2 \ll 2\omega_b \Rightarrow \frac{\epsilon_b}{U_b} \ll 4n^2.$$
(B21)

In summary, in a solvable limit where the three conditions (B7), (B17), and (B21) are all satisfied, we can perturbatively solve for the nearly degenerate ground states of the link Hamiltonian (B1). Since we have four independent parameters, this limit is always reachable. We find the n-fold trial ground states

$$\left|\tilde{\Psi}_{m^a}^{(0,0)}\right\rangle = \sqrt{\frac{W_a}{\pi W_b}} \sum_{j\in\mathbb{Z}} e^{i\frac{n}{2\pi}\bar{b}_j \bar{a}_{m^a}} e^{-\bar{b}_j^2(n/2\pi)^2 W_a^2/2}$$

$$\times \int db\, e^{-(b-\bar{b}_j)^2/2W_b^2}|b\rangle$$
(B22)

with $\bar{b}_j = (2\pi/n)j + \bar{b}_0$ and $a_{m^a} = (2\pi/n)m^a + \bar{a}_0$, $m^a \in \mathbb{Z}_n$; note that when m^a shifts by n, the wave function changes by an overall phase that depends on $\bar{b}_0$, which commutes with the link operators b, a. We can also take the linear combination that is more localized in the b basis:

$$\left|\Psi_{m^b}^{(0,0)}\right\rangle \equiv \frac{1}{\sqrt{n}} \sum_{m^a=0}^{n-1} e^{-i\frac{n}{2\pi}\bar{a}_{m^a}\bar{b}_{m^b}} \left|\Psi_{m^a}^{(0,0)}\right\rangle$$

$$= \sqrt{\frac{W_a}{n\pi W_b}} \sum_{z^b\in\mathbb{Z}} e^{inz^b \bar{a}_0} e^{-\bar{b}_{m^b+nz^b}^2(n/2\pi)^2 W_a^2/2}$$

$$\times \int db\, e^{-(b-\bar{b}_{m^b+nz^b})^2/2W_b^2}|b\rangle.$$
(B23)

Again, m_b is $\mathbb{Z}_n$ valued in the sense that, when m_b shifts by n, the wave function changes by an overall phase that depends

on $\bar{a}_0$, which commutes with the link operators b, a. With two small widths, $W_b = \frac{\sqrt{2\pi}}{n}(\frac{\epsilon_a}{U_b})^{\frac{1}{4}}$, $W_a = \frac{\sqrt{2\pi}}{n}(\frac{\epsilon_b}{U_a})^{\frac{1}{4}}$, we can separate neighboring Gaussian wave packets in either the a or b basis. Those Gaussian envelopes are very narrow and resemble delta function in one space but are very broad in the other basis. Illustrations of $\Psi_{m^b}^{(0,0)}$ in b basis can be found in Fig. 3. Denoting the n-dimensional subspace spanned by the trial ground-state wave functions as $\mathcal{T}$, the true n-fold ground subspace $\mathcal{L}$ of H_{link} differs from $\mathcal{T}$ by a bounded error $O(W_b^2, W_a^2, W_a W_b)$, and the split among the actual ground states in $\mathcal{L}$ is bounded by the same order compared to the gap with $\mathcal{L}_\perp$, which is of order $\min(\omega_a, \omega_b)$.

Note that we can also express the trial ground states in the a basis:

$$|\tilde{\Psi}_{m^a}^{(0,0)}\rangle = \sqrt{\frac{W_b}{n\pi W_a}} \sum_{z^a \in \mathbb{Z}} \int da\, e^{-i\frac{n}{2\pi}\bar{b}_0(a-\bar{a}_{m^a})} e^{-W_b^2(n/2\pi)^2 a^2/2}$$
$$\times e^{-(a-\bar{a}_{m^a+nz^a})^2/2W_a^2}|a\rangle \qquad (B24)$$

and

$$|\Psi_{m^b}^{(0,0)}\rangle = \sqrt{\frac{W_b}{\pi W_a}} \sum_{j \in \mathbb{Z}} \int da\, e^{-i\frac{n}{2\pi}\bar{b}_{m^b}a} e^{-W_b^2(n/2\pi)^2 a^2/2}$$
$$\times e^{-(a-\bar{a}_j)^2/2W_a^2}|a\rangle. \qquad (B25)$$

Note that these expressions are not entirely symmetric with those in the b basis. This is because we started with $H_{\text{link},b}$ and only separated the excitation energy in $H_{\text{link},b}$ from the perturbation strength. If we also impose a constraint similar to (B21), $\epsilon_a/U_a \ll 4n^2$, we would be able to neglect the difference between those different forms and conveniently write the trial ground states in the form of (B23) or (B24) for both Ψ_{m^b} in the b basis and $\tilde{\Psi}_{m^a}$ in the a basis. Let us simplify the discussion by taking

$$\epsilon_b = \epsilon_a = \epsilon, \quad U_a = U_b = U, \quad W \equiv \frac{\sqrt{2\pi}}{n}\left(\frac{\epsilon}{U}\right)^{\frac{1}{4}} \to 0 \qquad (B26)$$

while fixing $\Delta_0 = 2\pi\sqrt{U\epsilon}$. The asymmetry between the trial wave-function expressions in the b basis and the a basis is of order $O(W^2)$ (where W is the width of both types of Gaussian pockets), which is indeed the order that our trial ground subspace differs from the actual ground subspace, and therefore there is no contradiction with apparent symmetry between a and b.

APPENDIX C: JUSTIFICATION FOR PROJECTING MANY-BODY COUPLING TERMS INTO LOCAL LOW-ENERGY SUBSPACE

After obtaining the local low-energy subspace for H_{link} on each link, our next step of solving the entire Hamiltonian is to project the V_a and V_b terms into the nearly degenerate subspace formed by tensoring these local low-energy spaces. At first sight this seems problematic because the V_a, V_b terms, proportional to $(d^\star a - 2\pi s^a)_{\mathbf{p}^\star}^2$ and $(db - 2\pi s^b)_{\mathbf{p}}^2$, respectively, are unbounded, and hence it seems the mixing they cause between the local low-energy states and the local excited states (the gaps are of or greater than order $\sqrt{\epsilon U}$) would be unboundedly

large even if we assumed $V \ll \sqrt{\epsilon U}$. Now, we show this is not the case.

First of all, the remaining steps to solve for the low-energy many-body topological physics does not rely on the V_a, V_b terms taking this particular quadractic expression. They can be modified into other functions $F[(d^\star a - 2\pi s^a)_{\mathbf{p}^\star}^2]$ and $F[(db - 2\pi s^b)_{\mathbf{p}}^2]$, where F is a monotonically increasing function but bounded above when the argument is large, e.g., $F[x] = \xi[1 - \exp(-x/\xi)]$ with arbitrary positive ξ. Hence, the potential problem is immediately circumvented.

Even with our current choice of the quadratic expression, we can show the undesired mixing is indeed controlled by $V/\sqrt{\epsilon U} \ll 1$ as desired. It suffices to focus on the V_b term since any argument would work for the V_a term in a similar manner. To prove this claim, we first note that the sufficient condition of treating V_b term as a perturbation (compared to the energy scales in the link local Hilbert space) is to require that, for any excited state $|\Psi_k\rangle$ and any low-energy state $|\Psi_0\rangle$ of local Hamiltonian $\sum_{\mathbf{l}} H_{\text{link},\mathbf{l}}$ to satisfy

$$\left|\langle\Psi_k|V_b(db - 2\pi s^a)_{\mathbf{p}}^2|\Psi_0\rangle\right| \ll |E_k - E_0|. \qquad (C1)$$

We now take a specific $|\Psi_0\rangle = |\{m_{\mathbf{l}}^b\}, \{s_{\mathbf{p}}^b\}, \{\theta_{\mathbf{v}}^b\}\rangle$, whose explicit expression is given in Eq. (29). Applying the V_b term yields

$$(db - 2\pi s^b)_{\mathbf{p}}^2 \left|\{m_{\mathbf{l}}^b\}, \{s_{\mathbf{p}}^b\}, \{\theta_{\mathbf{v}}^b\}\rangle\right.$$
$$= \sum_{\{z_{\mathbf{l}}^b\}} \int_{\{b_{\mathbf{l}}\}} \left[d(b - \bar{b}_{m^b, z^b}) + \frac{2\pi}{n}\left(\left[: \frac{qdA}{2\pi} :\right] - v^b\right)\right]_{\mathbf{p}}^2$$
$$\times \left(\prod_{\mathbf{l}} e^{iz_{\mathbf{l}}^b \cdot (pA^\star)_{\mathbf{l}^\star}} e^{-\bar{b}_{m_{\mathbf{l}}^b, z_{\mathbf{l}}^b}^2 (n/2\pi)^2 W_a^2/2} e^{-(b_{\mathbf{l}} - \bar{b}_{m_{\mathbf{l}}^b, z_{\mathbf{l}}^b})^2/2W_b^2}\right)$$
$$\bigotimes_{\mathbf{l},\mathbf{p},\mathbf{v}} |b_{\mathbf{l}}\rangle|(s^b + dz^b)_{\mathbf{p}}\rangle|\theta_{\mathbf{v}}^b\rangle, \qquad (C2)$$

where v^b is the number of anyonic excitaitons defined in Eq. (34), and $[: x :] \equiv x - [x]$ always has magnitude that is less than $\frac{1}{2}$. Then we recognize that, up to an exponentially small error, $b_{\mathbf{l}} - \bar{b}_{m_{\mathbf{l}}^b, z_{\mathbf{l}}^b} = \frac{2\pi}{n}\frac{W_b}{\sqrt{2}}(c_{\mathbf{l}, m_{\mathbf{l}}^b + nz_{\mathbf{l}}^b}^\dagger + c_{\mathbf{l}, m_{\mathbf{l}}^b + nz_{\mathbf{l}}^b})$ with $c_{\mathbf{l}, j}^\dagger$ the ladder operators introduced in Eq. (B3) on link $\mathbf{l}$. Therefore, the effect of V_b operator on this state can be described as

$$(db - 2\pi s^b)_{\mathbf{p}}^2$$
$$\to \left(\frac{2\pi}{n}\right)^2 \left[\frac{W_b}{\sqrt{2}} d(c^\dagger + c) + \left(\left[: \frac{qdA}{2\pi} :\right] - v^b\right)\right]_{\mathbf{p}}^2, \qquad (C3)$$

where $c_{\mathbf{l}}^\dagger \equiv \sum_j c_{\mathbf{l}, j}^\dagger$ change the excitation number N_b of $H_{\text{link},b}$ at $\mathbf{l}$ [defined in Eq. (B2)] by 1 so that the energy is changed by ω_b. Since $([: \frac{qdA}{2\pi} :] - v^b)_{\mathbf{p}}$ will be an $O(1)$ number for the final solution we are interested in, the amplitude of perturbing this $|\Psi_0\rangle$ out of the low-energy subspace of $\sum_{\mathbf{l}} H_{\text{link},\mathbf{l}}$ will be $\sim V_b W_b \sim V_b(\epsilon_a/U_b)^{1/4}$, which is small compared to the local energy gap $\omega_b \sim \sqrt{\epsilon_a U_b}$. The above description also suggests that the eigenvalues of V_b term has a systematic error $2V_b W_b^2(2\pi/n)^2$ originating from $c_{\mathbf{l}} c_{\mathbf{l}}^\dagger$ terms, relative to the expression in Eq. (32). This constant shift of energy does not affect any discussion in the main text.

APPENDIX D: DETAILS FOR MANY-BODY GROUND STATES AND EXCITATIONS

Following the solving procedure towards the many-body problem in Sec. III B, we need to consider the equivalent con-

figurations generated by the $\mathbb{Z}/n\mathbb{Z}$ reminiscence of the 0-form $\mathbb{R}/2\pi\mathbb{Z}$ gauge invariance (8) rescaled by $n/2\pi$ [combined with (6)], which do not alter the conditions (34) and (35) with given values of v^b and v^a. The explicit transformations, labeled by $t^b_{\mathbf{v}} \in \mathbb{Z}_n$ and $t^a_{\mathbf{v}^\star} \in \mathbb{Z}_n$, respectively, are

$$\mathsf{T}(\{t^b_{\mathbf{v}}\})\big|\{m^b_{\mathbf{l}}\},\{s^b_{\mathbf{p}}\},\{\theta^b_{\mathbf{v}}\}\big\rangle \equiv e^{-i\sum_{\mathbf{l}}\lfloor\frac{m^b+dt^b}{n}\rfloor_{\mathbf{l}}\cdot(pA^\star)_{\mathbf{l}^\star}}\bigg|\Big\{n\Big\lfloor:\frac{m^b+dt^b}{n}:\Big\rfloor\Big\}_{\mathbf{l}}\Big\},\Big\{\Big(s^b-d\Big\lfloor\frac{m^b+dt^b}{n}\Big\rfloor\Big)_{\mathbf{p}}\Big\},\{\theta^b_{\mathbf{v}}\}\bigg\rangle \tag{D1}$$

$$\mathsf{T}^\star(\{t^a_{\mathbf{v}^\star}\})\big|\{m^a_{\mathbf{l}^\star}\},\{\theta^a_{\mathbf{v}^\star}\},\{s^a_{\mathbf{p}^\star}\}\big\rangle \equiv e^{i\sum_{\mathbf{l}}(qA)_{\mathbf{l}}\cdot\lfloor\frac{m^a+d^\star t^a}{n}\rfloor_{\mathbf{l}^\star}}\bigg|\Big\{n\Big\lfloor:\frac{m^a+d^\star t^a}{n}:\Big\rfloor_{\mathbf{l}^\star}\Big\},\{\theta^a_{\mathbf{v}^\star}\},\Big\{\Big(s^a-d^\star\Big\lfloor\frac{m^a+d^\star t^a}{n}\Big\rfloor\Big)_{\mathbf{p}^\star}\Big\}\bigg\rangle, \tag{D2}$$

where we define $\lfloor x\rfloor$ to be the nearest integer that is not larger than x, and $\lfloor: x:\rfloor \equiv x - \lfloor x\rfloor$. For integer x, $n\lfloor:\frac{x}{n}:\rfloor$ simply means $x \bmod n$ in the range $\{0, 1, \ldots, n-1\}$. Formally, these transformations are equivalent to substituting $\{\theta\} \to \{\theta + 2\pi t\}$. We claim that the suitable linear combinations of all gauge-equivalent states that simultaneously diagonalize H_a and H_b are $|C\rangle$ and $|C^\star\rangle$ given in Eqs. (38) and (39). These states are manifestly the eigenstates of H_b and H_a, respectively, but we need to show that they are actually simultaneous eigenstates of both terms. To verify this, we show $|C\rangle$ is a linear combination of all $|C^\star\rangle$. We express $|C\rangle$ in the m_a basis:

$$|C\rangle = \sum_{\{t^b_{\mathbf{v}}\},\{m^a_{\mathbf{l}^\star}\},\{s^a_{\mathbf{p}^\star}\}} \prod_{\mathbf{l},\mathbf{v},\mathbf{p}} \int d\theta^b_{\mathbf{v}}\, d\theta^a_{\mathbf{v}^\star}\, e^{i\left(s^{b,C\,rep}-d\lfloor\frac{m^{b,C\,rep}+dt^b}{n}\rfloor\right)_{\mathbf{p}}\cdot\theta^a_{\mathbf{v}^\star}}\, e^{-i\theta^b_{\mathbf{v}}\cdot(s^a)_{\mathbf{p}^\star}}$$

$$\times \exp\left\{-i\frac{2\pi}{n}\left(n\Big\lfloor:\frac{m^{b,C\,rep}+dt^b}{n}:\Big\rfloor + \frac{d\theta^b+qA}{2\pi}\right)_{\mathbf{l}}\cdot\left(m^a+\frac{d^\star\theta^a+pA^\star}{2\pi}\right)_{\mathbf{l}^\star}\right\}$$

$$\times \exp\left\{\frac{-i(2\pi t^b_{\mathbf{v}}+\theta^b_{\mathbf{v}})\cdot\left(\big[:\frac{pd^\star A^\star}{2\pi}:\big]-v^a\right)_{\mathbf{p}^\star}}{n} - i\Big\lfloor\frac{m^{b,C\,rep}+dt^b}{n}\Big\rfloor_{\mathbf{l}}\cdot(pA^\star)_{\mathbf{l}^\star}\right\}\big|\{m^a_{\mathbf{l}^\star}\},\{\theta^a_{\mathbf{v}^\star}\},\{s^a_{\mathbf{p}^\star}\}\big\rangle. \tag{D3}$$

We first treat the summation over $t^b_{\mathbf{v}} \in \mathbb{Z}_n$ on each vertex $\mathbf{v}$:

$$\sum_{t^b_{\mathbf{v}}} \exp\left\{i\frac{2\pi}{n}t^b_{\mathbf{v}}\cdot\left(d^\star m^a + \Big[\frac{pd^\star A^\star}{2\pi}\Big] + v^a\right)_{\mathbf{p}^\star}\right\} \tag{D4}$$

which yields the constraint $(1/n)(d^\star m^a + [\frac{p\phi^\star}{2\pi}] - v^a)_{\mathbf{p}^\star} \in \mathbb{Z}$ on each dual plaquette. Next we do the integral over $\theta^b_{\mathbf{v}} \in (-\pi, \pi]$ on each vertex $\mathbf{v}$:

$$\int d\theta^b_{\mathbf{v}} \exp\left\{i\theta^b_{\mathbf{v}}\cdot\left(\frac{(d^\star m^a + [\frac{pd^\star A^\star}{2\pi}] + v^a)_{\mathbf{p}^\star}}{n} - s^a_{\mathbf{p}^\star}\right)\right\} \tag{D5}$$

which complete the constraint (35). Then, we can rewrite the summation over $\{m^a_{\mathbf{l}^\star}\}$ and $\{s^a_{\mathbf{p}^\star}\}$, as a combined summation over both topological class $C^\star$ and the gauge transformations $\{t^a_{\mathbf{v}^\star}\}$ on the representative configurations $\{m^{a,C^\star\,rep}_{\mathbf{l}^\star}\}$ and $\{s^{a,C^\star\,rep}_{\mathbf{p}^\star}\}$. We thus reach the linear combination

$$|C\rangle = \sum_{C^\star} \prod_{\mathbf{l}} e^{-i\frac{2\pi}{n}(m^{b,C\,rep}+\frac{qA}{2\pi})_{\mathbf{l}}\cdot(m^{a,C^\star\,rep}+\frac{pA^\star}{2\pi})_{\mathbf{l}^\star}}|C^\star\rangle \tag{D6}$$

as desired.

APPENDIX E: DETAILS FOR HALL CONDUCTIVITY AND FRACTIONALIZED ELECTRIC CHARGE

We first calculate the expectation value of the local electric charge density (10). In the presence of a background A and arbitrary anyon excitations. assembling terms in (28), (D1), and (38), we have

$$\langle C|ps^b_{\mathbf{p}}|C\rangle = \frac{1}{\mathcal{N}} \sum_{\{t^b_{\mathbf{v}}\},\{z^b_{\mathbf{l}}\}} \prod_{\mathbf{v},\mathbf{l}} \int_{-\pi}^{\pi} d\theta^b_{\mathbf{v}}\, p\left(s^{b,C\,rep} + dz^b - d\Big\lfloor\frac{m^{b,C\,rep}+dt^b}{n}\Big\rfloor\right)_{\mathbf{p}} e^{-\bar{b}^2_{n\lfloor:\frac{m^{b,C\,rep}+dt^b}{n}:\rfloor_{\mathbf{l}}\cdot z^b_{\mathbf{l}}}(n/2\pi)^2 W^2}, \tag{E1}$$

where we formally recovered a normalization factor $\mathcal{N}$. The integral over $\{\theta_{\mathbf{v}}\}$ along with the summation over $\{t^b_{\mathbf{v}}\}$, $\{z^b_{\mathbf{l}}\}$ is equivalent to a infinite integral over the $\bar{b}$ variable on every link, as long as each term of the integrand only depends on $\bar{b}$ variable on one link (a similar and detailed derivation of this can be found below in the calculation of the Berry curvature). Therefore,

the expectation can be evaluated as

$$\langle C|ps_{\mathbf{p}}^{b}|C\rangle = \frac{1}{\mathcal{N}}\prod_{\mathbf{l}}\int d\bar{b}_{\mathbf{l}} \cdot \frac{p}{n}\left(d\bar{b} + ns^{b,C\,rep} - dm^{b,C\,rep} - \frac{qdA}{2\pi}\right)_{\mathbf{p}} e^{-\bar{b}^{2}(n/2\pi)^{2}W^{2}}$$

$$= -\frac{p}{n}\frac{q(dA)_{\mathbf{p}}}{2\pi} + \frac{p}{n}\left(v^{b} + \left[\frac{qdA}{2\pi}\right]\right)_{\mathbf{p}}. \tag{E2}$$

A similar derivation gives the response on dual plaquette $\mathbf{p}_{0}^{\star}$:

$$\langle C|qs_{\mathbf{p}^{\star}}^{a}|C\rangle = -\frac{q}{n}\frac{p(d^{\star}A^{\star})_{\mathbf{p}^{\star}}}{2\pi} + \frac{q}{n}\left(v^{a} + \left[\frac{pd^{\star}A^{\star}}{2\pi}\right]\right)_{\mathbf{p}^{\star}}. \tag{E3}$$

The results are interpreted in the main text in Sec. IV.

Then we calculate the globally defined Hall conductivity. It is given by the Chern number over the $\alpha_{x}, \alpha_{y} \in [0, 2\pi n)$ space of holonomies, averaged over n^{2} visited states [22]:

$$2\pi\sigma_{H} = \frac{1}{n^{2}}\int_{0}^{2\pi n}d\alpha_{x}\int_{0}^{2\pi n}d\alpha_{y}\frac{\mathcal{B}}{2\pi}, \qquad \mathcal{B} \equiv -i\left(\left\langle\frac{\partial C_{0}}{\partial\alpha_{x}}\Big|\frac{\partial C_{0}}{\partial\alpha_{y}}\right\rangle - \left\langle\frac{\partial C_{0}}{\partial\alpha_{y}}\Big|\frac{\partial C_{0}}{\partial\alpha_{x}}\right\rangle\right). \tag{E4}$$

Importantly, the integrand is a Berry curvature $\mathcal{B}/2\pi$ and the integral is the Chern number that must be an integer. Now we evaluate this integrand. For simplicity we start with the class C_{0} in which $m_{\mathbf{l}}^{b,C\,rep} = 0$, $s_{\mathbf{p}}^{b,C\,rep} = 0$. First note the terms that give rise to the Berry curvature come from an average over $\langle\frac{\partial C_{0}}{\partial A_{\mathbf{l}^{\star}}}|\frac{\partial C_{0}}{\partial A_{\mathbf{l}}}\rangle$ (or its conjugation, depending on the orientation of $\mathbf{l}$) for all unit cells. Because of translational invariance of the problem, the calculation reduces to a single link, e.g., $\mathbf{l}_{0}$:

$$\mathcal{B} = 4\sum_{z_{\mathbf{l}_{0}}^{b}}\int_{0}^{2\pi n}d\theta_{\mathbf{v}_{1}}^{b}d\theta_{\mathbf{v}_{2}}^{b}\sum_{t_{\mathbf{v}_{1}}^{b},t_{\mathbf{v}_{2}}^{b}}\left(-\frac{pq}{n}\right)z_{\mathbf{l}_{0}}^{b}\bar{b}_{\mathbf{l}_{0}}\left(\frac{n}{2\pi}\right)^{2}W^{2}e^{-\bar{b}_{\mathbf{l}_{0}}^{2}(n/2\pi)^{2}W^{2}}/\mathcal{N}',$$

$$\bar{b}_{\mathbf{l}_{0}} \equiv 2\pi\left(z_{\mathbf{l}_{0}}^{b} + \frac{2\pi t_{\mathbf{v}_{1}}^{b} + \theta_{\mathbf{v}_{1}}^{b} - 2\pi t_{\mathbf{v}_{2}}^{b} - \theta_{\mathbf{v}_{2}}^{b}}{2\pi n}\right), \tag{E5}$$

where $\mathbf{v}_{1}$ and $\mathbf{v}_{2}$ are the two end points of link $\mathbf{l}_{0}$, $\mathcal{N}' = \frac{\sqrt{\pi}}{nW}(2\pi n)^{2}$ is a normalization factor, and we have neglected the infinitesimal background field. Then, the summation of t^{b} and the integral over θ^{b} on $[0, 2\pi)$ on each vertex can make up to an integral over θ^{b} on $[0, 2\pi n)$:

$$\mathcal{B} = 4\sum_{z_{\mathbf{l}_{0}}^{b}}\int_{0}^{2\pi n}d\theta_{\mathbf{v}_{1}}^{b}d\theta_{\mathbf{v}_{2}}^{b}\left(-\frac{pq}{n}\right)z_{\mathbf{l}_{0}}^{b}\bar{b}_{\mathbf{l}_{0}}\left(\frac{n}{2\pi}\right)^{2}W^{2}e^{-\bar{b}_{\mathbf{l}_{0}}^{2}(n/2\pi)^{2}W^{2}}/\mathcal{N}',$$

$$\bar{b}_{\mathbf{l}_{0}} \equiv 2\pi\left(z_{\mathbf{l}_{0}}^{b} + \frac{\theta_{\mathbf{v}_{1}}^{b} - \theta_{\mathbf{v}_{2}}^{b}}{2\pi n}\right). \tag{E6}$$

Integrating over $\theta_{\mathbf{v}_{2}}^{b}$ yields

$$\mathcal{B} = 2\sum_{z_{\mathbf{l}_{0}}^{b}}\int_{0}^{2\pi n}d\theta_{\mathbf{v}_{1}}^{b}(-pq)z_{\mathbf{l}_{0}}^{b}\left(e^{-\left(z_{\mathbf{l}_{0}}^{b}-1+\frac{\theta_{\mathbf{v}_{1}}^{b}}{2\pi n}\right)^{2}n^{2}W^{2}} - e^{-\left(z_{\mathbf{l}_{0}}^{b}+\frac{\theta_{\mathbf{v}_{1}}^{b}}{2\pi n}\right)^{2}n^{2}W^{2}}\right)\bigg/\mathcal{N}'$$

$$= (4\pi n)(-pq)\int_{-\infty}^{\infty}dz\,e^{-z^{2}n^{2}W^{2}}/\mathcal{N}' = -\frac{pq}{\pi n}, \tag{E7}$$

where in the second line we have combined the summation over $z_{\mathbf{l}_{0}}^{b}$ and integral over $\theta_{\mathbf{v}_{1}}^{b}$ to make up an infinite integral. We thus find a constant Berry curvature all over the domain of integral. Substituting into (E4) confirms that the Hall conductivity $2\pi\sigma_{H} = -2pq/n$.

In the calculation above, the only approximation we made is that the overlap between different Gaussian wave packets is negligible. The errors introduced by such overlaps are always exponentially suppressed by a factor of $e^{-\pi^{2}/n^{2}W^{2}}$, and thus can be arbitrarily small in $W \sim \sqrt{\epsilon/U} \to 0$ limit. The errors introduced by using our trial wave functions rather than the exact ones, $O(\sqrt{\epsilon/U}, \epsilon/V, V/\sqrt{\epsilon U})$, are also controlled to be small in the said limit. Since the integral (E4) is the first Chern class of a U(1) principal bundle of the ground-state wave functions on the base manifold of a torus parametrized by the holonomies α_{x} and α_{y}, it must be quantized. So the

Hall conductivity must be the calculated fraction, as long as the excitation gap is not closed. This completes the proof of the exactness and robustness of the Hall conductivity.

APPENDIX F: IDENTIFYING THE TOPOLOGICAL ORDERS FOR THE MODELS IN REF. [15]

At the end of Sec. II we briefly mentioned the relation between this work and [15]. A family of Hamiltonians were introduced in [15], with the "$c = 1$" cases (in their notations) coinciding with the Hamiltonians we constructed (their "d" corresponds to our n), although the motivating reasons are quite different. In a followup study [16], the relation between the "$c = 1$" cases (in the absence of electromagnetic background) and the $\mathbb{Z}_n$ toric code was argued. On the other hand, the other cases of integer "$c \neq 1$" were more mysterious; the nature of their topological orders was not identified. Both the "$c = 1$" and "$c \neq 1$" models were studied with sign-free Monte Carlo numerical simulation; however, a connection to Chern-Simons theory was not established and solutions (or a path towards the solutions) were not obtained. In this Appendix, note the following:

(i) We first explain how to suitably identify the topological orders for the models considered in [15], especially the mysterious "$c \neq 1$" cases. We will show that in the topological limit the "$c \neq 1$" models do not lead to new topological orders; they reduce to certain "$c = 1$" cases.

(ii) Then, based on this observation, we sketch how to solve the "$c \neq 1$" Hamiltonians in the topological limit.

In Appendix B2 of [15] a mapping between their lattice models and Chern-Simons–type theories was attempted. However, even in the "$c = 1$" cases which coincide with our Hamiltonians, the attempted mapping disagrees with our (A1). More exactly, it was suggested in [15] that the lattice Hamiltonians map to Chern-Simons–type space-time lattice theories similar to our (A1), but with 1 instead of n in front of our second and third terms: the Dirac string couplings. Here we emphasize that the Dirac string coupling coefficient being n instead of 1 is important, for it determines the topological order. To see this, for simplicity let us turn off the electromagnetic coupling; then, if we sum out the Dirac strings s^a and s^b, the gauge fields a, b are indeed restricted to $(2\pi/n)\mathbb{Z} \mod 2\pi$, as expected for a $\mathbb{Z}_n$ toric code. On the other hand, if the Dirac string coupling coefficient were 1 instead of n, then a, b would have been reduced to multiples of 2π, but then the adb term would always have dropped out for being a multiple of 2π and the theory would have become topologically trivial.

To suitably identify the topological order for the models in [15], we should use their Eq. (20) instead. One may note that, upon rescaling their real valued α_1, α_2 by d (our n) and taking the small λ (our ϵ) limit, the topological]part of their Eq. (20) takes the form [in the following our integers (m, n) correspond to (c, d) in [15]]

$$\frac{mn}{2\pi} \sum_{plaq\ p} a_p(db)_p - n \sum_{plaq\ p} a_p s_p^b - n \sum_{link\ l} s_l^a b_l$$
$$+ \sum_{cube\ c} \theta_c^a(ds^b)_c + \sum_{vert\ v} \theta_v^b(\partial s^a)_v \qquad (F1)$$

which, when $m = 1$, is our (A1) (dropping the electromagnetic coupling for now). This addresses the nature of the topological orders for the $m = 1$ cases.

How about the more mysterious $m \neq 1$ cases? To understand them, we first consider the level $n' = mn$ doubled Chern-Simons theory on lattice, i.e., (F1) but with the Dirac string couplings being $n' = mn$ instead of n; this is simply the theory for the $\mathbb{Z}_{n'}$ toric code. Then, what does it mean to reduce the Dirac string coupling from $n' = mn$ to n? This corresponds to condensing those Wilson loops L^a and L^b in (A1) when they take values (charges) $n\mathbb{Z} \mod n'$ under a and b, respectively. Equivalently, this corresponds to gauging the 1-form $\mathbb{Z}_m \times \mathbb{Z}_m$ subgroup out of the global 1-form $\mathbb{Z}_{n'} \times \mathbb{Z}_{n'}$ symmetry [11]. When m does not divide n, this is in fact a classic example of gauging a 1-form symmetry with mixed anomaly. In general, when a theory contains gauge anomaly, some sectors of the anomalous theory will vanish, while some sectors will remain as a nonanomalous theory. Now we show that in the present case, the remaining nonanomalous theory is simply the theory of level $\tilde{n} = n/\gcd(m, n)$.

In the present case, the anomaly is manifested by the fact that the Wilson loops that we are trying to condense have a mutual statistics of

$$\exp\left[i\frac{2\pi}{n'}(nl^a)(nl^b)\right] = \exp\left[i\frac{2\pi n}{m}l^a l^b\right], \qquad (F2)$$

where $nl^a \mod n'$ is the a charge of some Wilson loop L^a, and $nl^b \mod n'$ is the b charge of some L^b that has linking number 1 with L^a. When m divides n, the mutual statistics is trivial and the Wilson loops can indeed be simultaneously condensed. In this case we arrive at a level $\tilde{n} = n/m$ theory: this can be seen by simply rescaling the real valued a, b in (F1) by $1/m$. In the opposite scenario where m, n are coprime and $m \neq 1$, for fixed l^a, summing over all possible l^b would yield a canceled contribution to the partition function unless $l^a = 0 \mod m$, i.e., $nl^a = 0 \mod n'$. This is equivalent to replacing $s^b \rightarrow ms^b$, i.e., restoring the coefficient $n' = mn$ in front of the as^b term in (F1), while still keeping n in front of the $s^a b$ term. This theory, however, is nothing but a level n theory, as can be seen from rescaling the real valued a by $1/m$. (The "particle-hole dual" $m = n - 1$ case mentioned in [16] is a special case of this scenario.) Now we are ready to explain the situation for general values of m and n. Define the integers $\tilde{m}$, $\tilde{n}$ to be the ones such that $m/\tilde{m} = n/\tilde{n} = \gcd(m, n)$, the greatest common divisor of m and n. For fixed l^a, summing over all possible l^b would yield a canceled contribution to the partition function unless $l^a = 0 \mod \tilde{m}$. But, this is equivalent to replacing $s^b \rightarrow \tilde{m}s^b$, i.e.,

$$\frac{mn}{2\pi} \sum_{plaq\ p} a_p(db)_p - \tilde{m}n \sum_{plaq\ p} a_p s_p^b - n \sum_{link\ l} s_l^a b_l$$
$$+ \sum_{cube\ c} \theta_c^a(ds^b)_c + \sum_{vert\ v} \theta_v^b(\partial s^a)_v. \qquad (F3)$$

Now we may rescale the real valued a and b by $1/m$ and $\tilde{n}/n$, respectively, and we find the theory in fact reduces to a level $\tilde{n}$ theory. This completes the identification of the topological orders for the models in [15]: the general values of "c" (our m) are topologically equivalent to "$c = 1$" but with a changed value of "d" (our n).

In the above we only considered the intrinsic topological order. But, then the coupling to electromagnetic background is straightforward. If the s^a and s^b in the original anomalous theory (F1) are coupled to A and $A^\star$ with coefficients q and p, respectively, then in the remaining nonanomalous level $\tilde{n}$ theory, the couplings are $\tilde{q} = q$ and $\tilde{p} = \tilde{m}p$, respectively. The Hall conductivity is then $-2\tilde{m}pq/\tilde{n} = -2mpq/n$, in agreement with the numerically computed current-current correlation from [15,16] when setting $q = p = 1$.

Now that we have understood the topological nature of the Hamiltonians for $m \neq 1$ ("i.e. $c \neq 1$" in [15]), we show that these Hamiltonians, which have the integer n replaced by the fraction n/m in both the commutation relations (2) and the Hamiltonian H, can still be controllably solved (applying the bounding function F to the V terms is understood) by our method with only slight modification. A potential problem is that the $\cos\left((n/m)a_{\mathbf{l}^\star}\right)$ and $\cos\left((n/m)b_{\mathbf{l}}\right)$ in the Hamiltonian no longer commute (now $[b_{\mathbf{l}}, a_{\mathbf{l}^\star}] = i2\pi m/n$) if m does not divide n. In the solving procedure, on each link we first solve $H_{\text{link},b}$ and find the narrow Gaussians (in the b basis) peaked around $\bar{b}_j = (2\pi m/n)j + \bar{b}_0$, $j \in \mathbb{Z}$. The noncommutativity

issues are manifested by the fact that the set of such narrow Gaussians with $j \in \mathbb{Z}$ no longer form an invariant subspace when we consider the shift terms $e^{\pm i(n/m)a}$ in $H_{\text{link},a}$, which shift a narrow Gaussian by a distance of 2π. However, this is not a substantial problem. It is not hard to see that the only consequence at low energy is that the low-energy states are then the ones formed by the peaks that are commensurate with each other. That is, the shifts only matter at low energy when they act for $\tilde{m}$ times, i.e., $(e^{\pm i(n/m)a})^{\tilde{m}}$, shifting a narrow Gaussian peak to another one located $2\pi\tilde{m}$ away, separated by $\tilde{n} - 1$ other peaks in-between. Thus, the only change is that effective period for the low-energy states is $2\pi\tilde{m}$ instead of 2π, and there are $\tilde{n}$ low-energy states on each link, as expected based on the previous identification of topological order. Corresponding, since we are only keeping $(e^{\pm i(n/m)a})^{\tilde{m}}$, the coupling of $e^{\pm i(n/m)a}$ to $e^{\mp ipA^\star}$ means p is effectively rescaled to $\tilde{m}p$. The remaining parts of the solution are unchanged. Hence, we still can controllably solve the Hamiltonian, and the solution, as expected from the reasoning before, is essentially the same as that for a level $\tilde{n}$ theory, with electromagnetic couplings q and $\tilde{m}p$.

[1] K. v. Klitzing, G. Dorda, and M. Pepper, Phys. Rev. Lett. **45**, 494 (1980).

[2] D. C. Tsui, H. L. Stormer, and A. C. Gossard, Phys. Rev. Lett. **48**, 1559 (1982).

[3] R. B. Laughlin, Phys. Rev. Lett. **50**, 1395 (1983).

[4] A. Kitaev, Ann. Phys. **303**, 2 (2003).

[5] M. A. Levin and X.-G. Wen, Phys. Rev. B **71**, 045110 (2005).

[6] X. Chen, Z.-C. Gu, Z.-X. Liu, and X.-G. Wen, Phys. Rev. B **87**, 155114 (2013).

[7] V. G. Turaev and O. Y. Viro, Topology **31**, 865 (1992).

[8] A. Kirillov, Jr, arXiv:1106.6033.

[9] A. Kitaev and L. Kong, Commun. Math. Phys. **313**, 351 (2012).

[10] A. Kapustin and L. Fidkowski, Commun. Math. Phys. **373**, 763 (2020).

[11] J.-Y. Chen, Commun. Math. Phys. **381**, 293 (2021).

[12] A. Kapustin and N. Saulina, Nucl. Phys. B **845**, 393 (2011).

[13] C.-H. Lin and M. Levin, Phys. Rev. B **89**, 195130 (2014).

[14] A. Kitaev, Ann. Phys. **321**, 2 (2006).

[15] S. D. Geraedts and O. I. Motrunich, Ann. Phys. **334**, 288 (2013).

[16] S. D. Geraedts and O. I. Motrunich, Phys. Rev. B **96**, 115137 (2017).

[17] M. DeMarco and X.-G. Wen, arXiv:2102.13057.

[18] E. Witten, arXiv:hep-th/0307041.

[19] X.-G. Wen, Adv. Phys. **44**, 405 (1995).

[20] D. Gaiotto, A. Kapustin, N. Seiberg, and B. Willett, J. High Energy Phys. 02 (2015) 172.

[21] M. Levin, F. J. Burnell, M. Koch-Janusz, and A. Stern, Phys. Rev. B **84**, 235145 (2011).

[22] Q. Niu, D. J. Thouless, and Y.-S. Wu, Phys. Rev. B **31**, 3372 (1985).

[23] R. Bezrukavnikov and A. Kapustin, Arnold Math. J. **5**, 15 (2019).

[24] Q.-R. Wang and M. Cheng, arXiv:2103.13399.